



Article Machine Learning-Based Prediction of Glioma *IDH* Gene Mutation Status Using Physio-Metabolic MRI of Oxygen Metabolism and Neovascularization (A Bicenter Study)

Andreas Stadlbauer^{1,2,3,*}, Katarina Nikolic^{1,4}, Stefan Oberndorfer^{1,4}, Franz Marhold^{1,5}, Thomas M. Kinfe^{3,6}, Anke Meyer-Bäse⁷, Diana Alina Bistrian⁸, Oliver Schnell³ and Arnd Doerfler⁹

- ¹ Karl Landsteiner University of Health Sciences, 3500 Krems, Austria; katarina.nikolic@stpoelten.lknoe.at (K.N.); stefan.oberndorfer@stpoelten.lknoe.at (S.O.); franz.marhold@stpoelten.lknoe.at (F.M.)
- ² Institute of Medical Radiology, Diagnostics, Intervention, University Hospital St. Pölten, 3100 St. Pölten, Austria
- ³ Department of Neurosurgery, Universitätsklinikum Erlangen, Friedrich-Alexander University (FAU) Erlangen-Nürnberg, 91054 Erlangen, Germany; thomasmehari.kinfe@uk-erlangen.de (T.M.K.); oliver.schnell@uk-erlangen.de (O.S.)
- ⁴ Division of Neurology, University Hospital St. Pölten, 3100 St. Pölten, Austria
- ⁵ Division of Neurosurgery, University Hospital St. Pölten, 3100 St. Pölten, Austria
- ⁶ Division of Functional Neurosurgery and Stereotaxy, Friedrich-Alexander University (FAU) Erlangen-Nürnberg, 91054 Erlangen, Germany
- ⁷ Department of Scientific Computing, Florida State University, 400 Dirac Science Library Tallahassee, Tallahassee, FL 32306-4120, USA; ameyerbaese@fsu.edu
- ⁸ Department of Electrical Engineering and Industrial Informatics, Politehnica University of Timisoara, 300006 Timisoara, Romania; diana.bistrian@fih.upt.ro
- ⁹ Department of Neuroradiology, Universitätsklinikum Erlangen, Friedrich-Alexander University (FAU) Erlangen-Nürnberg, 91054 Erlangen, Germany; arnd.doerfler@uk-erlangen.de
- Correspondence: andi@nmr.at; Tel.: +43-2742-9004-14185; Fax: +43-2742-9004-49300

Simple Summary: Early characterization of the isocitrate dehydrogenase (*IDHIDH*) gene mutation status of glioma is crucial for personalized decision making and prognosis in clinical neurooncological treatment. Based on the known differences in energy metabolism between *IDHIDH*-mutated and *IDHIDH*-wildtype gliomas, we assessed physio-metabolic magnetic resonance imaging-based measures along with machine learning for potential reliable presurgical characterization of *IDHIDH* gene status. Traditional machine learning algorithms and simple deep learning models trained in analyzing physio-metabolic parameters demonstrated the best performance in classifying the *IDHIDH* gene status of gliomas in independent internal testing. In contrast, external testing revealed that traditional machine learning differences in data acquisition methodology between the two sites. Our results outline the necessity of independent internal and external testing, thus calling for standardized and robust protocols for clinical MRI-driven AI applications.

Abstract: The mutational status of the isocitrate dehydrogenase (*IDH*) gene plays a key role in the treatment of glioma patients because it is known to affect energy metabolism pathways relevant to glioma. Physio-metabolic magnetic resonance imaging (MRI) enables the non-invasive analysis of oxygen metabolism and tissue hypoxia as well as associated neovascularization and microvascular architecture. However, evaluating such complex neuroimaging data requires computational support. Traditional machine learning algorithms and simple deep learning models were trained with radiomic features from clinical MRI (cMRI) or physio-metabolic MRI data. A total of 215 patients (first center: 166 participants + 16 participants for independent internal testing of the algorithms versus second site: 33 participants for independent external testing) were enrolled using two different physio-metabolic MRI protocols. The algorithms trained with physio-metabolic data demonstrated the best classification performance in independent internal testing: precision, 91.7%; accuracy, 87.5%; area under the receiver operating curve (AUROC), 0.979. In external testing, traditional machine learning models trained with cMRI data exhibited the best *IDH* classification results: precision, 84.9%;



Citation: Stadlbauer, A.; Nikolic, K.; Oberndorfer, S.; Marhold, F.; Kinfe, T.M.; Meyer-Bäse, A.; Bistrian, D.A.; Schnell, O.; Doerfler, A. Machine Learning-Based Prediction of Glioma *IDH* Gene Mutation Status Using Physio-Metabolic MRI of Oxygen Metabolism and Neovascularization (A Bicenter Study). *Cancers* 2024, *16*, 1102. https://doi.org/10.3390/ cancers16061102

Academic Editor: David Wong

Received: 29 January 2024 Revised: 21 February 2024 Accepted: 23 February 2024 Published: 8 March 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). accuracy, 81.8%; and AUROC, 0.879. The poor performance for the physio-metabolic MRI approach appears to be explainable by site-dependent differences in data acquisition methodologies. The physio-metabolic MRI approach potentially supports reliable classification of *IDH* gene status in the presurgical stage of glioma patients. However, non-standardized protocols limit the level of evidence and underlie the need for a reproducible framework of data acquisition techniques.

Keywords: artificial intelligence; deep learning; glioma; isocitrate dehydrogenase gene; *IDH* gene mutation; neurooncology; machine learning; physio-metabolic MRI; preoperative classification

1. Introduction

Glioma represents the most common primary brain tumor and arises from glial cells that surround and support neurons such as astrocytes, oligodendrocytes, and ependymal cells. Treatment planning for glioma depends on the World Health Organization (WHO) classification system of central nervous system (CNS) tumors, with gliomas designated as "low grade" (WHO grade 1 or 2) or "high grade" (WHO grade 3 or 4) based on the growth potential and aggressiveness of the tumor. Glioblastoma (WHO grade 4) is the most aggressive primary brain tumor in adults. The median glioblastoma survival time is 14–17 months [1]. Only 25% of patients survive more than one year after initial diagnosis and only 5% survive more than five years [2]. The latest (fifth) edition of the WHO classification system of CNS tumors published in 2021 [3] considered the isocitrate dehydrogenase (*IDH*) gene mutation status relevant for the diagnostic principles and nomenclature, with important implications for clinical practice such as treatment, patient prognosis, and the conceptualization/interpretation of clinical trials.

In the new classification, glioblastomas comprise only IDH-wildtype tumors. In addition, IDH-wildtype diffuse astrocytic tumors in adults without the histologic features of glioblastoma but with one or more of three genetic parameters (TERT promoter mutation, EGFR gene amplification, or combined gain of entire chromosome 7 and loss of entire chromosome 10) are also classified as glioblastoma. This has important implications for the prognosis and therapy of IDH-wildtype astrocytoma WHO grade 2 or 3. On the other hand, IDH-mutant gliomas are characterized by a class-defining and possibly tumor-initiating clonal IDH1 or IDH2 gene mutation [4,5]. These encompass astrocytomas (IDH-mutant, WHO grade 2, 3, or 4) and oligodendrogliomas (IDH-mutant and 1p/19q-codeleted, WHO grade 2 or 3) [3]. Patients with IDH-mutant glioma have better clinical outcomes compared with patients suffering from *IDH*-wildtype glioma [5,6], suggesting that there are unique physiological characteristics of *IDH*-mutant gliomas that may make them particularly vulnerable to specific therapies. The *IDH* gene mutation arises early during gliomagenesis, persists throughout the lifespan of the tumor, and confers neomorphic enzymatic activity [7], promoting cellular metabolism changes [8–10]. Like other malignant tumors, *IDH*-wildtype glioma is characterized by a high level of aerobic glycolysis, also known as the Warburg effect. Hypoxia induces upregulation of the transcription factor HIF1 α , leading to overexpression of glycolysis-related proteins and enzymes, thereby accelerating glycolysis. In *IDH*-mutated glioma, however, the HIF1 α response to hypoxia is blunted, shifting the metabolism to oxidative phosphorylation. A 13 C metabolic flux analysis [10] on a panel of IDH-mutated glioma cell lines observed increased oxidative metabolism in the Krebs cycle as well as reduced tumor growth rates under hypoxic conditions, suggesting *IDH*1 mutant gliomas prefer a more oxygenated microenvironment for proliferation. This was confirmed by the results of an imaging study in glioma patients [11] using simultaneous pH- and oxygen-sensitive molecular magnetic resonance imaging (MRI), which suggest that mutation of the *IDH* gene is associated with lower tumor acidity and lower vascular hypoxia.

This separation into *IDH*-wildtype and *IDH*-mutated gliomas is an important advance. However, it places particular emphasis on neuropathology laboratories having access to adequate molecular testing and the ability to obtain results in a timely manner.

Molecular imaging is a promising method for non-invasively investigating the metabolic features of different types of tumors, permitting the distinction between tumor progression and treatment effects [12]. The known physiological connection between neovascularization and tissue hypoxia in brain tumors [13] reflected our rationale for developing a physiometabolic MRI mapping tool for a combined quantitative assay of neovascularization activity, microvascular architecture, oxygen metabolism, and CNS tissue hypoxia [14–16].

In order to obtain deeper insight into the pathophysiology of brain tumors, our approach combines multiparametric quantitative blood oxygenation level-dependent (qBOLD) MRI [17] and vascular architecture mapping (VAM) [18]. The physical basis for the VAM approach is the vessel diameter-dependent sensitivity of spin echo and gradient echo MRI sequences [19]: spin echo perfusion measurements exhibit a peak sensitivity to the microvasculature at a vessel diameter of approximately 10 mm, whereas gradient echo perfusion measurements, which are commonly used in clinical routine, are dominated by larger vessel diameters (>20 mm) [19]. Consequently, the skillful combination of both spin echo and gradient echo perfusion MRI, as used in our physio-metabolic MRI approach, allows for the examination of the entire physiological and pathophysiological range of vessel diameters and structures in the human brain.

Inclusion of innovative neuroimaging methods into clinical routine, however, requires the assessment of a large amount of additional complex neuroimaging data and is time consuming. A timely evaluation of this helpful diagnostic information requires computational support. Methods from artificial intelligence (AI) such as deep learning (DL) and traditional machine learning (ML) may offer new options for timely support of clinicians. AI-based analysis is independent of the individual experience level of the evaluating clinician, helps to increase the comparability of results, and makes it possible to cope with large amounts of data. Finally, radiomics extracts semi-quantitative and quantitative features from neuroimaging data that are beyond human perception [20–22].

In recent years, deep learning has developed tremendously and has become widely used in various fields such as natural language processing or image processing [23,24]. Convolutional neural networks (CNNs) execute tasks by scanning the data with various filters in layers for convolution, pooling, and normalization to extract abundant features [25]. CNNs, however, are limited in learning dependencies between distant data points; thus, it is challenging for CNNs to directly learn the most important and representative features from data. Recurrent neural networks (RNNs) are deep learning models typically used to solve problems with longer sequential input data such as time series or radiomic feature vectors [26,27], thereby circumventing the mentioned limitations of CNNs. However, one of the biggest challenges with RNNs is vanishing or exploding gradients. These occur when the gradients of the loss function with respect to the parameters of the RNN become very small or very large as they propagate through the sequential input data. This can make it difficult to train the RNN effectively, as the updates to the parameters will be very small or very large and the network will not learn effectively. Long short-term memory (LSTM) is an RNN architecture [28,29] that has been widely adopted for sequential data processing as its design specifically solves the gradient vanishing/exploding problem of RNNs and aims to provide a short-term memory for RNNs that can last thousands of time steps, hence the name "long short-term memory". LSTMs have been employed for medical image analysis recently [30,31]. For example, Zou et al. [32] used an LSTM model for pharmacokinetic parameter estimation in head and neck cancer using a dynamic contrast-enhanced MRI sequence.

Building on these findings and our preliminary work, we aimed to determine whether the combined use of deep learning algorithms and physio-metabolic MRI data enables a reliable preoperative classification of the *IDH* gene mutation status of glioma patients in a clinical setting. To test this hypothesis, traditional ML algorithms and deep learning algorithms including a CNN and an LSTM were trained with conventional MRI data and physiometabolic MRI data for oxygen metabolism and neovascularization from 166 consecutive glioma patients. The algorithms were tested with data from an independent internal test cohort as well as with data from an independent external test cohort using a different MRI data acquisition protocols.

2. Materials and Methods

2.1. Patients

The prospectively populated institutional MR neuroimaging databases of two sites were searched to identify patients with untreated brain tumors that were newly diagnosed between January 2016 and May 2023 (University Clinic St. Pölten) and June 2016 and October 2018 (University Clinic Erlangen), respectively. Inclusion criteria were as follows: (i) age \geq 18 years; (ii) histopathological confirmation of adult-type diffuse glioma based on the 2021 WHO Classification of Tumors of the CNS [3], including conclusive information about isocitrate dehydrogenase 1 and 2 (*IDH*1 and 2) gene mutation status, i.e., glioblastoma *IDH*-wildtype WHO grade 4, astrocytoma *IDH*-mutant WHO grade 2–4, or oligodendroglioma *IDH*-mutant WHO grade 2–3; (iii) no previous treatment of the brain tumor; and (iv) a preoperative MRI examination with the study protocol.

The brain tumor MRI database at the University Clinic St. Pölten contained a total of 2156 MRI examinations according to the study protocol in 681 brain tumor patients for the period from January 2016 to May 2023. A total of 182 patients (89 females; 93 males; mean age 57.5 \pm 15.5 years; 19–89 years) with newly diagnosed, untreated glioma fulfilled/satisfied the inclusion criteria.

Patients examined from January 2016 to January 2023 at the University Clinic St. Pölten were selected as the training cohort. Of the total of 166 patients (81 females; 85 males; mean age 58.1 \pm 15.2 years; 19–89 years) during this period, 123 patients (74%; 59 females; 64 males; mean age 63.2 \pm 12.3 years; 23–89 years) were diagnosed with a glioma without mutation of the *IDH* gene (*IDH*wt), and 43 patients (26%; 22 females; 21 males; mean age 43.3 \pm 12.8 years; 19–73 years) were diagnosed with a glioma with mutation of the *IDH* gene (*IDH*wt). The difference in mean age between these two groups was significantly different (*p* < 0.001), as determined by a Student's *t* test.

Of the 123 patients with a diagnosis of an *IDH*wt glioma, 116 patients had a glioblastoma WHO grade 4, five patients had an astrocytoma WHO grade 3, and two patients had an astrocytoma WHO grade 2. The 43 patients suffering from *IDH*-mutated glioma had the following histological diagnoses: eleven patients had an oligodendroglioma WHO grade 2, five patients had an oligodendroglioma WHO grade 3, six patients had an astrocytoma WHO grade 2, thirteen patients had an astrocytoma WHO grade 3, and eight patients had an astrocytoma WHO grade 4.

Patients who were examined at the University Clinic St. Pölten from February 2023 to May 2023 were selected as an independent internal test cohort. Of the total 16 patients (eight females; eight males; mean age 51.9 ± 18.6 years; 19–76 years) during this time period, twelve patients (75%; five females; seven males; mean age 53.4 ± 20.8 years; 19–76 years) had a diagnosis of *IDH*wt glioma (ten glioblastoma, one astrocytoma WHO grade 3, and one astrocytoma WHO grade 2) and four patients (25%; three females; one male; mean age 47.3 ± 10.9 years; 31-54 years) had a diagnosis of *IDH*mut glioma (two oligodendroglioma WHO grade 3).

All patients examined at the FAU Erlangen–Nürnberg (from June 2016 to October 2018) who met the inclusion criteria were selected as an independent external test cohort. Of the total 33 patients (seven females; twenty-six males; mean age 54.7 ± 15.3 years; 27–78 years), 21 patients (64%; four females; seventeen males; mean age 62.7 ± 10.5 years; 45–78 years) had a diagnosis of *IDH*wt glioma (twenty glioblastoma and one astrocytoma WHO grade 2) and twelve patients (36%; three females; nine males; mean age 40.6 ± 11.6 years; 27–67 years) had a diagnosis of *IDH*wt glioma (four oligodendroglioma

WHO grade 3, three astrocytoma WHO grade 2, three astrocytoma WHO grade 3, and two astrocytoma WHO grade 4). The characteristics of the patient cohorts are summarized in Supplementary Table S1.

2.2. MRI Data Acquisition

2.2.1. The MRI Study Protocol at the University Clinic St. Pölten

All data of the MRI study protocol were acquired on a clinical 3 Tesla scanner (Trio, Siemens, Erlangen, Germany) using the conventional 12-channel head coil provided by the vendor.

Clinical MRI (cMRI) for routine diagnosis of brain tumors included, among others, an axial fluid-attenuated inversion–recovery (FLAIR) sequence, high-resolution pre- and post-contrast enhanced T1-weighted (ceT1w) sequences, an axial diffusion-weighted imaging (DWI) sequence, and a gradient echo dynamic susceptibility contrast (GE-DSC) perfusion MRI sequence with 60 dynamic imaging volumes during administration of 0.1 mmol/kg-bodyweight gadoterate–meglumine (Dotarem, Guerbet, Villepinte, France).

MRI-based oxygen metabolism was assessed using the quantitative blood-oxygenlevel-dependent (qBOLD) imaging approach [16], which additionally included the acquisition a multi-echo GE sequence and a multi-echo SE sequence to map the transverse relaxation rates, R_2^* (=1/ T_2^*) and R_2 (=1/ T_2), respectively.

MRI-based assessment of microvascular architecture and neovascularization activity was performed with the vascular architecture mapping (VAM) technique [18], which additionally included a spin echo DSC (SE-DSC) perfusion MRI sequence with the same parameters and contrast agent injection protocol as described for the routine GE-DSC perfusion MRI. Adverse effects due to contrast media leakage and differences in time to first-pass peak that may significantly affect the data evaluation were minimized as described previously [33,34].

All MRI sequences for qBOLD and VAM were performed with identical geometric parameters (number of slices, voxel size, etc.) and slice position as the routine GE-DSC perfusion sequence. The sequences required a total of seven minutes of additional scanning time.

2.2.2. The MRI Study Protocol at the FAU Erlangen-Nürnberg

The MRI study protocol at the FAU Erlangen-Nürnberg differed from that at the University Clinic St. Pölten. The sequence parameters at the two sites are compared in Supplementary Table S2. MRI data acquisition at the FAU was also performed with a 3 Tesla scanner (Trio, Siemens, Erlangen, Germany), but in combination with a 32-channel head coil.

The cMRI protocol for routine diagnosis at the FAU also included axial FLAIR, ceT1w, and DWI sequences. Instead of performing the GE-DSC and the SE-DSC sequences separately with two administrations of contrast media, a hybrid single-shot gradient echo–spin echo (GESE) DSC sequence was used at the FAU. GESE-DSC perfusion examinations were performed with 80 dynamic measurements during administration of a double dose (0.2 mmol/kg bodyweight) of gadoterate meglumine (Dotarem, Guerbet, France). The GESE-DSC raw data were then separated into a GE-DSC and an SE-DSC dataset, respectively. We refer to Supplementary Table S2 for more details.

This GESE-DSC sequence has the advantage that two separate SE- and GE-DSC acquisitions are not required. This reduces sensitivity to patient motions and variations in injection timing. The separate acquisition of the SE and GE-DSC sequences, on the other hand, enables the measurement of VAM data with a high signal-to-noise ratio (SNR), high spatial resolution, and coverage of the entire brain. This is due to the fact that with separate acquisition, the full data acquisition potential is available exclusively for the respective sequence and does not have to be shared. The combined GESE-DSC sequence is therefore less suitable for use in clinical routine.

The sequences for the qBOLD imaging approach for MRI-based oxygen metabolism assessment differed at the FAU, especially in terms of repetition times (TR) and the number of slices (Supplementary Table S2). TR has an especially significant impact on the data.

2.3. MRI Data Processing and Calculation of MRI Biomarker Maps

At both sites, MRI data processing and calculation of MRI biomarker maps were performed using in-house software (PrecisionMRI version 6.1 beta) developed with MatLab (MathWorks, Natick, MA, USA). Processing of the cMRI data included calculation of cerebral blood flow (CBF) and volume (CBV) maps from the GE-DSC perfusion MRI data using automatic identification of the arterial input functions (AIFs) [35,36]. Furthermore, apparent diffusion coefficient (ADC) maps were calculated from DWI data using a mono-exponential model (two-point b values of 0 and 1000 s/mm²) and the following equation:

$$ADC = -\frac{1}{b} \cdot \ln\left(\frac{S}{S_0}\right) \tag{1}$$

where S_0 is the signal intensity of no diffusion gradients (b = 0 s/mm²), S is the signal intensity with diffusion gradients (b = 1000 s/mm²), and b is the b value.

Processing of qBOLD data for MRI-based oxygen metabolism included corrections for background field artifacts in the R₂*-mapping data [37] and for stimulated echoes in the R₂-mapping data [38], followed by calculation of R₂* and R₂ maps. Biomarker maps of tissue oxygen metabolism including the oxygen extraction fraction (OEF) and cerebral metabolic rate of oxygen (CMRO₂) [39] were calculated using the following equations:

$$OEF = \frac{R_2^* - R_2}{\frac{4}{3} \cdot \pi \cdot \gamma \cdot \Delta \chi \cdot Hct \cdot B_0 \cdot CBV}$$
(2)

and

$$CMRO_2 = \frac{C_a \cdot CBF}{k \cdot CBV} \cdot (R_2^* - R_2)$$
(3)

where $\Delta \chi = 0.264 \cdot 10^{-6}$ is the difference between the magnetic susceptibilities of fully oxygenated and fully deoxygenated haemoglobin; Hct = 0.42 \cdot 0.85 is the microvascular hematocrit fraction (the factor 0.85 stands for a correction factor of systemic Hct for small vessels); $\gamma = 2.67502 \cdot 10^8 \text{ rad/s/T}$ is the nuclear gyromagnetic ratio; and Ca = 8.68 mmol/mL is the arterial blood oxygen content [40]. Maps of capillary oxygen tension (capiPO₂) and mitochondrial oxygen tension (mitoPO₂) [41,42] were determined with:

$$capiPO_2 = P_{50} \sqrt[h]{\left(\frac{2}{OEF} - 1\right)}$$
(4)

and

mitoPO₂ = P₅₀
$$\sqrt[h]{\left(\frac{2}{OEF} - 1\right)} - \frac{CMRO_2}{L}$$
 (5)

where h is the Hill coefficient of oxygen binding to hemoglobin (h = 2.7), P_{50} is the hemoglobin half-saturation tension of oxygen (27 mmHg), and L (4.4 mmol/Hg per minute) is the tissue oxygen conductivity as defined by Vafaee and Gjedde [43].

Processing VAM data for MRI-based microvascular architecture and neovascularization activity included correction for remaining contrast agent extravasation as described previously [33,34], fitting of the first bolus curves for each voxel of the GE- and SE-DSC perfusion MRI data with a previously described gamma-variate function [44], and calculation of the the $\Delta R_{2,GE}$ versus ($\Delta R_{2,SE}$)^{3/2} diagram, also known as the vascular hysteresis loop (VHL). [45]. Biomarker maps of microvascular architecture [46] including the microvessel density (MVD) and vessel size index (VSI, i.e., microvessel radius) were calculated using the VHL curve data and the following equations:

$$MVD = \frac{Q_{max}}{\beta} \cdot \left(\frac{CBV^2}{4\pi^2 \cdot ADC \cdot R}\right)^{1/3}$$
(6)

and

$$VSI = \left(\frac{CBV \cdot ADC \cdot \beta^3}{2\pi \cdot Q_{max}^3}\right)^{1/2}$$
(7)

where $Q_{max} = max[\Delta R_{2,GE}]/max[(\Delta R_{2,GE})^{3/2}]$, β is a numerical constant ($\beta = 1.6781$), and $-\frac{4}{R} \approx 3.0 \ \mu$ m is the mean vessel lumen radius [46]. Neovascularization activity was esti-

mated by the microvessel type indicator (MTI), which was previously [18] defined as the area of the VHL curve signed with the rotational direction of the VHL curve (a clockwise VHL direction was identified with a plus-sign and a counter-clockwise VHL direction was identified with a plus-sign and a counter-clockwise VHL direction was identified with a plus-sign and a counter-clockwise VHL direction was identified with a plus-sign and a counter-clockwise VHL direction was identified with a minus-sign [18]). Previous studies [18,47] demonstrated that a positive MTI value (assigned to warm colors in the MTI maps) is associated with a vascular system that is dominated by arterioles, whereas a negative MTI value (cool colors in the MTI maps) is associated with venule- and capillary-like vessel components. Finally, the map for the microvascular cerebral blood volume (μ CBV) was calculated from the SE-DSC perfusion MRI data via a separate automatic identification of AIFs [36].

In summary, MRI data processing resulted in the following data sets for each patient for further data analysis:

- Four cMRI data sets: FLAIR and ceT1w MRI data as well as the ADC maps for microstructural density and the CBV maps for macrovascular perfusion.
- Four biomarker maps for oxygen metabolism: MRI-based tissue oxygen metabolism (OEF and CMRO₂) as well as MRI-based capillary oxygen tension (capiPO₂) and mitochondrial (tissue) oxygen tension (mitoPO₂).
- Four biomarker maps for microvascular architecture and neovascularization activity: microvascular density (MVD), microvascular diameter (VSI), microvascular perfusion (μCBV), and microvascular type (MTI).

2.4. Radiomic Feature Extraction

The data processing pipeline is summarized in Supplementary Figure S1, and the procedures have been described in more detail in the Supplementary Materials as well as in previous publications [48]. Summarized briefly, the anatomical cMRI data and biomarker maps of a patient were loaded into the open-source software platform 3D Slicer (v. 4.11). The tumor volume was segmented on ceT1w and FLAIR MRI data, respectively. The anatomical MRI data were z-score normalized [49,50], resampled into a uniform voxel size of $1 \times 1 \times 1$ mm³, and grey-level discretization was performed [51] with a bin width value of 0.1 resulting in histograms with approximately 60 bins. Biomarker maps, however, represented quantitative imaging data with a range of physiological reasonable values. Individually adapted thresholds were applied to the biomarker maps, and biomarker value discretization was performed with adapted bin width values in order to obtain histograms with 60–67 bins. Supplementary Table S3 summarizes the value ranges and bin sizes used for the biomarker maps.

Next, 107 radiomic features were extracted for every data set from the segmented tumor volume using the Python package PyRadiomics [52], with 14 shape features, 18 first-order features, and 75 texture features, respectively. A detailed overview and description of these radiomic features can be found in the Supplementary Materials (Supplementary Table S4) and elsewhere [48]. Procedures and features were in accordance with the Imaging Biomarker Standardization Initiative (IBSI) [53]. Mathematical formulas have been described on the website of the package (https://pyradiomics.readthedocs.io; accessed on 22 February 2023).

The feature vectors for ceT1w, FLAIR, ADC, and CBV were concatenated to the cMRI feature vector, those for OEF, CMRO₂, capiPO₂, and mitoPO₂ were concatenated to the feature vector of oxygen metabolism, and the feature vectors for vasculature and neovascularization included the features for μ CBV, MVD, VSI, or MTI. Thus, all three of these resulting feature vectors contained 428 features per patient.

2.5. Traditional Machine Learning

Traditional machine learning, including radiomic feature selection and model development, was performed using the open-source software package Weka (version 3.8.5, University of Waikato, Hamilton, New Zealand). For the selection of radiomic features, we proceeded in two steps and followed the strategy combining advantages of both the ranking methods and the learner-based methods as described previously [54].

Firstly, we applied the RefiefF attribute evaluation filter in combination with the Ranker search method using the following parameters: number of nearest neighbors for attribute estimation = 10; random seed for sampling instances; sigma that sets the influence of nearest neighbors = 2; and no weighting of nearest neighbors by their distance. The top 25% of features in the ranking were selected and used in the next step.

Secondly, we applied the learner-based feature selection method Wrapper using the respective ML algorithm that was used for model development (we refer to the next chapter for a detailed description) as classifier. The BestFirst search method was used in combination with 10-fold cross-validation. Furthermore, we used accuracy as a measure to evaluate the performance of attribute combinations and a threshold of 0.01. Those features that were selected at least once in the rankings were selected as the most suitable features specifically for each ML algorithm, i.e., an individual feature list was generated for each ML classifier.

As we expected unbalanced classes due to different prevalence and well-known differences in patient numbers at our institution for the brain tumor entities, the synthetic minority oversampling technique (SMOTE) [55] was employed in order to balance the classes using five nearest neighbors and one seed for random sampling. This was followed by the application of a randomize filter.

For model development, only those algorithms were considered that had shown top performance in brain tumor classification in a previous study, namely:

- A multilayer perceptron (MLP) with one hidden layer and number of neurons = number of features + number of classes;
- Adaptive boosting (ABoost) using decision tree "J48" as classifier; and
- Random forest (RF).

Short descriptions of these traditional ML algorithms are summarized in the Supplementary Materials. A list of optimized hyperparameters is provided in Supplementary Table S5. Four categories of data were used as training data sets:

- cMRI (ceT1w, FLAIR, ADC, CBV);
- MRI-based oxygen metabolism (CMRO₂, OEF, capiPO₂, mitoPO₂);
- MRI-based vascular architecture (µCBV, MVD, MTI, VSI); and
- The combination of MRI-based oxygen metabolism and vascular architecture.

A 10-fold cross-validation procedure was used to train and validate the models.

2.6. Deep Learning

For the development of the deep learning algorithms, i.e., 1D-CNN and LSTM networks, the open-source data analysis platform KNIME (Konstanz Information Miner; version 4.7.4) was used, which is widely known and performs well in biomedical and healthcare applications [56–58]. The input data for the deep learning algorithms were the full radiomic feature vectors for cMRI, MRI-based oxygen metabolism, and MRI-based vascular architecture, respectively. Each feature vector included 428 (4 \times 107) features.

We additionally used the combination of MRI-based oxygen metabolism and vascular architecture, which included 856 (2×428) features.

CNNs are the most widely used architectures in deep learning approaches and are generally composed of an input layer, several hidden layers for convolution, pooling, and fully connected "dense" layers, and an output layer. Our CNN for prediction of the IDH gene mutation status of glioma was based on a one-dimensional CNN (1D-CNN) [59]. It was characterized by convolution kernels that slide in only one dimension over the elements of the input pattern, i.e., the feature vectors for cMRI, oxygen metabolism, vascular architecture, or the combination of the latter two. The architecture of the 1D-CNN included nine hidden layers: three convolutional layers, three max-pooling layers, one flatten layer, and two fully connected (dense) layers. A rectified linear unit (ReLU) was used as an activation function for all convolutional layers and the first dense layer [60]. A sigmoid function was applied in the final dense layer for activation and calculation of an output with one unit for binary classification: *IDH*-wildtype (*IDH*wt) or *IDH*-mutant (*IDH*mut). After the second and third convolution layer as well as after the first dense layer, three dropout rates of 0.2 were used in order to avoid overfitting [61]. The binary cross-entropy was considered a loss function, and Adam, an adaptive learning rate optimization algorithm specifically designed for deep neural network training, was exploited as an optimizer. The parameters are summarized in Supplementary Table S6. The CNN was trained with 80% and validated with 20% of the SMOTE balanced training data (the same as used for the ML models). The data were shuffled before each epoch (up to 300) in order to ensure the robustness of the models.

LSTMs learn (long) distant dependencies within the sequential input data using a gating mechanism that controls the memorizing process. An LSTM unit is composed of a cell that can store, read, and write information via a forget gate, an input gate, and an output gate that regulate the flow of information into and out of the cell. The forget gate decides which information from a previous state needs attention and which can be ignored. The input gate learns new sets of information from the incoming cell and decides which pieces of new information to store in the current state. The output gate controls which pieces of information in the current state to output as the next hidden state.

The inputs to LSTM were again the combined feature vectors from the cMRI data, MRI-based oxygen metabolism, and MRI-based vascular architecture, respectively, of the *IDH*wt and *IDH*mut glioma, with 4×107 (428) features each.

The LSTM models consisted of an input layer, four LSTM [62] layers (eight layers for the combination of MRI-based oxygen metabolism and vascular architecture), and two fully connected (dense) layers. The serial direction of an LSTM layer was occupied by the 107 radiomic features from one of the biomarker volumes. For all LSTM layers, the hyperbolic tangent (tanh) was the activation function, the hard sigmoid was the recurrent activation function, and 0.2 was the forward and recurrent dropout rate, respectively. The dense layers applied a sigmoid function for activation with a dropout rate of 0.2. The second dense layer had an output with one unit for binary classification of *IDHwt* or *IDH*mut. The LSTM network was also trained with 80% and validated with 20% of the SMOTE balanced training data using the Adam optimizer with an initial learning rate of 0.001, beta1 of 0.9, beta2 of 0.999, and epsilon of 10^{-8} . The training batch size was set to 28, the validation batch size was set to 10, and each adapted to the number of cases in the training and validation data sets. The number of epochs was 300. A list of optimized hyperparameters is provided in Supplementary Table S5.

2.7. Model Performance Testing

Data from the University Clinic St. Pölten were used as independent internal test cohort, and data from the FAU Erlangen-Nürnberg were used as independent external test cohort with different MRI examination protocols. For performance evaluation, confusion matrix-derived metrics including accuracy, sensitivity (aka the recall or true positive rate), specificity (aka the true negative rate), precision (aka the positive predictive value), and the

10 of 23

F-score, as well as the area under the receiver operating characteristic curve (AUROC) [63], were calculated. Details are provided in Supplementary Figure S2. Because the data sets for both independent internal and independent external testing were unbalanced, we calculated the weighted mean over the performance metrics for the *IDH*-wildtype and *IDH*-mutated gliomas.

3. Results

3.1. The Selected Radiomic Features

The feature selection in the training data for the traditional ML algorithms resulted in the following feature distributions. For cMRI: 62% ceT1w features, 1% FLAIR features, 16% ADC features, and 21% CBV features. For MRI-based oxygen metabolism: 32% CMRO2 features, 9% OEF features, 27% capiPO2 features, and 32% mitoPO2 features. For MRI-based vascular architecture: 14% μ CBV features, 37% MVD features, 31% MTI features, and 18% VSI features.

During training, the loss curves of all selected algorithms showed that the training models converged. The accuracy values for the validation of the trained models are summarized in Table 1. The LSTM algorithm in combination with the cMRI data showed the highest accuracy in validating the trained models, followed by the MLP with the biomarkers for oxygen metabolism and the RF model with the data for vascular architecture, respectively. In general, MLP and RF showed the best accuracy in training validation, while ABoost and the CNN showed the worst accuracy.

Table 1. Accuracy values for validation of the trained models.

	ABoost	MLP	RF	CNN	LSTM
Clinical MRI	0.858	0.866	0.907	0.891	0.94
Oxygen Metabolism	0.87	0.907	0.902	0.855	0.88
Vascular Architecture Mapping	0.85	0.902	0.907	0.891	0.88
Oxy Met & VAM	0.829	0.902	0.898	0.873	0.86

Oxy Met = Oxygen metabolism; VAM = vascular architecture mapping; ABoost = adaptive boosting; MLP = multilayer perceptron; RF = random forest, CNN = convolutional neural network; LSTM = long short-term memory.

3.2. Testing with an Independent Internal Cohort

For the independent internal test cohort of 16 patients examined at the University Clinic St. Pölten, both the LSTM trained with the MRI-based oxygen metabolism and the MLP trained with the combination of oxygen metabolism and vascular architecture showed the best classification performance: two errors (2 *IDH*wt glioma), an AUROC of 0.979, an accuracy of 0.875, and a precision of 0.917. The CNN trained with the data for vascular architecture and the RF model trained with the combination of oxygen metabolism and vascular architecture also showed only two errors, although one was for *IDH*wt glioma and one was for *IDH*mut glioma. This means that all of the best-performing algorithms were trained with physio-metabolic MRI data. Furthermore, most algorithms (five out of six) with medium classification quality (three errors) were also trained with physio-metabolic data. Only the LSTM trained on cMRI data also showed medium-rank classification quality. Poor classification quality (four and five errors out of 16 patients) in the tests with the independent internal cohort was particularly evident in algorithms trained with cMRI data (four out of five). However, six out of fifteen algorithms trained with physio-metabolic data also showed poor classification quality when tested with the independent internal cohort. These primarily included CNN in combination with oxygen metabolism and ABoost in combination with VAM.

The values for accuracy, sensitivity, specificity, precision, F-score, AUROC, and classification errors for the five *IDH* gene status classifiers and the four MRI data sets (cMRI; oxygen metabolism; VAM; combination of oxygen metabolism & VAM) when tested with the independent internal test cohort are presented as heat maps in Figure 1. Confusion matrices are provided in Supplementary Figure S3.

	Clinical MRI					Oxygen Metabolism					Vascula		Oxy Met & VAM									
	ABoost	MLP	RF	CNN	LSTM	ABoost	MLP	RF	CNN	LSTM	ABoost	MLP	RF	CNN	LSTM	ABoost	MLP	RF	CNN	LSTM		
Accuracy	0.688	0.750	0.750	0.750	0.813	0.813	0.813	0.750	0.688	0.875	0.688	0.813	0.813	0.875	0.813	0.750	0.875	0.875	0.688	0.750		0.98
Sensitivity	0.688	0.750	0.750	0.750	0.813	0.813	0.813	0.750	0.688	0.875	0.688	0.813	0.813	0.875	0.813	0.750	0.875	0.875	0.688	0.750		
Specificity	0.729	0.750	0.750	0.750	0.813	0.813	0.813	0.750	0.563	0.958	0.729	0.813	0.813	0.792	0.813	0.750	0.958	0.792	0.563	0.750		
Precision	0.774	0.800	0.800	0.800	0.832	0.832	0.832	0.800	0.714	0.917	0.774	0.832	0.832	0.875	0.832	0.800	0.917	0.875	0.714	0.800		
F score	0.708	0.764	0.764	0.764	0.819	0.819	0.819	0.764	0.698	0.882	0.708	0.819	0.819	0.875	0.819	0.764	0.882	0.875	0.698	0.764		
AUROC	0.771	0.750	0.750	0.750	0.833	0.833	0.833	0.750	0.625	0.979	0.771	0.833	0.833	0.875	0.833	0.750	0.979	0.875	0.625	0.750		0.56
Errors	5	4	4	4	3	3	3	4	5	2	5	3	3	2	3	4	2	2	5	4		2

Figure 1. Overview of the results for the testing with the independent internal patient cohort. Heat maps depicting (top down) accuracy, sensitivity, specificity, precision, F-score, AUROC, and classification errors for the traditional ML algorithms adaptive boosting (ABoost), multilayer perceptron (MLP), and random forest (RF), as well as for the DL algorithms CNN and LSTM trained with the for the four MRI data sets (clinical MRI, MRI-based oxygen metabolism, MRI-based vascular architecture mapping (VAM), and the combination of oxygen metabolism and VAM). The color codes are listed on the far right.

Representative cases of the independent internal patient testing cohort for patients with an *IDH*-mutated glioma and an *IDH*-wildtype glioma that were correctly classified by all algorithms are shown in Figures 2 and 3, respectively. Figure 4, on the other hand, shows the case of a patient who was misclassified by all algorithms due to strong motion artifacts in the perfusion and relaxometry MRI sequences.



Figure 2. Representative case of the independent internal patient cohort shows a 54-year-old male patient who suffered from an *IDH*-mutated oligodendroglioma WHO grade 2 that was correctly classified by all algorithms. Clinical MRI data included contrast-enhanced (CE) T1w MRI, FLAIR MRI, and the quantitative maps of the apparent diffusion coefficient (ADC) and cerebral blood volume (CBV). MRI biomarker maps for oxygen metabolism included the oxygen extraction fraction (OEF), cerebral metabolic rate of oxygen (CMRO₂), capillary oxygen tension (capiPO₂), and mitochondrial oxygen tension (mitoPO₂), respectively. MRI biomarker maps for microvascular architecture and neovascularization activity included the microvascular perfusion (μ CBV), microvascular density (MVD), microvascular diameter (aka vessel size index, VSI), and microvascular type (MTI). The color codes are presented on the right-hand part of the images.



Figure 3. Representative case of the independent internal patient cohort shows a 60-year-old female patient who suffered from an *IDH*-wildtype astrocytoma WHO grade 3 that was correctly classified by all algorithms. Clinical MRI data included contrast-enhanced (CE) T1w MRI, FLAIR MRI, and the quantitative maps of the apparent diffusion coefficient (ADC) and cerebral blood volume (CBV). MRI biomarker maps for oxygen metabolism included the oxygen extraction fraction (OEF), cerebral metabolic rate of oxygen (CMRO₂), capillary oxygen tension (capiPO₂), and mitochondrial oxygen tension (mitoPO₂), respectively. MRI biomarker maps for microvascular architecture and neovascularization activity included the microvascular perfusion (μ CBV), microvascular density (MVD), microvascular diameter (aka vessel size index, VSI), and microvascular type (MTI). The color codes are presented on the right-hand part of the images.



Figure 4. Case of a 36-year-old female patient who suffered from an *IDH*-wildtype glioblastoma WHO grade 4 that was misclassified by all algorithms. The reason for this was significant movement artifacts in the perfusion and relaxometry data. Clinical MRI data included contrast-enhanced (CE) T1w MRI, FLAIR MRI, and the quantitative maps of the apparent diffusion coefficient (ADC) and cerebral blood volume (CBV). MRI biomarker maps for oxygen metabolism included the oxygen extraction fraction (OEF), cerebral metabolic rate of oxygen (CMRO₂), capillary oxygen tension (capiPO₂), and mitochondrial oxygen tension (mitoPO₂), respectively. MRI biomarker maps for microvascular architecture and neovascularization activity included the microvascular perfusion (μ CBV), microvascular density (MVD), microvascular diameter (aka vessel size index, VSI), and microvascular type (MTI). The color codes are presented on the right-hand part of the images.

3.3. Testing with an Independent External Cohort

The tests of the algorithms with the independent external patient cohort from the FAU Erlangen-Nürnberg produced completely opposite results compared to the tests with the independent internal cohort. Algorithms trained with the cMRI data from the

University Clinic St. Pölten showed the best classification quality when tested with the independent external cMRI data from the FAU Erlangen-Nürnberg. ABoost performed best with six errors out of 33 patients, a precision of 0.849, and an accuracy of 0.818. All other algorithms trained with cMRI data followed with seven to nine errors in the middle range. Only the MLP algorithm, which was trained with the oxygen metabolism and VAM data combination, and the ABoost model, which was trained with oxygen metabolism, were also in the middle range of classification quality, with seven and nine errors. All other algorithms that were trained with physio-metabolic data from St. Pölten showed poor classification quality, with 10 to 13 errors in the 33 patients when tested with the independent external data from FAU Erlangen-Nürnberg.

The values for the quality parameters of the five *IDH* gene status classifiers and the four MRI data sets when tested with the independent internal test cohort are presented as heat maps in Figure 5. Confusion matrices are provided in Supplementary Figure S3.

	Clinical	MRI			Oxygen Metabolism						Vascula	r Archit	ecture M	lapping		Oxy Met & VAM						
	ABoost	MLP	RF	CNN	LSTM	ABoost	MLP	RF	CNN	LSTM	ABoost	MLP	RF	CNN	LSTM	ABoost	MLP	RF	CNN	LSTM		
Accuracy	0.818	0.758	0.788	0.727	0.758	0.727	0.606	0.667	0.606	0.697	0.636	0.727	0.727	0.636	0.697	0.606	0.788	0.727	0.606	0.727).88
Sensitivity	0.818	0.758	0.788	0.727	0.758	0.727	0.606	0.667	0.606	0.697	0.636	0.727	0.727	0.636	0.697	0.606	0.788	0.727	0.606	0.727		
Specificity	0.86	0.683	0.736	0.701	0.719	0.666	0.525	0.56	0.525	0.577	0.506	0.63	0.63	0.614	0.613	0.489	0.7	0.523	0.561	0.594		
Precision	0.849	0.753	0.785	0.733	0.758	0.723	0.599	0.653	0.599	0.685	0.615	0.72	0.72	0.651	0.689	0.586	0.785	0.733	0.613	0.725		
F score	0.822	0.752	0.786	0.729	0.758	0.725	0.602	0.654	0.602	0.680	0.616	0.717	0.717	0.641	0.690	0.591	0.786	0.729	0.609	0.706		
AUROC	0.879	0.77	0.863	0.714	0.738	0.661	0.607	0.641	0.607	0.637	0.599	0.694	0.694	0.625	0.655	0.567	0.863	0.714	0.583	0.661).52
Errors	6	8	7	9	8	9	13	11	13	10	12	9	9	12	10	13	7	9	13	9		6 13

Figure 5. Overview of the results for the testing with the independent external patient cohort. Heat maps depicting (top down) accuracy, sensitivity, specificity, precision, F-score, AUROC, and classification errors for the traditional ML algorithms adaptive boosting (ABoost), multilayer perceptron (MLP), and random forest (RF), as well as for the DL algorithms CNN and LSTM trained with the for the four MRI data sets (clinical MRI, MRI-based oxygen metabolism, MRI-based vascular architecture mapping (VAM), and the combination of oxygen metabolism and VAM). The color codes are listed on the far right.

Cases of the independent external patient cohort that were misclassified by all algorithms are depicted in Figures 6 and 7, respectively. The reasons for these misclassifications were MRI signal changes that were atypical for an *IDH*-wildtype glioblastoma (Figure 6) and susceptibility artifacts due to bleeding in the *IDH*-mutated oligodendroglioma WHO grade 3 (Figure 7), respectively.



Figure 6. Case of the independent external patient cohort: A 45-year-old male patient who suffered from an *IDH*-wildtype glioblastoma WHO grade 4 that was misclassified by all algorithms. The reason for this was signal changes that are atypical for a glioblastoma, such as no contrast enhancement in

large parts of the tumor, low hyperperfusion and neovascularization, and high oxygen tension. Clinical MRI data included contrast-enhanced (CE) T1w MRI, FLAIR MRI, and the quantitative maps of the apparent diffusion coefficient (ADC) and cerebral blood volume (CBV). MRI biomarker maps for oxygen metabolism included the oxygen extraction fraction (OEF), cerebral metabolic rate of oxygen (CMRO₂), capillary oxygen tension (capiPO₂), and mitochondrial oxygen tension (mitoPO₂), respectively. MRI biomarker maps for microvascular architecture and neovascularization activity included the microvascular perfusion (μ CBV), microvascular density (MVD), microvascular diameter (aka vessel size index, VSI), and microvascular type (MTI). The color codes are presented on the right-hand part of the images.



Figure 7. Case of the independent external patient cohort: A 67-year-old male patient who suffered from an *IDH*-mutated oligodendroglioma WHO grade 3 that was misclassified by all algorithms. The reason for this was susceptibility artifacts due to bleeding in the tumor. Clinical MRI data included contrast-enhanced (CE) T1w MRI, FLAIR MRI, and the quantitative maps of the apparent diffusion coefficient (ADC) and cerebral blood volume (CBV). MRI biomarker maps for oxygen metabolism included the oxygen extraction fraction (OEF), cerebral metabolic rate of oxygen (CMRO₂), capillary oxygen tension (capiPO₂), and mitochondrial oxygen tension (mitoPO₂), respectively. MRI biomarker maps for microvascular architecture and neovascularization activity included the microvascular perfusion (μ CBV), microvascular density (MVD), microvascular diameter (aka vessel size index, VSI), and microvascular type (MTI). The color codes are presented on the right-hand part of the images.

4. Discussion

Early and reliable characterization of the *IDH* gene mutation status of glioma is crucial for personalized treatment decisions and prognosis in clinical neurooncology. The gold standard, histopathological analysis of invasively obtained tissue samples, requires additional time and financial resources. A reliable preoperative, non-invasive determination of *IDH* gene status would therefore be of great benefit.

In this study, we trained traditional ML algorithms and simple DL models with radiomic features from both physio-metabolic MRI data and routine MRI data based on differences in energy metabolism between *IDH*mut and *IDH*wt gliomas. The algorithms trained with physio-metabolic data showed the best classification performance when tested with an independent internal cohort, with precision, accuracy, and AUROC up to 91.7%, 87.5%, and 0.979, respectively. The LSTM model in combination with oxygen metabolism and the MLP in combination with oxygen metabolism along with microvascular architecture demonstrated the highest performance parameters, which in part was in line with our hypothesis: combining physio-metabolic MRI and deep learning algorithms enables a noninvasive and reliable classification of the *IDH* gene mutation status of gliomas.

In contrast, CNN showed poor classification performance for *IDH* status in combination with oxygen metabolism, which can be explained by the well-known limitations of CNNs in learning dependencies between distant data points, which appear to be particularly relevant to the radiomic features of the biomarkers for oxygen metabolism, i.e., OEF, CMRO₂, capiPO₂, and mitoPO₂. However, when combined with the microvascular architecture features, CNN performed very well, indicating greater redundancy of the biomarkers (μ CBV, MVD, MTI, VSI). Training the DL algorithms (CNN and LSTM) with the combination of oxygen metabolism plus microvascular architecture resulted in a significant deterioration/reduction in diagnostic performance, which in turn may be caused by the extended feature vector (856 features) for the relatively simple DL models, while MLP and RF performed very well when trained on the combined OxyMet plus VAM data.

To our surprise, the results for independent external testing were clear, with the models trained with the cMRI data from UK St. Pölten consistently yielding the best results for *IDH* classification for the external test cohort from FAU Erlangen with precision, accuracy, and AUROC of up to 84.9%, 81.8%, and 0.879. The poor performance of the models (ML and DL) trained with the physio-metabolic data are related to relevant differences in the innovative physiologic and metabolic MRI techniques for data acquisition between the study sites.

These results also reflect a limitation in applying AI models in combination with experimental imaging data for classification. There is often no uniform standard for experimental imaging methods. Rather, each institution often follows its own technical approach, sometimes with major differences in MR technology, protocols, and post-processing. This means a great challenge for multicenter studies and building up large public databases.

A growing number of studies have been published using traditional ML or DL methods combined with imaging data in order to classify glioma *IDH* gene status and can essentially be divided into two groups: studies that used innovative imaging techniques in combination with traditional ML or simple DL algorithms and studies that used exclusively anatomic MRI data from large, publicly available, multicenter databases in combination with advanced DL models. Our study clearly belongs to the former group. The previously published papers in this group, however, have not independently tested their algorithms on data unseen by the model.

There exists a broad variety of experimental neuroimaging methods, with only two studies using metabolic imaging methods for *IDH* classification.

Ozturk-Isik et al. [64] used short-echo time proton MR spectroscopy (1H-MRS) to examine metabolite levels in 112 glioma patients. They used this metabolic information in combination with traditional ML algorithms for noninvasive preoperative classification of *IDH* mutation subgroups. Using the logistic regression algorithm, they were able to predict the presence of an *IDH* mutation with an accuracy of 88.4%. Tatekawa et al. [65] developed a method for voxel-wise clustering of multiparametric MRI and 18F-dihydroxyphenylalanine (FDOPA) positron emission tomography (PET) images using an unsupervised 16-class clustering approach. Combined with a support vector machine algorithm, they classified the *IDH*-gene mutation status of 62 glioma patients with a classification accuracy of 76% or an AUROC of 0.81, respectively. However, due to the small number of patients in these two studies, the authors did not carry out tests with an independent cohort, but only used leave-one-out cross-validation.

Kesler et al. [66] followed an interesting approach by determining gray matter connectomes from preoperative non-contrast T1w MRI scans and demonstrated that gray matter connectomes are useful for distinguishing between *IDH* variants of high-grade glioma. In a subsequent study [67], the authors extracted 93 connectome features from the preoperative T1w MRIs of 234 glioma patients and evaluated the performance of four traditional ML algorithms for predicting the *IDH* genotype. Using features for the connectome efficiency, brain volume, network degree, network size, tumor hemisphere, and tumor lobe in combination with an RF model, the authors achieved an accuracy and AUROC of 89% and 0.95, respectively. Based on the assumption that *IDH*-mutant gliomas have lower cellular density and increased interstitial edema, Karami et al. [68] evaluated the potential of multi-shell diffusion MRI in combination with transfer learning of a pretrained DL model for *IDH* gene status classification. The multi-shell diffusion MRI approach additionally uses significantly higher b-values, which enables more informative diffusion models. By adding these experimental data to cMRI data, the authors achieved an increase in accuracy from 74% to 81% for prediction of the *IDH* gene status in 146 glioma patients. However, in these two studies, no tests were carried out on independent cohorts, but rather leave-one-out [67] and five-fold cross-validation [68] were used, respectively. In our study, accuracy values for training validation ranged from 89.9% to 94.0%, which is significantly higher than for most previous studies using experimental imaging techniques. At this point, it is important to emphasize that independent testing of the algorithms may present an essential quality criterion for such studies, and we strongly recommend its use.

To our knowledge, the only study that has independently tested DL models using experimental MRI methods to classify *IDH* gene status is that of Choi et al. [69], who performed DSC perfusion MRI to obtain intensity-time curves of the T_2 *-susceptibility signal from tumor subregions segmented by a CNN. These T_2 *-time curves and the corresponding arterial input functions from 395 patients were used as multidimensional input to train a convolutional LSTM model with an attention mechanism. They were able to predict the *IDH* genotype in an independent test cohort of 50 patients with 91.7% accuracy. We also achieved this accuracy value with the best-performing models in our study, LSTM with oxygen metabolism and MLP with oxygen metabolism and VAM, respectively, but were only able to include half as many patients in our study. Data from more patients and the use of more up-to-date algorithms could further improve the performance of our approach. However, in the study by Choi et al. [69], non-independent external testing was carried out, which should be the standard for future studies on clinical AI applications.

Independent external testing has been carried out in most of the studies of the abovementioned second group, i.e., studies that have performed the classification of the *IDH* gene status using anatomical MRI data from large publicly available databases in combination with more advanced DL methods. These studies usually used several of the following public databases: The Cancer Imaging Archive (TCIA) [70], the Repository of Molecular Brain Neoplasia Data (REMBRANDT) [71], the Clinical Proteomic Tumor Analysis Consortium Glioblastoma Multiforme (CPTAC-GBM) collection [72], the Ivy Glioblastoma Atlas Project (Ivy GAP) collection [73], the Brain Tumor Segmentation (BraTS) challenge dataset [74], and the Erasmus Glioma Database [75]. In addition, databases from the participating universities were often used for training and/or independent testing.

The three studies with the best performance in *IDH* classification used data from over 1000 patients in combination with advanced CNNs. Chakrabarty et al. [76] used a 2.5D hybrid CNN to simultaneously locate the glioma and classify its *IDH* status. By leveraging MRI features and knowledge features from clinical records and tumor location, the authors achieved an accuracy and AUROC of 93.5% and 0.925 for independent internal testing and 94.1% and 0.933 for independent external testing, respectively. Choi et al. [77] combined a 3D U-shaped CNN for tumor segmentation with a CNN classifier (34-layer ResNet) and achieved with this hybrid model an accuracy and AUROC of 93.8% and 0.96 for the internal test, as well as 87.9% and 0.94 for the external test. Finally, Bangalore Yogananda et al. [78] used a so-called nnU-Net architecture, which is characterized by excellent generalizability and reduced risk of overfitting by transferring information from each preceding layer to the subsequent ones. Additionally, the nnU-Net automatically configures itself, including pre-processing, network architecture, training, and post-processing [79]. In independent internal testing, they achieved an overall accuracy of 92.8% with an AUROC of 0.96. Other studies [80–82], which also used multi-site data from public databases but with fewer patients or which only used a simple CNN, achieved significantly lower accuracy values (84.0-85.7%).

The accuracy of the three studies [76–78] with the top performance in the independent internal testing (92.8–93.8%) was significantly higher than our best result (87.5%); the

AUROC values (0.925–0.96), however, were lower than our result (0.979). Our independent internal testing cohort was quite small, at only 16 patients, which is a limitation of our study. However, during the prespecified four-month period for enrolling patients into the testing cohort, only these 16 patients with untreated gliomas were examined.

The results of the independent external tests clearly demonstrate the advantages of the approach of training more complex DL models with multi-site data from many patients from public databases. Chakrabaty et al. [76] and Choi et al. [77] achieved accuracies of 94.1% and 87.9% and AUROC values of 0.933 and 0.94, respectively, in their independent external tests. In our study, the traditional ML models trained with cMRI data achieved the best classification results, with an accuracy of max. 81.8% and AUROC of max. 0.879, respectively. The models trained with the physio-metabolic MRI data, especially the DL algorithms, performed far worse in our independent external testing. In addition to the insufficient generalization of the models due to training with single-site data, the differences in the data acquisition techniques of the physio-metabolic MRI data between the two sites led to a further increase in overfitting.

These case studies demonstrated that motion and susceptibility artifacts were also a reason for some misclassifications, as presented in our physio-metabolic MRI data. Susceptibility artifacts caused by intratumoral bleeding (Figure 7), for example, cannot be corrected. Nalawade et al. [83], however, demonstrated that classifier performance molecular glioma markers including *IDH* status decreased markedly with increasing motion corruption of the MRI data, but applying motion correction effectively restored classification accuracy for even the most motion-corrupted images.

There are some biasing concerns in our present study worthy of mention. Calculation of ADC values was performed by linear approximation with two-point b-values $(b = 0 \text{ and } 1000 \text{ s/mm}^2)$ and did not take into account higher b-values or the calculation of more quantitative biomarkers for diffusion. However, this procedure is still widely used in clinical routine diagnostics. As noted above, the patient cohorts for both training/validation (166 patients) and independent internal testing (16 patients) were quite small, particularly compared to studies using large-scale public databases. Additionally, all data were collected using a single MRI scanner at a single site, leading to insufficient generalization of the AI algorithms due to differences in MR scanner setups from different manufacturers, different magnetic field strengths, and variations in examination protocols that were not considered. The differences in data acquisition techniques for quantitative physio-metabolic MRI data between the two sites involved had a greater impact on classification performance than expected. The independent external testing clearly demonstrated the limitations of using AI models in combination with innovative imaging methods developed and applied at a single site. To prevent these pronounced effects of model overfitting, the construction of public databases for innovative imaging data from different sites with different settings is necessary.

Two recently published multicenter studies have shown how to prevent model overfitting. Prof. Marias' group used DSC-MRI perfusion data from four MRI scanners at three different sites, from three different manufacturers, and operating at two different field strengths to predict the *IDH* gene mutation of gliomas. Data from one site were used for training and data from the other two sites were used for independent external testing. In the study by Manikis et al. [84], the authors were able to increase the accuracy of *IDH* status classification from 54.4% to 70.6% using dynamics-based standardization of images. In the study by Ioannidis et al. [85] using mathematical models to quantify the DSC data, the accuracy of *IDH* mutation prediction was increased to 75%.

Another possible solution to prevent the limited generalization of the applied models is to use self-supervised learning-based foundation models, which represent large AI models trained on a vast quantity of unlabeled data at scale, resulting in models that can be adapted to a wide range of downstream tasks [86]. The self-supervised learning approach reduces data inefficiency by deriving supervisory signals directly from data rather than relying on expert knowledge using labels [87,88]. Self-supervised learning uses huge amounts of unlabeled data to learn general-purpose feature representations that can be easily adapted easily to more specific tasks. After this pretraining phase, the models are fine-tuned to specific downstream tasks such as segmentation, classification, or prediction. Recently, Zhou et al. [89] presented a foundation model for retinal images using trained selfsupervised learning on 1.6 million unlabeled retinal images. They adapted this foundation model to diagnosis and prognosis of sight-threatening eye diseases as well as incident prediction of complex systemic disorders such as heart failure and myocardial infarction, sometimes with just a few hundred labeled data. They demonstrated that the adapted foundation model provides a generalizable solution to improve model performance and consistently outperforms several traditional benchmark models. Creating a baseline model for brain imaging is time- and labor-intensive and requires powerful hardware resources and huge amounts of data. However, once generated and made publicly available, it could be adapted for a variety of sub-applications such as neurooncology, neurodegeneration, or stroke. Finally, we used the radiomics approach for feature extraction, which is more transparent and easier to understand compared to deep learning approaches. However, all steps must be carried out manually and hence may have unfavorable time/labor profiles. Therefore, the next step is to implement deep neural networks for brain tumor segmentation and feature extraction. In the next step, the application of innovative AI frameworks [90,91] would then be conceivable to uncover deep interrelationships between gene expression, neuropathology, and clinical data.

5. Conclusions

In summary, innovative physio-metabolic imaging methods provide valuable functional and metabolic insights into tumor pathophysiology and enable a reliable preoperative classification of the *IDH* gene status of gliomas even with a smaller amount of data. However, this advantage is partly lost in multisite clinical AI applications compared to anatomical or clinically advanced MRI due to varying data acquisition techniques. The large amounts of anatomical MRI data in public databases permit exploration of more complex AI algorithms more suitable for this purpose. A two-step protocol capable of counterbalancing these disadvantages and variabilities in experimental imaging techniques could be a foundation model trained in advance through self-supervised learning for brain imaging in general and then adapted for specific tasks using much smaller amounts of data.

Supplementary Materials: The following supporting information can be downloaded at: https:// www.mdpi.com/article/10.3390/cancers16061102/s1, Supplementary Table S1: Patient characteristics of the cohorts. Supplementary Table S2: Sequence parameters of the MRI study protocols at the University Clinic St. Pölten and the Friedrich-Alexander University Erlangen-Nürnberg. Supplementary Table S3: Value ranges, bin sizes, and bin numbers for discretization of biomarker maps. Supplementary Table S4: List of radiomic features. Supplementary Table S5: List of optimized hyperparameters. Supplementary Table S6: Parameters of one-dimensional convolutional neural network (1D-CNN). Supplementary Figure S1: The overall study pipeline. Supplementary Figure S2: Calculation of the confusion matrix-derived metrics. Supplementary Figure S3: Confusion matrices for the independent internal and the independent external testing. References [49–53,92–96] are cited in Supplementary Materials.

Author Contributions: All authors have contributed to this study in a manner to justify authorship according to the ICMJE criteria for authorship. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted. Conceptualization, A.S., S.O., A.M.-B. and A.D.; methodology, A.S.; software, A.S.; formal analysis, A.S., K.N., F.M. and A.D.; investigation, A.S., K.N., S.O., F.M. and A.D.; validation, A.S., K.N., S.O., F.M. and A.D.; resources, A.S., S.O., F.M. and A.D.; data curation, A.S., K.N., S.O., F.M. and A.D.; writing—original draft preparation, A.S., T.M.K. and A.M.-B.; writing—review and editing, A.S., K.N., S.O., F.M., T.M.K., A.M.-B., D.A.B., O.S. and A.D.; visualization, A.S., S.O., F.M. and A.D.; authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the German Research Foundation (Deutsche Forschungsgemeinschaft—DFG; Grant Numbers STA 1331/3-1 and DO 721/9-1). This work was supported by Forschungsimpulse [project ID: SF_0045], a program of Karl Landsteiner University of Health Sciences funded by the Federal Government of Lower Austria. The authors want to appreciate the contribution of NÖ Landesgesundheitsagentur, legal entity of University Hospitals in Lower Austria, for providing the organizational framework to conduct this research. The authors also would like to acknowledge support by Open Access Publishing Fund of Karl Landsteiner University of Health Sciences, Krems, Austria.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Ethics Committee of the Lower Austrian Provincial Government (protocol code GS1-EK-4/339-2015) as well as of the Friedrich-Alexander University (FAU) Erlangen-Nürnberg (Nr. 3578 and 177_14B).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data available on request due to privacy and ethical restrictions.

Conflicts of Interest: The authors declared no conflicts of interest.

References

- Molinaro, A.M.; Taylor, J.W.; Wiencke, J.K.; Wrensch, M.R. Genetic and molecular epidemiology of adult diffuse glioma. *Nat. Rev. Neurol.* 2019, 15, 405–417. [CrossRef]
- 2. Vigneswaran, K.; Neill, S.; Hadjipanayis, C.G. Beyond the World Health Organization grading of infiltrating gliomas: Advances in the molecular genetics of glioma classification. *Ann. Transl. Med.* **2015**, *3*, 95. [CrossRef]
- Louis, D.N.; Perry, A.; Wesseling, P.; Brat, D.J.; Cree, I.A.; Figarella-Branger, D.; Hawkins, C.; Ng, H.K.; Pfister, S.M.; Reifenberger, G.; et al. The 2021 WHO Classification of Tumors of the Central Nervous System: A summary. *Neuro Oncol.* 2021, 23, 1231–1251. [CrossRef] [PubMed]
- Parsons, D.W.; Jones, S.; Zhang, X.; Lin, J.C.-H.; Leary, R.J.; Angenendt, P.; Mankoo, P.; Carter, H.; Siu, I.-M.; Gallia, G.L.; et al. An integrated genomic analysis of human glioblastoma multiforme. *Science* 2008, 321, 1807–1812. [CrossRef] [PubMed]
- 5. Yan, H.; Parsons, D.W.; Jin, G.; McLendon, R.; Rasheed, B.A.; Yuan, W.; Kos, I.; Batinic-Haberle, I.; Jones, S.; Riggins, G.J.; et al. *IDH*1 and *IDH*2 mutations in gliomas. *N. Engl. J. Med.* **2009**, *360*, 765–773. [CrossRef] [PubMed]
- Chen, J.-R.; Yao, Y.; Xu, H.-Z.; Qin, Z.-Y. Isocitrate Dehydrogenase (*IDH*)1/2 Mutations as Prognostic Markers in Patients with Glioblastomas. *Medicine* 2016, 95, e2583. [CrossRef]
- Dang, L.; White, D.W.; Gross, S.; Bennett, B.D.; Bittinger, M.A.; Driggers, E.M.; Fantin, V.R.; Jang, H.G.; Jin, S.; Keenan, M.C.; et al. Cancer-associated *IDH*1 mutations produce 2-hydroxyglutarate. *Nature* 2009, 462, 739–744. [CrossRef] [PubMed]
- 8. Tateishi, K.; Wakimoto, H.; Iafrate, A.J.; Tanaka, S.; Loebel, F.; Lelic, N.; Wiederschain, D.; Bedel, O.; Deng, G.; Zhang, B.; et al. Extreme Vulnerability of *IDH1* Mutant Cancers to NAD+ Depletion. *Cancer Cell* **2015**, *28*, 773–784. [CrossRef] [PubMed]
- Reitman, Z.J.; Duncan, C.G.; Poteet, E.; Winters, A.; Yan, L.-J.; Gooden, D.M.; Spasojevic, I.; Boros, L.G.; Yang, S.-H.; Yan, H. Cancer-associated isocitrate dehydrogenase 1 (*IDH*1) R132H mutation and d-2-hydroxyglutarate stimulate glutamine metabolism under hypoxia. *J. Biol. Chem.* 2014, 289, 23318–23328. [CrossRef]
- Grassian, A.R.; Parker, S.J.; Davidson, S.M.; Divakaruni, A.S.; Green, C.R.; Zhang, X.; Slocum, K.L.; Pu, M.; Lin, F.; Vickers, C.; et al. *IDH*1 mutations alter citric acid cycle metabolism and increase dependence on oxidative mitochondrial metabolism. *Cancer Res.* 2014, 74, 3317–3331. [CrossRef]
- Yao, J.; Chakhoyan, A.; Nathanson, D.A.; Yong, W.H.; Salamon, N.; Raymond, C.; Mareninov, S.; Lai, A.; Nghiemphu, P.L.; Prins, R.M.; et al. Metabolic characterization of human *IDH* mutant and wild type gliomas using simultaneous pH- and oxygen-sensitive molecular MRI. *Neuro. Oncol.* 2019, 21, 1184–1196. [CrossRef]
- Wu, B.; Warnock, G.; Zaiss, M.; Lin, C.; Chen, M.; Zhou, Z.; Mu, L.; Nanz, D.; Tuura, R.; Delso, G. An overview of CEST MRI for non-MR physicists. *EJNMMI Phys.* 2016, 3, 19. [CrossRef]
- 13. Hardee, M.E.; Zagzag, D. Mechanisms of Glioma-Associated Neovascularization. Am. J. Pathol. 2012, 181, 1126–1141. [CrossRef]
- 14. Stadlbauer, A.; Zimmermann, M.; Bennani-Baiti, B.; Helbich, T.H.; Baltzer, P.; Clauser, P.; Kapetas, P.; Bago-Horvath, Z.; Pinker, K. Development of a Non-invasive Assessment of Hypoxia and Neovascularization with Magnetic Resonance Imaging in Benign and Malignant Breast Tumors: Initial Results. *Mol. Imaging Biol.* **2019**, *21*, 758–770. [CrossRef]
- Stadlbauer, A.; Zimmermann, M.; Doerfler, A.; Oberndorfer, S.; Buchfelder, M.; Coras, R.; Kitzwögerer, M.; Roessler, K. Intratumoral heterogeneity of oxygen metabolism and neovascularization uncovers 2 survival-relevant subgroups of *IDH*1 wild-type glioblastoma. *Neuro Oncol.* 2018, 20, 1536–1546. [CrossRef] [PubMed]
- Stadlbauer, A.; Kinfe, T.M.; Eyüpoglu, I.; Zimmermann, M.; Kitzwögerer, M.; Podar, K.; Buchfelder, M.; Heinz, G.; Oberndorfer, S.; Marhold, F. Tissue Hypoxia and Alterations in Microvascular Architecture Predict Glioblastoma Recurrence in Humans. *Clin. Cancer Res.* 2021, 27, 1641–1649. [CrossRef]
- Christen, T.; Schmiedeskamp, H.; Straka, M.; Bammer, R.; Zaharchuk, G. Measuring brain oxygenation in humans using a multiparametric quantitative blood oxygenation level dependent MRI approach. *Magn. Reson. Med.* 2012, 68, 905–911. [CrossRef]

- Stadlbauer, A.; Zimmermann, M.; Heinz, G.; Oberndorfer, S.; Doerfler, A.; Buchfelder, M.; Rössler, K. Magnetic resonance imaging biomarkers for clinical routine assessment of microvascular architecture in glioma. *J. Cereb. Blood Flow Metab.* 2017, 37, 632–643. [CrossRef]
- 19. Boxerman, J.L.; Hamberg, L.M.; Rosen, B.R.; Weisskoff, R.M. Mr contrast due to intravascular magnetic susceptibility perturbations. *Magn. Reson. Med.* **1995**, *34*, 555–566. [CrossRef] [PubMed]
- Lambin, P.; Leijenaar, R.T.H.; Deist, T.M.; Peerlings, J.; de Jong, E.E.C.; van Timmeren, J.; Sanduleanu, S.; Larue, R.T.H.M.; Even, A.J.G.; Jochems, A.; et al. Radiomics: The bridge between medical imaging and personalized medicine. *Nat. Rev. Clin. Oncol.* 2017, 14, 749–762. [CrossRef] [PubMed]
- Gillies, R.J.; Kinahan, P.E.; Hricak, H. Radiomics: Images Are More than Pictures, They Are Data. Radiology 2016, 278, 563–577. [CrossRef]
- Aerts, H.J.W.L.; Velazquez, E.R.; Leijenaar, R.T.H.; Parmar, C.; Grossmann, P.; Carvalho, S.; Bussink, J.; Monshouwer, R.; Haibe-Kains, B.; Rietveld, D.; et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat. Commun.* 2014, *5*, 4006. [CrossRef] [PubMed]
- 23. SchmIDHuber, J. Deep learning in neural networks: An overview. Neural Netw. 2015, 61, 85–117. [CrossRef] [PubMed]
- 24. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. Nature 2015, 521, 436–444. [CrossRef]
- 25. Radenovic, F.; Tolias, G.; Chum, O. Fine-Tuning CNN Image Retrieval with No Human Annotation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *41*, 1655–1668. [CrossRef]
- 26. Kong, W.; Dong, Z.Y.; Jia, Y.; Hill, D.J.; Xu, Y.; Zhang, Y. Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network. *IEEE Trans. Smart Grid* 2019, *10*, 841–851. [CrossRef]
- 27. Choi, E.; Schuetz, A.; Stewart, W.F.; Sun, J. Using recurrent neural network models for early detection of heart failure onset. J. Am. Med. Inform. Assoc. 2017, 24, 361–370. [CrossRef]
- Jiang, Y.; Zhao, K.; Xia, K.; Xue, J.; Zhou, L.; Ding, Y.; Qian, P. A Novel Distributed Multitask Fuzzy Clustering Algorithm for Automatic MR Brain Image Segmentation. J. Med. Syst. 2019, 43, 118. [CrossRef] [PubMed]
- Hanson, J.; Yang, Y.; Paliwal, K.; Zhou, Y. Improving protein disorder prediction by deep bidirectional long short-term memory recurrent neural networks. *Bioinformatics* 2017, 33, 685–692. [CrossRef]
- Jing, X.; Dorrius, M.D.; Wielema, M.; Sijens, P.E.; Oudkerk, M.; van Ooijen, P. Breast Tumor Identification in Ultrafast MRI Using Temporal and Spatial Information. *Cancers* 2022, 14, 2042. [CrossRef]
- Antropova, N.; Huynh, B.; Li, H.; Giger, M.L. Breast lesion classification based on dynamic contrast-enhanced magnetic resonance images sequences with long short-term memory networks. J. Med. Imaging 2019, 6, 011002. [CrossRef]
- 32. Zou, J.; Balter, J.M.; Cao, Y. Estimation of pharmacokinetic parameters from DCE-MRI by extracting long and short time-dependent features using an LSTM network. *Med. Phys.* 2020, 47, 3447–3457. [CrossRef]
- Boxerman, J.L.; Schmainda, K.M.; Weisskoff, R.M. Relative cerebral blood volume maps corrected for contrast agent extravasation significantly correlate with glioma tumor grade, whereas uncorrected maps do not. *AJNR Am. J. Neuroradiol.* 2006, 27, 859–867. [PubMed]
- Boxerman, J.L.; Prah, D.E.; Paulson, E.S.; Machan, J.T.; Bedekar, D.; Schmainda, K.M. The Role of Preload and Leakage Correction in Gadolinium-Based Cerebral Blood Volume Estimation Determined by Comparison with MION as a Criterion Standard. *Am. J. Neuroradiol.* 2012, 33, 1081–1087. [CrossRef]
- 35. Smith, A.M.; Grandin, C.B.; Duprez, T.; Mataigne, F.; Cosnard, G. Whole brain quantitative CBF, CBV, and MTT measurements using MRI bolus tracking: Implementation and application to data acquired from hyperacute stroke patients. *J. Magn. Reson. Imaging* **2000**, *12*, 400–410. [CrossRef] [PubMed]
- Bjørnerud, A.; Emblem, K.E. A Fully Automated Method for Quantitative Cerebral Hemodynamic Analysis Using DSC–MRI. J. Cereb. Blood Flow Metab. 2010, 30, 1066–1078. [CrossRef] [PubMed]
- Preibisch, C.; Volz, S.; Anti, S.; Deichmann, R. Exponential excitation pulses for improved water content mapping in the presence of background gradients. *Magn. Reson. Med.* 2008, 60, 908–916. [CrossRef]
- Prasloski, T.; M\u00e4dler, B.; Xiang, Q.-S.; MacKay, A.; Jones, C. Applications of stimulated echo correction to multicomponent T 2 analysis. *Magn. Reson. Med.* 2012, 67, 1803–1814. [CrossRef]
- Stadlbauer, A.; Zimmermann, M.; Kitzwögerer, M.; Oberndorfer, S.; Rössler, K.; Dörfler, A.; Buchfelder, M.; Heinz, G. MR Imaging–derived Oxygen Metabolism and Neovascularization Characterization for Grading and *IDH* Gene Mutation Detection of Gliomas. *Radiology* 2017, 283, 799–809. [CrossRef]
- 40. Kennan, R.P.; Zhong, J.; Gore, J.C. Intravascular susceptibility contrast mechanisms in tissues. *Magn. Reson. Med.* **1994**, *31*, 9–21. [CrossRef]
- 41. Vafaee, M.S.; Vang, K.; Bergersen, L.H.; Gjedde, A. Oxygen Consumption and Blood Flow Coupling in Human Motor Cortex during Intense Finger Tapping: Implication for a Role of Lactate. *J. Cereb. Blood Flow Metab.* **2012**, *32*, 1859–1868. [CrossRef]
- 42. Gjedde, A. Cerebral Blood Flow Change in Arterial Hypoxemia Is Consistent with Negligible Oxygen Tension in Brain Mitochondria. *Neuroimage* **2002**, *17*, 1876–1881. [CrossRef]
- 43. Vafaee, M.S.; Gjedde, A. Model of Blood–Brain Transfer of Oxygen Explains Nonlinear Flow-Metabolism Coupling During Stimulation of Visual Cortex. *J. Cereb. Blood Flow Metab.* 2000, 20, 747–754. [CrossRef] [PubMed]

- 44. Ducreux, D.; Buvat, I.; Meder, J.F.; Mikulis, D.; Crawley, A.; Fredy, D.; TerBrugge, K.; Lasjaunias, P.; Bittoun, J. Perfusion-weighted MR imaging studies in brain hypervascular diseases: Comparison of arterial input function extractions for perfusion measurement. *AJNR Am. J. Neuroradiol.* **2006**, *27*, 1059–1069.
- Xu, C.; Kiselev, V.G.; Möller, H.E.; Fiebach, J.B. Dynamic hysteresis between gradient echo and spin echo attenuations in dynamic susceptibility contrast imaging. *Magn. Reson. Med.* 2013, 69, 981–991. [CrossRef] [PubMed]
- 46. Jensen, J.H.; Lu, H.; Inglese, M. Microvessel density estimation in the human brain by means of dynamic contrast-enhanced echo-planar imaging. *Magn. Reson. Med.* **2006**, *56*, 1145–1150. [CrossRef]
- Emblem, K.E.; Mouridsen, K.; Bjornerud, A.; Farrar, C.T.; Jennings, D.; Borra, R.J.H.; Wen, P.Y.; Ivy, P.; Batchelor, T.T.; Rosen, B.R.; et al. Vessel architectural imaging identifies cancer patient responders to anti-angiogenic therapy. *Nat. Med.* 2013, *19*, 1178–1183. [CrossRef]
- 48. Stadlbauer, A.; Marhold, F.; Oberndorfer, S.; Heinz, G.; Buchfelder, M.; Kinfe, T.M.; Meyer-Bäse, A. Radiophysiomics: Brain Tumors Classification by Machine Learning and Physiological MRI Data. *Cancers* **2022**, *14*, 2363. [CrossRef]
- Carré, A.; Klausner, G.; Edjlali, M.; Lerousseau, M.; Briend-Diop, J.; Sun, R.; Ammari, S.; Reuzé, S.; Alvarez Andres, E.; Estienne, T.; et al. Standardization of brain MR images across machines and protocols: Bridging the gap for MRI-based radiomics. *Sci. Rep.* 2020, 10, 12340. [CrossRef]
- Collewet, G.; Strzelecki, M.; Mariette, F. Influence of MRI acquisition protocols and image intensity normalization methods on texture classification. *Magn. Reson. Imaging* 2004, 22, 81–91. [CrossRef] [PubMed]
- Shafiq-Ul-Hassan, M.; Zhang, G.G.; Latifi, K.; Ullah, G.; Hunt, D.C.; Balagurunathan, Y.; Abdalah, M.A.; Schabath, M.B.; Goldgof, D.G.; Mackin, D.; et al. Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. *Med. Phys.* 2017, 44, 1050–1062. [CrossRef]
- van Griethuysen, J.J.M.; Fedorov, A.; Parmar, C.; Hosny, A.; Aucoin, N.; Narayan, V.; Beets-Tan, R.G.H.; Fillion-Robin, J.-C.; Pieper, S.; Aerts, H.J.W.L. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res.* 2017, 77, e104–e107. [CrossRef]
- 53. Zwanenburg, A.; Vallières, M.; Abdalah, M.A.; Aerts, H.J.W.L.; Andrearczyk, V.; Apte, A.; Ashrafinia, S.; Bakas, S.; Beukinga, R.J.; Boellaard, R.; et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. *Radiology* 2020, 295, 328–338. [CrossRef] [PubMed]
- Zacharaki, E.I.; Wang, S.; Chawla, S.; Soo Yoo, D.; Wolf, R.; Melhem, E.R.; Davatzikos, C. Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme. *Magn. Reson. Med.* 2009, 62, 1609–1618. [CrossRef] [PubMed]
- 55. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic Minority Over-sampling Technique. J. Artif. Intell. Res. 2002, 16, 321–357. [CrossRef]
- An, J.Y.; Seo, H.; Kim, Y.-G.; Lee, K.E.; Kim, S.; Kong, H.-J. Codeless Deep Learning of COVID-19 Chest X-Ray Image Dataset with KNIME Analytics Platform. *Healthc. Inform. Res.* 2021, 27, 82–91. [CrossRef] [PubMed]
- 57. Berthold, M.R.; Cebron, N.; Dill, F.; Gabriel, T.R.; Kötter, T.; Meinl, T.; Ohl, P.; Thiel, K.; Wiswedel, B. KNIME-the Konstanz information miner: Version 2.0 and beyond. *ACM SIGKDD Explor. Newsl.* **2009**, *11*, 26–31. [CrossRef]
- 58. Weigert, M.; Schmidt, U.; Boothe, T.; Müller, A.; Dibrov, A.; Jain, A.; Wilhelm, B.; Schmidt, D.; Broaddus, C.; Culley, S.; et al. Content-aware image restoration: Pushing the limits of fluorescence microscopy. *Nat. Methods* 2018, 15, 1090–1097. [CrossRef] [PubMed]
- Kiranyaz, S.; Avci, O.; Abdeljaber, O.; Ince, T.; Gabbouj, M.; Inman, D.J. 1D convolutional neural networks and applications: A survey. *Mech. Syst. Signal Process.* 2021, 151, 107398. [CrossRef]
- 60. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* 2017, 60, 84–90. [CrossRef]
- 61. Ng, W.; Minasny, B.; Montazerolghaem, M.; Padarian, J.; Ferguson, R.; Bailey, S.; McBratney, A.B. Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared, mid-infrared, and their combined spectra. *Geoderma* **2019**, *352*, 251–267. [CrossRef]
- 62. Hochreiter, S.; SchmIDHuber, J. Long short-term memory. Neural Comput. 1997, 9, 1735–1780. [CrossRef] [PubMed]
- 63. Bradley, A.P. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit.* **1997**, 30, 1145–1159. [CrossRef]
- 64. Ozturk-Isik, E.; Cengiz, S.; Ozcan, A.; Yakicier, C.; Ersen Danyeli, A.; Pamir, M.N.; Özduman, K.; Dincer, A. Identification of *IDH* and TERTp mutation status using 1 H-MRS in 112 hemispheric diffuse gliomas. *J. Magn. Reson. Imaging* **2020**, *51*, 1799–1809. [CrossRef]
- Tatekawa, H.; Hagiwara, A.; Uetani, H.; Bahri, S.; Raymond, C.; Lai, A.; Cloughesy, T.F.; Nghiemphu, P.L.; Liau, L.M.; Pope, W.B.; et al. Differentiating *IDH* status in human gliomas using machine learning and multiparametric MR/PET. *Cancer Imaging* 2021, 21, 27. [CrossRef] [PubMed]
- Kesler, S.R.; Noll, K.; Cahill, D.P.; Rao, G.; Wefel, J.S. The effect of *IDH*1 mutation on the structural connectome in malignant astrocytoma. *J. Neurooncol.* 2017, 131, 565–574. [CrossRef]
- 67. Kesler, S.R.; Harrison, R.A.; Petersen, M.L.; Rao, V.; Dyson, H.; Alfaro-Munoz, K.; Weathers, S.-P.; de Groot, J. Pre-surgical connectome features predict *IDH* status in diffuse gliomas. *Oncotarget* **2019**, *10*, 6484–6493. [CrossRef] [PubMed]
- Karami, G.; Pascuzzo, R.; Figini, M.; Del Gratta, C.; Zhang, H.; Bizzi, A. Combining Multi-Shell Diffusion with Conventional MRI Improves Molecular Diagnosis of Diffuse Gliomas with Deep Learning. *Cancers* 2023, 15, 482. [CrossRef]

- 69. Choi, K.S.; Choi, S.H.; Jeong, B. Prediction of *IDH* genotype in gliomas with dynamic susceptibility contrast perfusion MR imaging using an explainable recurrent neural network. *Neuro Oncol.* **2019**, *21*, 1197–1209. [CrossRef]
- Clark, K.; Vendt, B.; Smith, K.; Freymann, J.; Kirby, J.; Koppel, P.; Moore, S.; Phillips, S.; Maffitt, D.; Pringle, M.; et al. The Cancer Imaging Archive (TCIA): Maintaining and operating a public information repository. *J. Digit. Imaging* 2013, 26, 1045–1057. [CrossRef]
- Scarpace, L.; Flanders, A.; Jain, R.; Mikkelsen, T.; Andrews, D. Data from REMBRANDT. Version 1. *Cancer Imaging Arch.* 2015. [CrossRef]
- 72. National Cancer Institute Clinical Proteomic Tumor Analysis Consortium. (CPTAC) Radiology Data from the Clinical Proteomic Tumor Analysis Consortium Glioblastoma Multiforme CPTAC-GBM collection. Version 6. *Cancer Imaging Arch.* 2019. [CrossRef]
- 73. 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. An anatomic transcriptional atlas of human glioblastoma. *Science* **2018**, *360*, 660–663. [CrossRef]
- Menze, B.H.; Jakab, A.; Bauer, S.; Kalpathy-Cramer, J.; Farahani, K.; Kirby, J.; Burren, Y.; Porz, N.; Slotboom, J.; Wiest, R.; et al. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Trans. Med. Imaging* 2015, 34, 1993–2024. [CrossRef] [PubMed]
- 75. van der Voort, S.R.; Incekara, F.; Wijnenga, M.M.J.; Kapsas, G.; Gahrmann, R.; Schouten, J.W.; Dubbink, H.J.; Vincent, A.J.P.E.; van den Bent, M.J.; French, P.J.; et al. The Erasmus Glioma Database (EGD): Structural MRI scans, WHO 2016 subtypes, and segmentations of 774 patients with glioma. *Data Br.* **2021**, *37*, 107191. [CrossRef] [PubMed]
- Chakrabarty, S.; LaMontagne, P.; Shimony, J.; Marcus, D.S.; Sotiras, A. MRI-based classification of *IDH* mutation and 1p/19q codeletion status of gliomas using a 2.5D hybrid multi-task convolutional neural network. *Neuro-Oncol. Adv.* 2023, 5, vdad023. [CrossRef] [PubMed]
- 77. Choi, Y.S.; Bae, S.; Chang, J.H.; Kang, S.-G.; Kim, S.H.; Kim, J.; Rim, T.H.; Choi, S.H.; Jain, R.; Lee, S.-K. Fully automated hybrid approach to predict the *IDH* mutation status of gliomas via deep learning and radiomics. *Neuro. Oncol.* 2021, 23, 304–313. [CrossRef] [PubMed]
- Bangalore Yogananda, C.G.; Wagner, B.C.; Truong, N.C.D.; Holcomb, J.M.; Reddy, D.D.; Saadat, N.; Hatanpaa, K.J.; Patel, T.R.; Fei, B.; Lee, M.D.; et al. MRI-Based Deep Learning Method for Classification of *IDH* Mutation Status. *Bioengineering* 2023, 10, 1045. [CrossRef] [PubMed]
- 79. Isensee, F.; Jaeger, P.F.; Kohl, S.A.A.; Petersen, J.; Maier-Hein, K.H. nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* **2021**, *18*, 203–211. [CrossRef]
- van der Voort, S.R.; Incekara, F.; Wijnenga, M.M.J.; Kapsas, G.; Gahrmann, R.; Schouten, J.W.; Nandoe Tewarie, R.; Lycklama, G.J.; De Witt Hamer, P.C.; Eijgelaar, R.S.; et al. Combined molecular subtyping, grading, and segmentation of glioma using multi-task deep learning. *Neuro Oncol.* 2023, 25, 279–289. [CrossRef] [PubMed]
- Cheng, J.; Liu, J.; Kuang, H.; Wang, J. A Fully Automated Multimodal MRI-Based Multi-Task Learning for Glioma Segmentation and *IDH* Genotyping. *IEEE Trans. Med. Imaging* 2022, 41, 1520–1532. [CrossRef] [PubMed]
- Chang, K.; Bai, H.X.; Zhou, H.; Su, C.; Bi, W.L.; Agbodza, E.; Kavouridis, V.K.; Senders, J.T.; Boaro, A.; Beers, A.; et al. Residual Convolutional Neural Network for the Determination of *IDH* Status in Low- and High-Grade Gliomas from MR Imaging. *Clin. Cancer Res.* 2018, 24, 1073–1081. [CrossRef] [PubMed]
- Nalawade, S.S.; Yu, F.F.; Bangalore Yogananda, C.G.; Murugesan, G.K.; Shah, B.R.; Pinho, M.C.; Wagner, B.C.; Xi, Y.; Mickey, B.; Patel, T.R.; et al. Brain tumor *IDH*, 1p/19q, and MGMT molecular classification using MRI-based deep learning: An initial study on the effect of motion and motion correction. *J. Med. Imaging* 2022, *9*, 016001. [CrossRef] [PubMed]
- 84. Manikis, G.C.; Ioannidis, G.S.; Siakallis, L.; Nikiforaki, K.; Iv, M.; Vozlic, D.; Surlan-Popovic, K.; Wintermark, M.; Bisdas, S.; Marias, K. Multicenter DSC-MRI-Based Radiomics Predict *IDH* Mutation in Gliomas. *Cancers* **2021**, *13*, 3965. [CrossRef] [PubMed]
- Ioannidis, G.S.; Pigott, L.E.; Iv, M.; Surlan-Popovic, K.; Wintermark, M.; Bisdas, S.; Marias, K. Investigating the value of radiomics stemming from DSC quantitative biomarkers in *IDH* mutation prediction in gliomas. *Front. Neurol.* 2023, 14, 1249452. [CrossRef] [PubMed]
- Wiggins, W.F.; Tejani, A.S. On the Opportunities and Risks of Foundation Models for Natural Language Processing in Radiology. *Radiol. Artif. Intell.* 2022, 4, e220119. [CrossRef] [PubMed]
- Moor, M.; Banerjee, O.; Abad, Z.S.H.; Krumholz, H.M.; Leskovec, J.; Topol, E.J.; Rajpurkar, P. Foundation models for generalist medical artificial intelligence. *Nature* 2023, 616, 259–265. [CrossRef]
- Krishnan, R.; Rajpurkar, P.; Topol, E.J. Self-supervised learning in medicine and healthcare. *Nat. Biomed. Eng.* 2022, 6, 1346–1352.
 [CrossRef]
- 89. Zhou, Y.; Chia, M.A.; Wagner, S.K.; Ayhan, M.S.; Williamson, D.J.; Struyven, R.R.; Liu, T.; Xu, M.; Lozano, M.G.; Woodward-Court, P.; et al. A foundation model for generalizable disease detection from retinal images. *Nature* **2023**, *622*, 156–163. [CrossRef]
- Beebe-Wang, N.; Celik, S.; Weinberger, E.; Sturmfels, P.; De Jager, P.L.; Mostafavi, S.; Lee, S.-I. Unified AI framework to uncover deep interrelationships between gene expression and Alzheimer's disease neuropathologies. *Nat. Commun.* 2021, 12, 5369. [CrossRef]
- 91. Panizza, E. DeepOmicsAE: Representing Signaling Modules in Alzheimer's Disease with Deep Learning Analysis of Proteomics, Metabolomics, and Clinical Data. J. Vis. Exp. 2023. [CrossRef]
- Haralick, R.M.; Shanmugam, K.; Dinstein, I. Textural Features for Image Classification. *IEEE Trans. Syst. Man. Cybern.* 1973, SMC-3, 610–621. [CrossRef]

- 93. Sun, C.; Wee, W.G. Neighboring gray level dependence matrix for texture classification. *Comput. Vision Graph. Image Process.* **1983**, 23, 341–352. [CrossRef]
- 94. Galloway, M.M. Texture analysis using gray level run lengths. Comput. Graph. Image Process. 1975, 4, 172–179. [CrossRef]
- 95. Thibault, G.; Fertil, B.; Navarro, C.; Pereira, S.; Cau, P.; Levy, N.; Sequeira, J.; Mari, J.-L. Shape and texture indexes application tocell nuclei classification. *Int. J. Pattern Recognit. Artif. Intell.* **2013**, *27*, 1357002. [CrossRef]
- 96. Amadasun, M.; King, R. Textural features corresponding to textural properties. *IEEE Trans. Syst. Man. Cybern.* **1989**, *19*, 1264–1274. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.