




## Article

# Assessing Versatile Machine Learning Models for Glioma Radiogenomic Studies across Hospitals

Risa K. Kawaguchi <sup>1,2,3,†</sup>, Masamichi Takahashi <sup>4,\*,†</sup>, Mototaka Miyake <sup>5</sup>, Manabu Kinoshita <sup>6</sup>, Satoshi Takahashi <sup>2,7</sup>, Koichi Ichimura <sup>8</sup>, Ryuji Hamamoto <sup>2,7</sup>, Yoshitaka Narita <sup>4</sup> and Jun Sese <sup>2,9</sup>

- <sup>1</sup> Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology, 2-3-26 Aomi, Koto-ku, Tokyo 135-0064, Japan; rkawaguc@cshl.edu
- <sup>2</sup> Division of Molecular Modification and Cancer Biology, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan; satoshi.takahashi.fy@riken.jp (S.T.); rhamamot@ncc.go.jp (R.H.); sesejun@humanome.jp (J.S.)
- <sup>3</sup> Cold Spring Harbor Laboratory, 1 Bungtown Rd, Cold Spring Harbor, NY 11724, USA
- <sup>4</sup> Department of Neurosurgery and Neuro-Oncology, National Cancer Center Hospital, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan; yonarita@ncc.go.jp
- <sup>5</sup> Department of Diagnostic Radiology, National Cancer Center Hospital, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan; mmiyake@ncc.go.jp
- <sup>6</sup> Department of Neurosurgery, Osaka University Graduate School of Medicine, 2-2 Yamadaoka, Suita 565-0871, Japan; mail@manabukinoshita.com
- <sup>7</sup> Cancer Translational Research Team, RIKEN Center for Advanced Intelligence Project, 1-4-1 Nihonba-shi, Chuo-ku, Tokyo 103-0027, Japan
- <sup>8</sup> Division of Brain Tumor Translational Research, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan; kichimur@ncc.go.jp
- <sup>9</sup> Humanome Laboratory, 2-4-10 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan
- \* Correspondence: masataka@ncc.go.jp; Tel.: +81-3-3542-2511
- † Co-first authors owing to their equal participation in this study.



**Citation:** Kawaguchi, R.K.; Takahashi, M.; Miyake, M.; Kinoshita, M.; Takahashi, S.; Ichimura, K.; Hamamoto, R.; Narita, Y.; Sese, J. Assessing Versatile Machine Learning Models for Glioma Radiogenomic Studies across Hospitals. *Cancers* **2021**, *13*, 3611. <https://doi.org/10.3390/cancers13143611>

Academic Editors: Jorn Fierstra, Margarida Julia-Sape and Katja Pinker-Domenig

Received: 28 March 2021  
Accepted: 15 July 2021  
Published: 19 July 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 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/>).

**Simple Summary:** Radiogenomics enables prediction of the status and prognosis of patients using non-invasively obtained imaging data. Current machine learning (ML) methods used in radiogenomics require huge datasets, which involve the handling of large heterogeneous datasets from multiple cohorts/hospitals. In this study, two different glioma datasets were used to test various ML and image pre-processing methods to confirm whether the models trained on one dataset are universally applicable to other datasets. Our result suggested that the ML method that yielded the highest accuracy in a single dataset was likely to be overfitted. We demonstrated that implementation of standardization and dimension reduction procedures prior to classification, enabled the development of ML methods that are less affected by the multiple cohort difference. We advocate using caution in interpreting the results of radiogenomic studies of the training and testing datasets that are small or mixed, with a view to implementing practical ML methods in radiogenomics.

**Abstract:** Radiogenomics use non-invasively obtained imaging data, such as magnetic resonance imaging (MRI), to predict critical biomarkers of patients. Developing an accurate machine learning (ML) technique for MRI requires data from hundreds of patients, which cannot be gathered from any single local hospital. Hence, a model universally applicable to multiple cohorts/hospitals is required. We applied various ML and image pre-processing procedures on a glioma dataset from The Cancer Image Archive (TCIA,  $n = 159$ ). The models that showed a high level of accuracy in predicting glioblastoma or WHO Grade II and III glioma using the TCIA dataset, were then tested for the data from the National Cancer Center Hospital, Japan (NCC,  $n = 166$ ) whether they could maintain similar levels of high accuracy. Results: we confirmed that our ML procedure achieved a level of accuracy (AUROC = 0.904) comparable to that shown previously by the deep-learning methods using TCIA. However, when we directly applied the model to the NCC dataset, its AUROC dropped to 0.383. Introduction of standardization and dimension reduction procedures before classification without re-training improved the prediction accuracy obtained using NCC (0.804) without a loss in prediction accuracy for the TCIA dataset. Furthermore, we confirmed the same tendency in a model for IDH1/2 mutation prediction with standardization and application of dimension reduction

that was also applicable to multiple hospitals. Our results demonstrated that overfitting may occur when an ML method providing the highest accuracy in a small training dataset is used for different heterogeneous data sets, and suggested a promising process for developing an ML method applicable to multiple cohorts.

**Keywords:** glioma; machine learning; radiogenomics; IDH; MGMT

## 1. Introduction

Magnetic resonance imaging (MRI) is widely used for cancer diagnoses. It is most frequently used to diagnose the pathology of brain tumors [1,2]. Besides conventional diagnostic information, MRI data may also contain phenotypic features of brain tumors, which are potentially associated with the underlying biology of both the tumor and the patient [3,4]. Thus, MRI is drawing attention as a source of information that may be utilized to predict genomic or clinical backgrounds of brain tumor patients, leading to the development of treatment strategies [2]. This field of research is termed “radiogenomics” [5].

Glioma presents a predominant target for radiogenomics, because identification of biomarkers that may help improve glioma patient outcomes is considered to be an urgent task [2,6,7]. WHO Grade IV glioblastoma (GBM), in particular, results in distinctly severe outcomes, compared to other WHO Grade II and III gliomas (lower grade gliomas: LrGG) and the five-year survival rate of GBM is as low as 6.8%, while that of LrGG is higher such as 51.6% for diffuse astrocytoma or 82.7% for oligodendroglioma [8]. Recent reports indicate that numerous genetic mutations may play a role in the heterogeneity of gliomas. The molecular landscape has been remarkably transformed by the detection of key genetic mutations or epigenetic modifications, such as IDH1 and IDH2 mutations, the TERT promoter mutation, chr1p/19q codeletion, and O-6-Methylguanine-DNA methyltransferase (MGMT) promoter methylation [7]. IDH1 and IDH2 mutations (hereafter referred to as IDH mutation) are mutually exclusive “truncal mutations”, which are frequently associated with better prognoses for gliomas, such as Grade II and Grade III gliomas, and thereby considered as prognostic factors for gliomas [9,10]. MGMT promoter methylation status is also known to be a prognosis factor of GBM associated with temozolomide, which is used as a first-line, standard chemotherapeutic agent for GBM patients.

In addition, recently, the WHO grading system of central nervous system tumors was updated to require genetic testing including IDH mutation, TERT promoter mutation, H3K27M mutation, 1p/19q codeletion for the precise diagnosis of gliomas. Currently these amendments can offer better understanding of glioma biology, but also raise a new problem that all the tumor tissues supposed to be a glioma should be ideally taken by surgery or biopsy and analyzed for genetic testing, but not all countries have enough pathological diagnosis capacity. As such, it is clinically useful if genetic backgrounds could also be obtained at the same time when the first pre-operative images of glioma patients are captured.

Glioma is a rare cancer and, therefore, the number of patients at any given hospital is not necessarily large. Some large public databases, such as The Cancer Imaging Archive (TCIA) GBM and LGG [6] or Repository of Molecular Brain Neoplasia Data (REMBRANDT) were published [11] in order to enhance radiogenomic studies targeting glioma. Many recent studies have applied a wide variety of machine learning (ML) methods from volumetric features to cutting-edge deep learning (DL) methods, in order to predict the grade and grouping of tumors [12,13], including IDH mutation [14,15], MGMT methylation status [16–21], survival [22–26], as well as other combinations of patient backgrounds [27–34]. On the other hand, studies using local datasets showed significant diversity in prediction accuracy for the same prediction target. For example, radiogenomic features are found to predict MGMT methylation status with 83% accuracy using deep learning [28], although another study achieved the accuracy only around 62% for the same data source [35].

While numerous studies have validated the robustness of gene expression signatures for GBM [15,36], only a few have used multiple datasets from different sources [32,37]. Although, based on radiomic image features, these studies could not show consistent improvement in prediction accuracy without mixing cohorts, or using datasets where the number of patients was not sufficient to reduce the variance of classification performances obtained by cross validation. Furthermore, most studies examine the prediction accuracy of the best methods within one dataset and do not show the results from the viewpoint of the performance stability across the datasets, particularly for the difference between public anonymized dataset and local data [38]. The reproducibility of radiogenomic studies has already gathered attention for the evaluation of different preprocessing strategies and tumor annotations [39,40]. Image data that depend on the hospital-specific MRI sequence parameters or image processing for analyses tend to have many systematic and inevitable differences, called batch effects [41,42]. A spatial standardization, which is an approach to scale each brain image to a reference brain, is one promising way to reduce the impact of the systematic differences, while even the choice of standardization is impactful for the performance stability [43]. This raises a query as to whether a model developed using images obtained from a public database, which are generally pre-processed for anonymization, is applicable to images obtained from a local hospital in a practical way. This type of problem is the main subject of transfer learning in the ML domain because re-training of the entire models demands a huge resource as well as annotated training data from a local hospital [44].

The objective of the present study is to identify the important factors to establish ML models that exhibit a high level of performance, not only in a single-cohort, but also in other hospital datasets. We examined the stability of prediction accuracy using two different MRI datasets, wherein the public TCIA dataset was used to generate a model, which was then directly applied to the National Cancer Center Hospital Japan (NCC) dataset. Because the public dataset is already preprocessed, this comparison gives us the clarification whether the trained model is still effective for datasets from local hospitals beyond the various hospital or cohort-specific biases. According to our results, the best model is shown to suffer from the over-fitting problem for another dataset, especially when the model does not include either standardization or dimension reduction processes. It would be a robust and practical solution to choose the better model from multiple models trained for the public dataset using a partially annotated local dataset.

## 2. Materials and Methods

### 2.1. TCIA and NCC Cohort

This study was approved by the National Cancer Center Institutional Review Board (study number: 2013-042). The NCC cohort dataset consisted of MRI scans, clinical information, and genomic information collected from 90 GBM and 76 LrGG patients who were treated at NCC between 2004 and 2017. Patient information is summarized (Table 1 and Supplementary Table S1). MRI data contained 4 widely used data types as follows: T1-weighted imaging (T1WI); gadolinium-enhanced T1-weighted imaging (Gd-T1WI); T2-weighted imaging (T2WI); and fluid-attenuated inversion recovery (FLAIR) sequence. All images were acquired in an axial direction with thicknesses in the range of 4–6 mm.

As another cohort, the public image and background data of 102 GBM and 65 LGG patients were obtained from the training dataset of Brain Tumor Segmentation (BraTS) challenge 2018 [6], which is based on TCIA information. This public dataset is constructed for the contest of brain tumor segmentation methods and will hereafter be referred to as the TCIA dataset. The images were already co-registered, skull-stripped, and re-sampled to a size of  $240 \times 240 \times 155$ . The genomic and clinical information from these were described previously [45]; (Table 1 and Supplementary Table S1). Patients with a set of at least 3 types of MRI scans, tumor annotation, and genomic information were selected.

**Table 1.** Demography of glioma patients in two different datasets used in this study, The Cancer Image Archive (TCIA) and National Cancer Center Hospital Japan (NCC). Each row represents the statistics or number of people assigned to each category for GBM and lower grade glioma patients. For genomic alterations, the number of people carrying IDH mutations, TERT promoter mutations, chr 1p19q codeletions, and higher levels of MGMT promoter methylation was separated from those of others via slash.

	TCIA		NCC	
	GBM	LGG	GBM	LwGG
Total	102	65	90	76
Age (average)	57.5	44.0	61.9	45.5
Gender (F/M)	32/55	38/27	39/51	29/47
KPS (average)	80.4	90.0	76.6	86.7
IDH mut vs wt	4/66	53/11	6/84	52/22
TERT mut vs wt	3/1	21/40	55/35	33/43
chr1p19q codelet	0/83	13/52	2/48	25/50
MGMT met H vs L	20/29	52/13	34/56	46/30

The data provided by Arita et al., which consisted of IDH1/2 mutations, codeletion of chromosome 1p and 19q (1p19q codeletion), TERT promoter mutations, and MGMT methylation, were used to represent the genomic and epigenomic features of the NCC dataset [7]. All values were binarized according to a previous study.

## 2.2. Region of Interest (ROI) Extraction

Tumor regions of the NCC cohort dataset were examined by a skilled radiologist (MM), who has been practicing for 18-years, using ImageJ [46]. For all 4 MRI types, 2 categories of ROI information tumor (enhanced in Gd-T1WI) region and edema (high intensity lesion in T2WI and FLAIR) were contoured via single strokes. Image features were computed for the region inside of the outline. ROIs were also transformed via realigning, co-registering, and normalizing in the same manner. Tumor and edema ROI of the TCIA dataset were provided by BraTS. Pixels annotated as “necrosis” or “enhancing tumor region” were labeled as tumor, and regions annotated as “necrosis”, “edema”, “non-enhancing tumor”, and “enhancing tumor region” were treated as edema ROI, which corresponds to the edema region and its inside. All results for features computed for tumor ROI or without ROI are shown in Supplementary Figures S1–S5.

## 2.3. Preprocessing of Brain Images

NCC-derived MRI data was converted from digital imaging and communications in medicine (DICOM) to the Neuroimaging Informatics Technology Initiative (NIfTI) format with neurological space, using SPM12 [47]. Skull-stripping was carried out via the standard\_space\_roi command in FSL-BET library [48]. For the TCIA dataset, we applied normalization and brain standardization to the downloaded MRI data in the NIfTI format that was already skull-stripped and resampled after normalization and anonymization processes.

Two frequently used spatial standardization procedures were applied to adjust differences in individual brain sizes and shapes; one was Tissue Probability Maps (TPM) in MNI space using SPM12 (referred to SPM), while the other was average brain space compiled by MICCAI multi atlas challenge data using ANTs [49]. We also applied two normalization methods to regulate differences in pixel brightness. Details pertaining to an entire standardization, such as pixel normalization, mapping, and skull-stripping processes are shown in Supplementary Material.

## 2.4. Calculating Image and Patient Features for ML

To examine the classification efficiency of a wide variety of image features, a maximum of 16,221 features were generated for each patient, using both images and clinical records for the following ML training and prediction classifiers. Identical procedures were applied to both the NCC cohort and the TCIA dataset. The features are summarized below, where  $M$  represents the number of features generated in each category (see Supplementary Materials).

1. Basic features (MRI scans;  $M = 28 \times 4$ ): statistics of pixel values from 4 types of MRI scans. Twenty-eight features were generated from each MRI type, consisting of percentile pixel values of the ROI as well as whole brain regions as a control with steps increasing by 10%. Several other statistics, such as mean, median, min, max, dimension size (x, y, z), and centroid coordinates of the whole brain were also included to control patient-specific image pattern.
2. Pyradiomics-based features (MRI scans;  $M = 960 \times 4$ ): first order statistics, shapes, and textures were calculated by pyradiomics (v2.1.0), which is frequently used software for radiomic analysis [3]. Pyradiomics was used for radiomic feature extraction from 2D and 3D images with ROI information. All parameters were computed except for NGTDM due to the long running time.
3. Pre-trained DL-based features (MRI scans;  $M = 3072 \times 4$ ): the DL model Inception-ResNet v2, pre-trained on ImageNet database, was used to obtain general image features. Outputs of the second to last layer were used as image features. Because the number of slices containing tumors was different for each patient, averages and sums of the outputs along the z-axis were calculated.
4. Anatomical tumor location (MRI scans;  $M = 30$ ): a vector representing occupancies of each anatomical region calculated via FSL (v5.0.10) [48] was used to represent information pertaining to 3D tumor position.
5. Clinical information ( $M = 3$ ): features of each patient, such as sex, age, and Karnofsky Performance Status (KPS) were used as additional features.

Most image features were computed only for edema (tumor, or all brain) regions, while basic features contained the metrics for both edema and whole image. In addition, the computation of DL-based features used rectangular images to cover all target tumor regions with the margin of 1 pixel using the command line tool “ExtractRegionFromImageByMask” included in the ANTs library [49]. The results shown in the main script were obtained using edema ROIs because of the high prediction performances. Those with tumor ROIs and without ROIs are found in Supplementary Figures S1–S5.

### 2.5. Feature Selection and Dimension Reduction Methods

One feature selection and two-dimension reduction methods were tested to demonstrate the efficacy of feature generation for transfer learning on radiogenomics. Logistic regression with L1 regularization and linear discriminant analysis, with the maximum number of features set to 200, was used for feature selection. Principal component analysis (PCA) and non-negative matrix factorization (NMF) were used as dimension reduction methods. PCA and NMF approximate input data into the low-dimensional matrices, where the dimension of the column or the row is a user-defined value, thereby reducing noise and integrating similar features into a single dimension. In both methods, the number of dimensions following decomposition was set to 8, 20, 40, and 200, in order to extract the primary features in the 4 MRI scan types. These analyses were implemented using Python 3 scikit-learn library.

### 2.6. Classification and Performance Evaluation for ML Classifiers

In order to evaluate the robustness of ML classifiers to dataset differences, widely-used and accurate ML classifiers were applied to the classification procedure as follows: random forest, k-nearest neighbor (with k set to 1, 3, and 5), support vector machine, linear discriminant analysis, AdaBoost, and XGBoost. For purposes of classification, feature vectors were applied without normalization to avoid effects of patient imbalance on each dataset. Five-fold cross validation was performed for the TCIA dataset in order to evaluate these methods. Next, in order to select the best-fit model for NCC, the model was applied to half of the NCC data that had been randomly sampled to validate each prediction problem (referred to as NCC validation). Lastly, the final accuracy of the model was estimated using the remaining half of the NCC data (NCC test). The area under the receiver operating

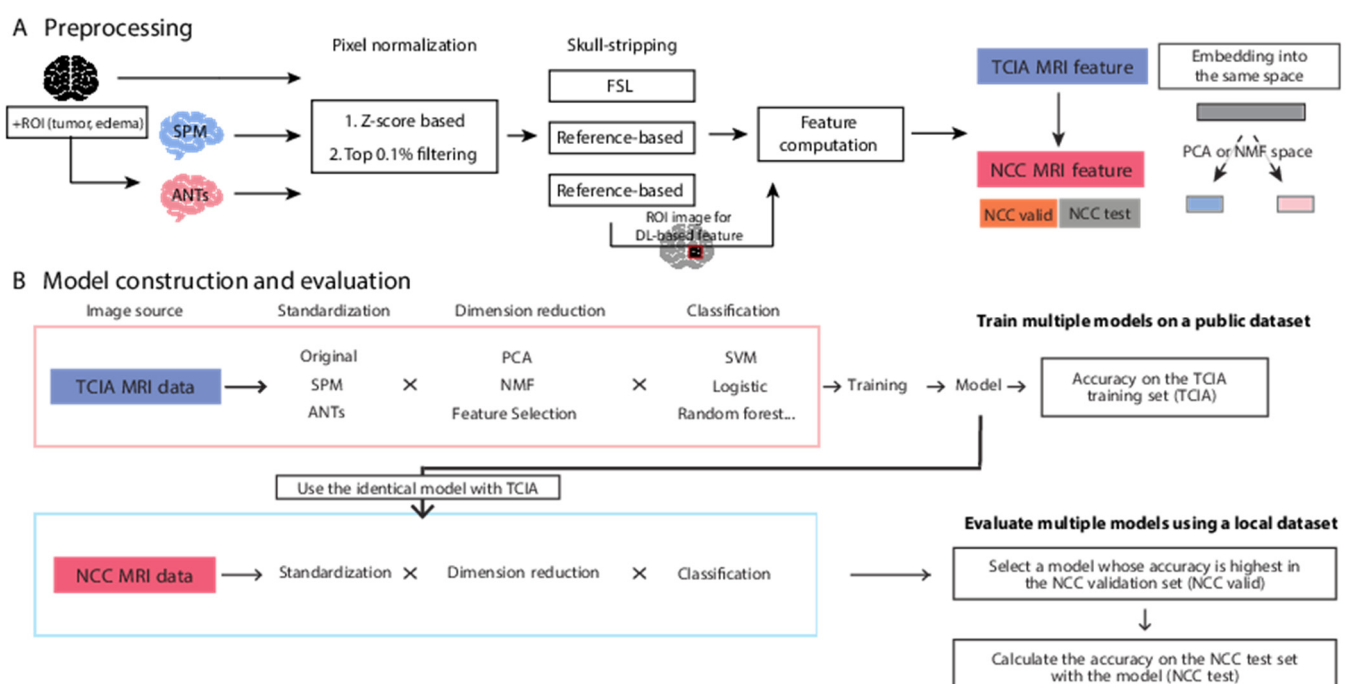
characteristic curve (AUROC) represented the measure of accuracy used to evaluate the general robustness of the models.

### 2.7. Publicly Available Brain Image Analysis Toolkit (PABLO)

The scripts used in this study are available at our repository, named PABLO (<https://github.com/carushi/PABLO>, accessed on 14 June 2021). PABLO is a versatile platform for analyzing combinations of methods, consisting of three steps widely used in MRI analysis; (1) standardization; (2) dimension reduction; and (3) classification. Standardization projects a patient head shape to the single standard to alleviate the individuality of brain images with adjusting the distribution of pixel brightness. Dimension reduction extracts or generates important features from the data, often resulting in stable classification results.

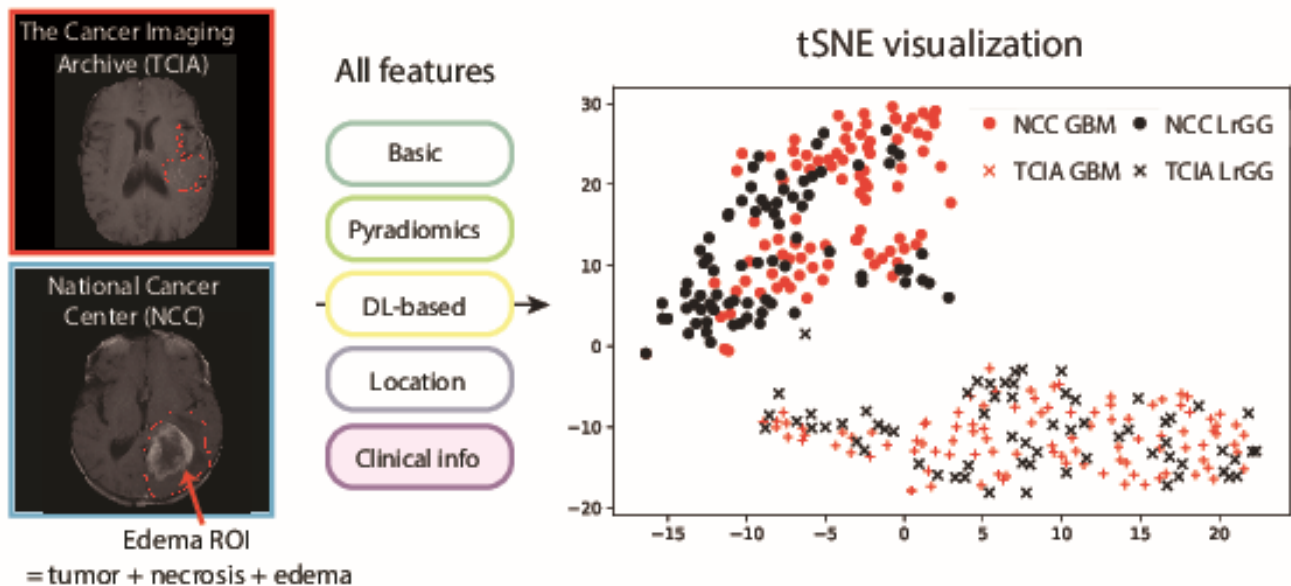
## 3. Results

Statistics of patients in the TCIA and NCC datasets are shown in Table 1. Both datasets displayed similar tendencies in terms of the number of GBM/LrGG, mutations and epigenomic status. On the other hand, the distribution of image feature values indicated that the distributions of the two datasets were clearly separated regardless of the property of patients (Figure 1). High-dimensional image features from each patient were calculated according to standard MRI analysis methods as described in Materials and Methods, and visualized in two dimensions via t-SNE [50]. While both TCIA and NCC datasets contained information from GBM, as well as LrGG patients, it is indicated that the patients were clustered according to the datasets that were obtained rather than their phenotypic properties, suggesting that the model generated for TCIA may not accurately work for NCC without additional training.



**Figure 1.** An overview of the dataset preprocessing and analytical methods used in this study. (A) Brain image data is applied to the multiple preprocessing steps; standardization of brain shape, pixel normalization, and skull-stripping. Then image features for radiogenomics are computed. To test the robustness of the trained models, we separated the whole NCC dataset to two independent partial NCC dataset: NCC validation and test set. The embedding space for dimension reduction is obtained from the TCIA dataset and applied to the partial NCC datasets independently. (B) In a machine learning workflow, all ML models were generated from TCIA data. The models were applied to TCIA (for cross validation), with half of the NCC data used for validation (termed NCC validation or NCC valid) and the other half of the NCC data (NCC test).

The application workflow of our ML procedures is depicted (Figure 2). In order to generate an ML procedure for the ML model that is sufficiently accurate for use in multiple cohorts or datasets, we developed a novel pipeline, termed PABLO. The purpose of classification was to generate a discriminative model between GBM/LrGG and other mutation/epigenetics types.

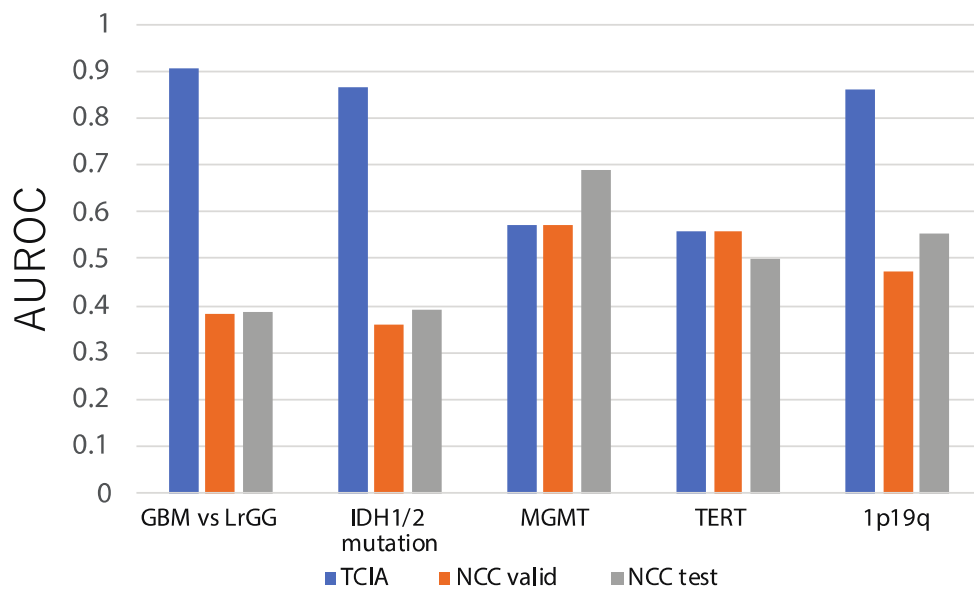


**Figure 2.** t-SNE visualization of all data from the TCIA and NCC datasets. Red and black colors depict GBM and LrGG patients, respectively, while crosses and circles indicate patients from TCIA and NCC datasets, respectively. The plots were based on image features calculated from MRI brain images and edema ROI annotation information and clinical features.

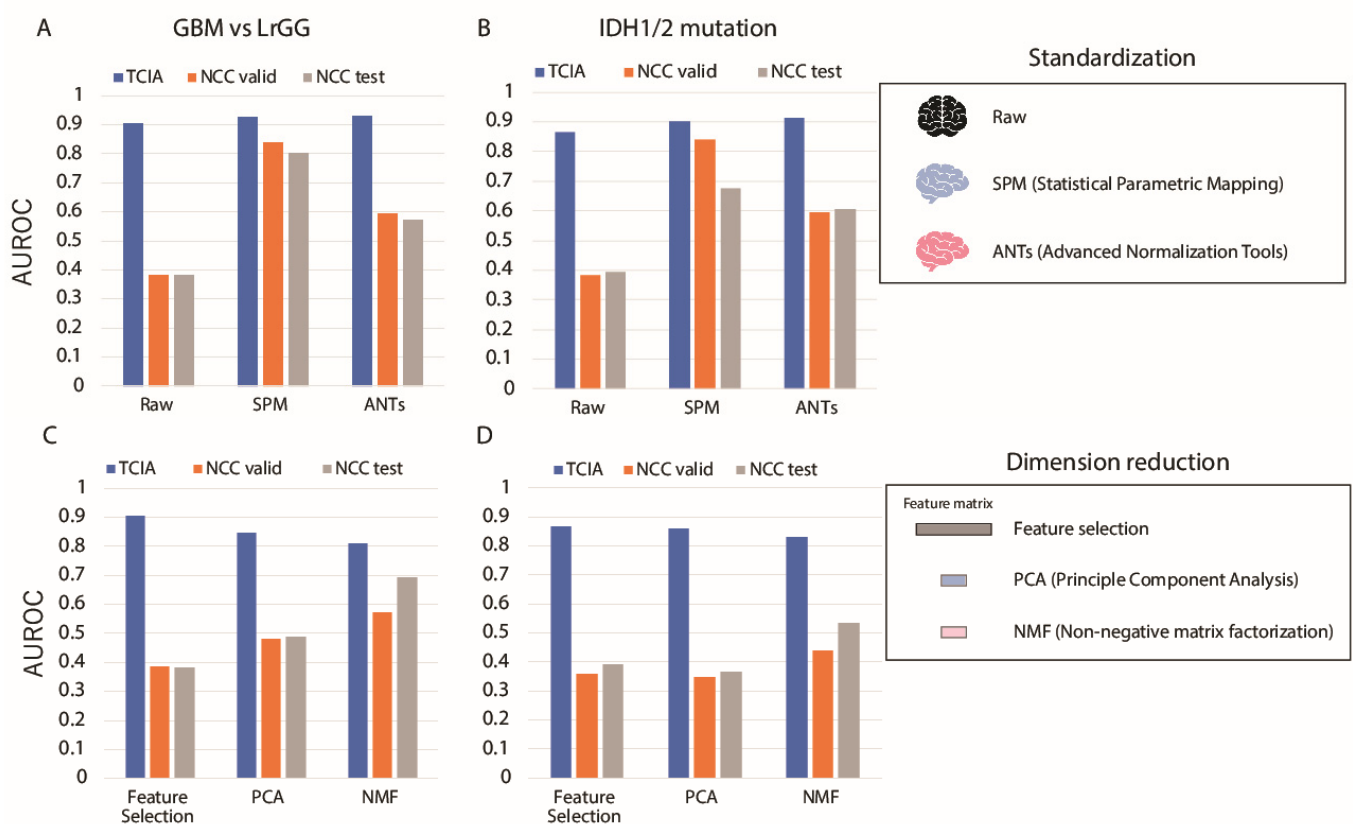
In Figure 3, the blue bars represent classification performances that evaluate model accuracies for the TCIA dataset by five-fold cross validation in the absence of normalization or dimension reduction. When the various classifiers implemented in PABLO were simply applied to predict available pathological or genetic characteristics related to glioma, such as GBM/LrGG classification (GBM prediction) and IDH mutation prediction (IDH prediction), the AUROC of the method was 0.904 for GBM prediction and 0.867 for IDH prediction; scores were comparable to those of previous methods, including a study based on DL [13,45]. Because of the significant clinical importance and high comparability with the previous radiogenomic studies, we focused on GBM and IDH predictions hereafter.

On the other hand, a model that is trained for TCIA and shows the highest accuracy in cross validation does not always achieve a high accuracy on other cohorts. The grey and orange bars seen in Figure 3 indicate the accuracies obtained when the best models developed in TCIA were applied to the NCC validation and test set (0.383 and 0.392, respectively), showing a significant decrease in accuracy for both cases.

Next, we applied standardization of brain shape prior to image feature computation, which was used to improve the accuracy of the model for radiogenomics in multiple cohorts. The accuracies obtained when two spatial standardization platforms SPM and ANTs were applied are compared with those without standardization for GBM and IDH prediction (Figure 4A,B). As with the case without standardization methods, the accuracies obtained for GBM prediction within the TCIA dataset were higher than 0.9 with SPM and ANTs, while those obtained via NCC changed drastically, depending on the standardization method used. In particular, the accuracy obtained via the SPM standardization method remained higher than 0.80, indicating that standardization may enhance applicability of a model to multi-cohort analyses. The accuracies estimated for IDH prediction also showed a similar tendency to those for GBM prediction. As such, it is suggested that standardization may improve the generalization performance of the generated model.



**Figure 3.** AUROCs were used as criteria to assess prediction accuracies of our machine learning workflow, without pre-processing, for 5 prediction problems: GBM/LrGG classification, IDH mutation existence, MGMT methylation status prediction, TERT promoter methylation prediction, and chr 1p19q co-deletion prediction. The accuracy of cross-validation results from TCIA (depicted as TCIA in Figure 2), as well as the accuracy of the application of the model to the NCC validation set (NCC valid) and the NCC test set (NCC test), are indicated by blue, orange, and grey bars, respectively.

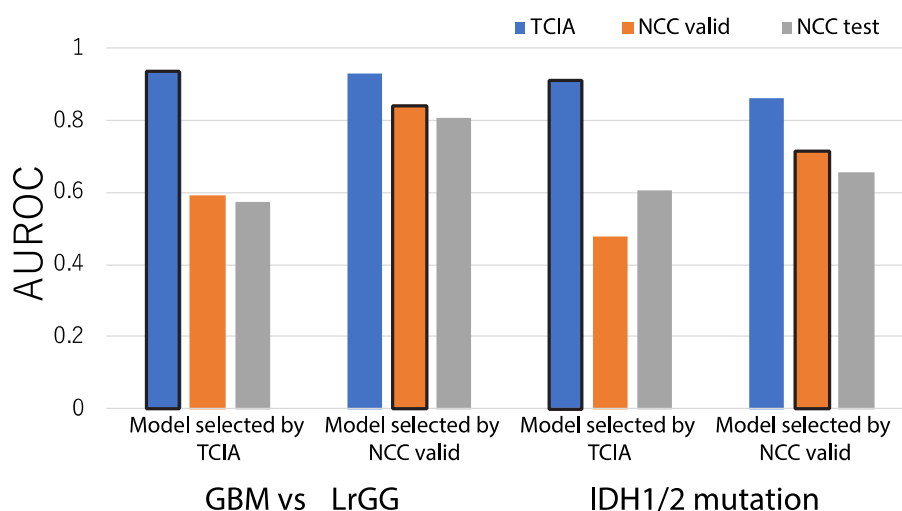


**Figure 4.** Changes in the classification performance of GBM (A,C) and IDH (B,D) prediction when standardization or dimension reduction methods were changed. (A,B), Changes in the accuracy of GBM and IDH prediction due to differences in the standardization method. Blue, orange, and gray bars correspond to TCIA cross validation, NCC validation, and NCC test accuracies, respectively. (C,D), Changes in the accuracy of GBM and IDH predictions due to differences in the dimension reduction method.



Dimension reduction is another pre-processing procedure to extract or generate lower dimensions representing the differences between different classes, such as GBM/LrGG or IDH mutation, beyond cohort differences. For the purpose to examine the influence of dimension reduction, simple feature selection methods, PCA, and NMF with varying the dimension size were performed before the classification (Figure 4C,D). All accuracies obtained for both prediction targets were high for TCIA when using dimension reduction methods. However, the accuracies for NCC test and validation using PCA and NMF showed substantially higher performances than those via feature selection in GBM prediction. For IDH prediction, the accuracies of NMF were higher than those of the others. These results indicate that dimension reduction has the potential to reduce the cohort-specific systematic differences and increase the accuracy for NCC (even when the model was generated using the TCIA dataset only).

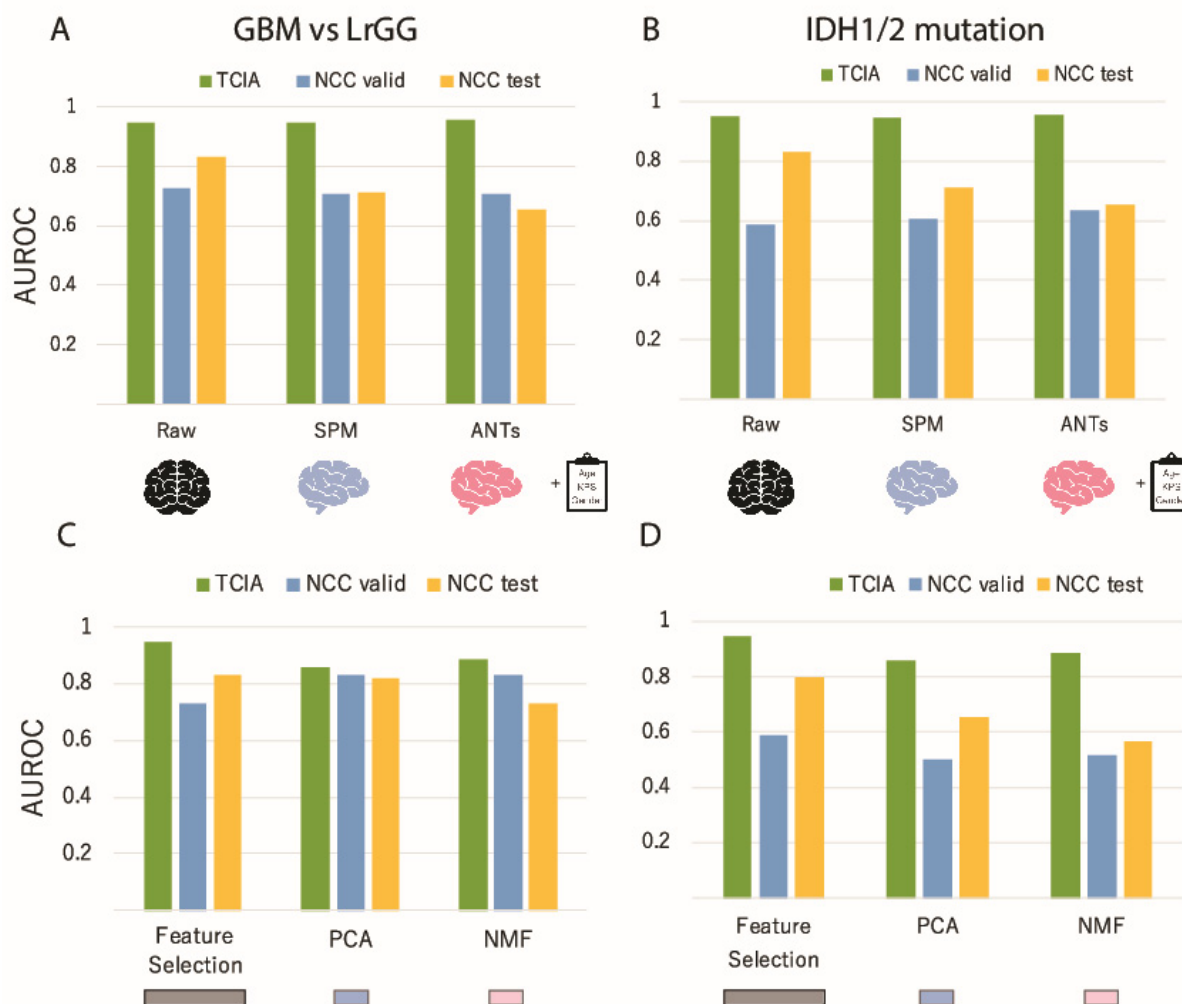
However, the model that achieved the highest accuracy for TCIA still did not achieve the highest accuracy for the NCC dataset, even after classification and dimension reduction were applied. Thus, we next assume a practical situation where a part of the local dataset, or NCC validation, is available to select a robust model among the generated models for TCIA. The accuracies of the two models are compared for the TCIA, NCC validation, and NCC test set to show which model is the best for the NCC test set (Figure 5). In either case, the model selected by the TCIA result was not the best, but worse for the NCC dataset, while the best model for the NCC validation set produced comparable accuracies for the TCIA dataset. As such, selecting the most robust model among all generated models according to their performances for partial validation data only is suggested to be an efficient way to obtain comparable prediction accuracy, even though the models are not re-trained for the new dataset.



**Figure 5.** Comparison of classification performances between the highest accuracy model in the TCIA and NCC validation dataset. The leftmost chart shows accuracies yielded by the model that showed the highest accuracy under cross validation within TCIA. The second leftmost chart shows the accuracies yielded by the model that showed the highest accuracy via the NCC validation set. The two graphs on the right depict the same settings as the previous examples, except that the prediction target is IDH mutation prediction.

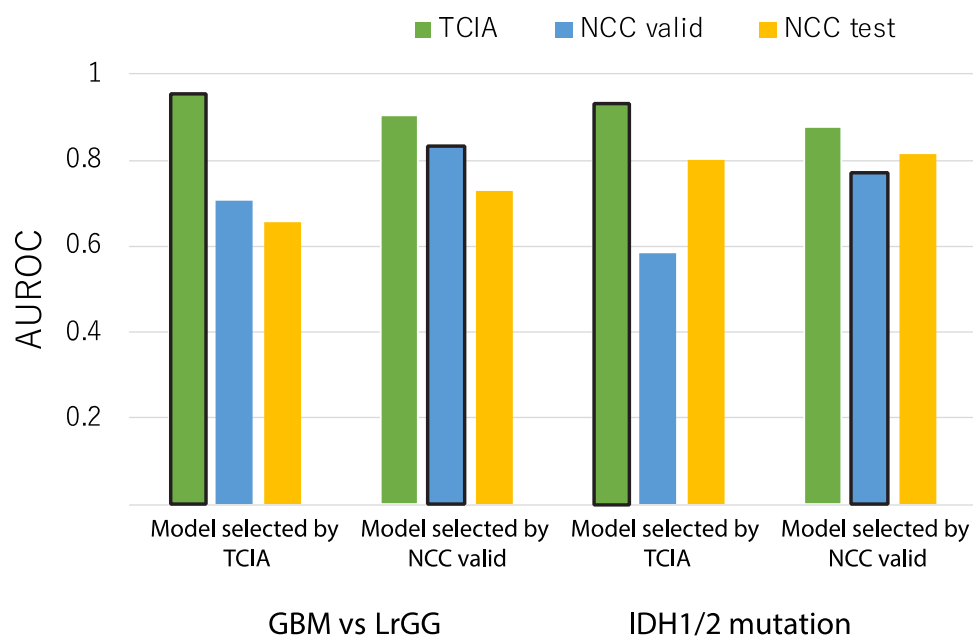
We further tested whether the accuracies showed similar tendencies even after clinical information was integrated into the radiomic features. We combined image information with clinical information, such as sex, age, and KPS, for classification, and examined the changes in accuracy in the same manner shown in Figures 3–5. Accuracies that were obtained using different standardization methods for GBM and IDH prediction are shown (Figure 6A,B). First, the accuracies obtained via TCIA and NCC overall tend to be higher than those obtained using only image information, indicating the significant contribution

made by clinical information. In addition, the accuracies were less sensitive for the difference of the standardization methods used. Similar calculations performed for dimension reduction methods are shown (Figure 6C,D). These results indicated that while combining image features with clinical information substantially increased prediction performance, dimension reduction was not sufficiently effective in bringing about improvement as much as that shown in classification based on image features alone.



**Figure 6.** Changes in classification performance for GBM (A,C) and IDH (B,D) prediction due to differences between standardization and dimension reduction methods. The difference of this figure from Figure 4 is that input data comprised images as well as clinical information, whereas in Figure 4 only image features are used. (A,B) Comparison of accuracies due to differences in the standardization method for GBM and IDH prediction, respectively. (C,D), Comparison of accuracies due to differences in dimension reduction methods for GBM and IDH prediction, respectively.

However, the importance of standardization and dimension reduction was revealed when the accuracies were compared between the models, showing the highest accuracy for the TCIA and NCC validation set (Figures 5 and 7). As a result, the selected models that showed the best accuracy for the NCC validation dataset were based on both SPM for standardization and NMF for dimension reduction. This result suggests that a combination of standardization and dimension reduction may enhance the stability and reproducibility of predictions based on image features and clinical information across the different cohorts, such as the public training dataset and test set from local hospitals.



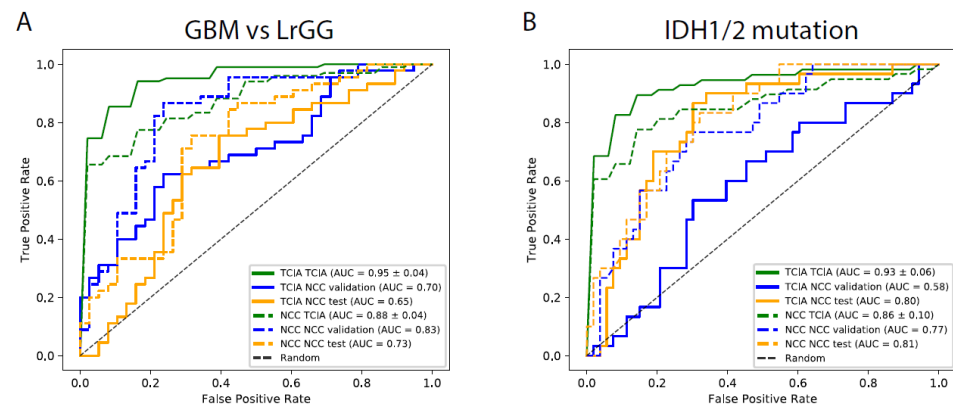
**Figure 7.** Comparison of classification performance between the highest accuracy models in TCIA and NCC validation. The experimental conditions were similar to those represented by Figure 5, except for the fact that input data contained clinical information. The two left and right graphs depict GBM and IDH prediction, respectively. In each group of 2 graphs, the left graph represents the model with the highest accuracy in TCIA while the right represents the model with the highest accuracy in the NCC validation set.

#### 4. Discussion

Radiogenomics is a rapidly developing field of research to predict the genetic backgrounds of neoplasms, including gliomas using medical images including computed tomography, MRI, and ultrasound, among others [51,52]. Precise diagnoses for gliomas require surgically obtained specimens for further investigation of immunohistochemistry, fluorescence in situ hybridization, and DNA sequencing. Although it is reported that the extent of resection (particularly when close to complete resection) is associated with better prognoses in GBM [53], the extent of resection was not prognostic for IDH 1/2 mutant and 1p/19q co-deleted tumors, which are known to be sensitive to chemotherapy as well as radiotherapy [54,55]. Hence, the potential impact of radiogenomics is that it offers a non-invasive method to predict these genetic markers important for diagnosis and prognosis. In fact, the WHO classification of tumors will be updated soon, in 2021, and include the need of analysis for chromosome 7 and loss of whole chromosome 10, and TERT promoter mutation and EGFR amplification for the diagnosis of glioblastoma even though the tumor has histologically low grade glioma features. Therefore, the importance of genetic background detection is increasing along with the discovery of more genes involved in modulating the tumor growth. Recent studies revealed the heterogeneity of genetic background within the tumor, which is expected to be decoded by radiogenomic studies [56].

The objective of this study was to discover important factors to establish a classifier that is able to perform better, not only in the context of single-cohort cross validation, but also when the model is used in other hospitals. Due to requirements associated with maintaining confidentiality regarding patient images and clinical information, we assumed a general situation, where the classifier was required to be applicable to a local dataset, using parameters trained only by a public dataset. Classification accuracies obtained via cross validation in TCIA were notably higher than those obtained by applying the model to either of the NCC validation or test sets, indicating a loss of generality in the model generated and selected from the TCIA dataset (Figures 6 and 8). However, the findings

also showed the presence of models having accuracies that were similar to those of the best models, demonstrating that pre-processing via brain-embedding yielded a much more robust model over cohorts/hospitals and clinical information was as important as image data. Thus standardization and dimension reduction, particularly for the combination of SPM and NMF, have potential to generate models that can be used over multiple hospitals.



**Figure 8.** Receiver operating characteristic (ROC) curves for classification results. The ROC curves enable the visualization of results in Figure 7 for GBM (A) and IDH prediction (B), respectively. Solid and dotted lines show the curves for the highest accuracy model in TCIA and NCC validation set, respectively. Each color corresponds to that used in Figure 7, where green, blue, and yellow depict TCIA, NCC validation, and NCC tests, respectively.

In addition to investigating different pre-processing and dimension reduction procedures, we also analyzed different brightness standardization methods and determined that changes in brightness did not completely fill the gap between TCIA and NCC images. Furthermore, the effect of ROI information was investigated and the results indicated that using tumor ROI did not improve prediction accuracies for IDH and MGMT mutations in GBM patients (Supplementary Figures S1 and S2). Prediction accuracies obtained without ROI information showed different tendencies for TCIA and NCC datasets. While accuracy obtained via cross validation was higher than 0.8 for GBM and IDH prediction in TCIA, the highest accuracy models obtained via the NCC dataset were only 0.4–0.6 AUROCs, regardless of the pre-processing methods used (Supplementary Figures S3–S5). This suggested the importance of using edema ROI information based on classical radiomic features for prediction improvement across cohorts. The classification performances of each single feature are also varied, depending on the source data types. Hence, the classification performance of each feature may be substantially changed according to the input and ROI (Supplementary Figure S6).

Furthermore, we examined the differences in accuracy associated with different ML classifiers and determined that the higher the accuracy of the model trained and evaluated on the same dataset was, the lower the accuracy of the model tested on the different dataset (Supplementary Figure S7). In addition, the procedures without dimension reduction methods tended to yield a lower accuracy for cross-dataset applications instead of a higher accuracy via cross validation, compared to the ones with dimension reduction methods, indicating the overfitting potential of each ML classifier in the cases without dimension reduction.

## 5. Conclusions

In this study, we performed a comprehensive assessment of radiogenomic classification models using two glioma datasets that are suffering from systematic biases for hospitals involved in the dataset acquisition and image pre-processing. Our results suggest that the ML method that yielded the highest accuracy in a single dataset was likely to be overfitted and showed the severe decrease of prediction accuracies for another dataset.

We further tested the impact of the implementation of standardization and dimension reduction procedures prior to classification. As a result, it enabled the development of ML methods that are less affected by the cohort differences. Our results advocate using caution in evaluating the accuracy of radiogenomic ML models in which the training and testing datasets are small or mixed, with a view to implement practical ML methods in radiogenomics.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/cancers13143611/s1>, Supplementary document: PDF document including the details of methods and supplementary data, Figure S1: AUROCs based on tumor ROI, Figure S2: the best models based on tumor ROI, Figure S3: AUROCs without ROI information with different standardization methods, Figure S4: AUROCs after applying dimension reduction methods without ROI information, Figure S5: the best models without ROI information, Figure S6: Comparison of classification performances of each pair of feature and image type by cross validation using the NCC dataset, Figure S7: Comparison of eight machine learning classifiers for the whole NCC dataset, Table S1: patient demographics in each cohort. Table S2: Summary of MRI scanners and scan protocols for each MRI data. Table S3: Summary of p-values for the comparison of ROC curves used in the main figures.

**Author Contributions:** Conceptualization, R.K.K., M.T., M.M., M.K., S.T., K.I., R.H., Y.N., and J.S.; resources and data curation, R.K.K., M.T., M.M., K.I., Y.N., and J.S.; software, R.K.K.; formal analysis and investigation, R.K.K., M.T., M.M., M.K., S.T., and J.S.; writing: R.K.K., M.T., M.M., M.K., S.T., K.I., R.H., Y.N., and J.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Japan Science and Technology agency (CREST JP-MJCR1689 to R.H.) and Japan Society for the Promotion of Science (17J01882 to R.K.K., JP18H04908 to R.H., and 20K17982 to S.T.).

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board of the National Cancer Center (protocol code: 2013-042, date of approval: 10 May 2013).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study. Written informed consent was obtained from the patients to publish this paper.

**Data Availability Statement:** Segmentation labels and radiomic features for the pre-operative scans of the TCGA-LGG and -GBM collection can be downloaded from the TCIA database: <https://www.cancerimagingarchive.net/> (accessed on 3 June 2021). Information on the biomarkers and demography of the patients were obtained from corresponding [40]. The statistics of the subjects in the NCC dataset were published in the previous study [7], whereas the NCC image dataset is not publicly available.

**Acknowledgments:** The experiment was conducted on a laptop computer (macOS Mojave 2.8GHz dual-core Intel core i7 CPU) and AIST AI Cloud system (AAIC). We sincerely thank the patients treated at NCC and their families for allowing us to analyze their clinical information and image data and their willingness to promote scientific understanding of glioma. We thank Hidenori Sakanashi, Hirokazu Nosato, and Masahiro Murakawa for the instructive comments about dimension reduction methods for medical images. We also appreciate Kyoko Fujioka, Yukiko Ochi, and Motoko Tsuji for their dedicated support.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Zhou, M.; Scott, J.; Chaudhury, B.; Hall, L.; Goldgof, D.; Yeom, K.; Iv, M.; Ou, Y.; Kalpathy-Cramer, J.; Napel, S.; et al. Radiomics in Brain Tumor: Image Assessment, Quantitative Feature Descriptors, and Machine-Learning Approaches. *Am. J. Neuroradiol.* **2018**, *39*, 208–216. [[CrossRef](#)] [[PubMed](#)]
2. Pope, W.B.; Brandal, G. Conventional and advanced magnetic resonance imaging in patients with high-grade glioma. *Q. J. Nucl. Med. Mol. Imaging* **2018**, *62*, 239–253. [[CrossRef](#)]
3. 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](#)]

4. Rudie, J.D.; Rauschecker, A.M.; Bryan, R.N.; Davatzikos, C.; Mohan, S. Emerging Applications of Artificial Intelligence in Neuro-Oncology. *Radiology* **2019**, *290*, 607–618. [[CrossRef](#)] [[PubMed](#)]
5. Kickingereder, P.; Bonekamp, D.; Nowosielski, M.; Kratz, A.; Sill, M.; Burth, S.; Wick, A.; Eidel, O.; Schlemmer, H.; Radbruch, A.; et al. Radiogenomics of Glioblastoma: Machine Learning-based Classification of Molecular Characteristics by Using Multi-parametric and Multi-regional MR Imaging Features. *Radiology* **2016**, *281*, 907–918. [[CrossRef](#)] [[PubMed](#)]
6. Bakas, S.; Akbari, H.; Sotiras, A.; Bilello, M.; Rozycki, M.; Kirby, J.; Freymann, J.B.; Farahani, K.; Davatzikos, C. Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Sci. Data* **2017**, *4*, 170117. [[CrossRef](#)] [[PubMed](#)]
7. Arita, H.; Yamasaki, K.; Matsushita, Y.; Nakamura, T.; Shimokawa, A.; Takami, H.; Tanaka, S.; Mukasa, A.; Shirahata, M.; Shimizu, S.; et al. A combination of TERT promoter mutation and MGMT methylation status predicts clinically relevant subgroups of newly diagnosed glioblastomas. *Acta Neuropathol. Commun.* **2016**, *4*, 1–14. [[CrossRef](#)]
8. Ostrom, Q.T.; Cioffi, G.; Gittleman, H.; Patil, N.; Waite, K.; Kruchko, C.; Barnholtz-Sloan, J.S. CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2012–2016. *Neuro-Oncology* **2019**, *21* (Suppl. 5), v1–v100. [[CrossRef](#)]
9. Suzuki, H.; Aoki, K.; Chiba, K.; Sato, Y.; Shiozawa, Y.; Shiraishi, Y.; Shimamura, T.; Niida, A.; Motomura, K.; Ohka, F.; et al. Mutational landscape and clonal architecture in grade II and III gliomas. *Nat. Genet.* **2015**, *47*, 458–468. [[CrossRef](#)]
10. Louis, D.N.; Perry, A.; Reifenberger, G.; von Deimling, A.; Figarella-Branger, D.; Cavenee, W.K.; Ohgaki, H.; Wiestler, O.D.; Kleihues, P.; Ellison, D.W. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: A summary. *Acta Neuropathol.* **2016**, *131*, 803–820. [[CrossRef](#)]
11. Gusev, Y.; Bhuvaneshwar, K.; Song, L.; Zenklusen, J.C.; Fine, H.; Madhavan, S. The REMBRANDT study, a large collection of genomic data from brain cancer patients. *Sci. Data* **2018**, *5*, 180158. [[CrossRef](#)]
12. Alberts, E.; Tetteh, G.; Trebeschi, S.; Bieth, M.; Valentinitsch, A.; Wiestler, B.; Zimmer, C.; Menze, B. Multi-modal image classification using low-dimensional texture features for genomic brain tumor recognition. In *Graphs in Biomedical Image Analysis, Computational Anatomy and Imaging Genetics*; Springer: Basel, Switzerland, 2017; pp. 201–209.
13. Cho, H.-H.; Park, H. Classification of low-grade and high-grade glioma using multi-modal image radiomics features. In Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jeju, Korea, 11–15 July 2017; Volume 2017, pp. 3081–3084.
14. Kinoshita, M.; Sakai, M.; Arita, H.; Shofuda, T.; Chiba, Y.; Kagawa, N.; Watanabe, Y.; Hashimoto, N.; Fujimoto, Y.; Yoshimine, T.; et al. Introduction of High Throughput Magnetic Resonance T2-Weighted Image Texture Analysis for WHO Grade 2 and 3 Gliomas. *PLoS ONE* **2016**, *11*, e0164268. [[CrossRef](#)] [[PubMed](#)]
15. Liu, X.; Li, Y.; Sun, Z.; Li, S.; Wang, K.; Fan, X.; Liu, Y.; Wang, L.; Wang, Y.; Jiang, T. Molecular profiles of tumor contrast enhancement: A radiogenomic analysis in anaplastic gliomas. *Cancer Med.* **2018**, *7*, 4273–4283. [[CrossRef](#)]
16. Ahn, S.S.; Shin, N.-Y.; Chang, J.H.; Kim, S.H.; Kim, E.H.; Kim, D.W.; Lee, S.-K. Prediction of methylguanine methyltransferase promoter methylation in glioblastoma using dynamic contrast-enhanced magnetic resonance and diffusion tensor imaging. *J. Neurosurg.* **2014**, *121*, 367–373. [[CrossRef](#)]
17. Drabycz, S.; Roldán, G.; de Robles, P.; Adler, D.; McIntyre, J.B.; Magliocco, A.M.; Cairncross, J.G.; Mitchell, J.R. An analysis of image texture, tumor location, and MGMT promoter methylation in glioblastoma using magnetic resonance imaging. *NeuroImage* **2010**, *49*, 1398–1405. [[CrossRef](#)]
18. Ellingson, B.M.; Cloughesy, T.F.; Pope, W.; Zaw, T.M.; Phillips, H.; Lalezari, S.; Nghiemphu, P.L.; Ibrahim, H.; Naeini, K.M.; Harris, R.J.; et al. Anatomic localization of O6-methylguanine DNA methyltransferase (MGMT) promoter methylated and unmethylated tumors: A radiographic study in 358 de novo human glioblastomas. *NeuroImage* **2012**, *59*, 908–916. [[CrossRef](#)] [[PubMed](#)]
19. Kanas, V.G.; Zacharaki, E.I.; Thomas, G.A.; Zinn, P.O.; Megalooikonomou, V.; Colen, R.R. Learning MRI-based classification models for MGMT methylation status prediction in glioblastoma. *Comput. Methods Programs Biomed.* **2017**, *140*, 249–257. [[CrossRef](#)] [[PubMed](#)]
20. Li, Z.-C.; Bai, H.; Sun, Q.; Li, Q.; Liu, L.; Zou, Y.; Chen, Y.; Liang, C.; Zheng, H. Multi-regional radiomics features from multi-parametric MRI for prediction of MGMT methylation status in glioblastoma multiforme: A multicentre study. *Eur. Radiol.* **2018**, *28*, 3640–3650. [[CrossRef](#)] [[PubMed](#)]
21. Xi, Y.-B.; Guo, F.; Xu, Z.; Li, C.; Wei, W.; Tian, P.; Liu, T.-T.; Liu, L.; Cheng, G.; Ye, J.; et al. Radiomics signature: A potential biomarker for the prediction of MGMT promoter methylation in glioblastoma. *J. Magn. Reson. Imaging* **2018**, *47*, 1380–1387. [[CrossRef](#)] [[PubMed](#)]
22. Chaddad, A.; Sabri, S.; Niazi, T.; Abdulkarim, B. Prediction of survival with multi-scale radiomic analysis in glioblastoma patients. *Med. Biol. Eng. Comput.* **2018**, *56*, 2287–2300. [[CrossRef](#)] [[PubMed](#)]
23. Gevaert, O.; Mitchell, L.A.; Achrol, A.S.; Xu, J.; Echeagaray, S.; Steinberg, G.K.; Cheshier, S.H.; Napel, S.; Zaharchuk, G.; Plevritis, S.K. Glioblastoma Multiforme: Exploratory Radiogenomic Analysis by Using Quantitative Image Features. *Radiology* **2014**, *273*, 168–174. [[CrossRef](#)]
24. Macyszyn, L.; Akbari, H.; Pisapia, J.M.; Da, X.; Attiah, M.; Pigrish, V.; Bi, Y.; Pal, S.; Davuluri, R.V.; Roccograndi, L.; et al. Imaging patterns predict patient survival and molecular subtype in glioblastoma via machine learning techniques. *Neuro-Oncology* **2015**, *18*, 417–425. [[CrossRef](#)]

25. Rao, A.; Rao, G.; Gutman, D.A.; Flanders, A.E.; Hwang, S.N.; Rubin, D.L.; Colen, R.R.; Zinn, P.O.; Jain, R.; Wintermark, M.; et al. A combinatorial radiographic pheno-type may stratify patient survival and be associated with invasion and proliferation characteristics in glioblastoma. *J. Neurosurg.* **2016**, *124*, 1008–1017. [[CrossRef](#)]
26. Rios Velazquez, E.; Meier, R.; Dunn, W.D., Jr.; Alexander, B.; Wiest, R.; Bauer, S.; Gutman, D.A.; Reyes, M.; Aerts, H.J.W.L. Fully automatic GBM segmentation in the TCGA-GBM dataset: Prognosis and correlation with VASARI features. *Sci. Rep.* **2015**, *5*, 16822. [[CrossRef](#)]
27. Arita, H.; Kinoshita, M.; Kawaguchi, A.; Takahashi, M.; Narita, Y.; Terakawa, Y.; Tsuyuguchi, N.; Okita, Y.; Nonaka, M.; Moriuchi, S.; et al. Lesion location implemented magnet-ic resonance imaging radiomics for predicting IDH and TERT promoter mutations in grade II/III gliomas. *Sci. Rep.* **2018**, *8*, 11773. [[CrossRef](#)]
28. Chang, P.; Grinband, J.; Weinberg, B.D.; Bardis, M.; Khy, M.; Cadena, G.; Su, M.-Y.; Cha, S.; Filippi, C.G.; Bota, D.; et al. Deep-Learning Convolutional Neural Networks Accurately Classify Genetic Mutations in Gliomas. *Am. J. Neuroradiol.* **2018**, *39*, 1201–1207. [[CrossRef](#)] [[PubMed](#)]
29. Chen, L.; Zhang, H.; Thung, K.-H.; Liu, L.; Lu, J.; Wu, J.; Wang, Q.; Shen, D. Multi-label Inductive Matrix Completion for Joint MGMT and IDH1 Status Prediction for Glioma Patients. In *Lecture Notes in Computer Science*; Springer: Basel, Switzerland, 2017; Volume 10434, pp. 450–458.
30. Gutman, D.A.; Cooper, L.A.; Hwang, S.N.; Holder, C.A.; Gao, J.; Aurora, T.D.; Dunn, W.D., Jr.; Scarpace, L.; Mikkelsen, T.; Jain, R.; et al. MR Imaging Predictors of Mo-lecular Profile and Survival: Multi-institutional Study of the TCGA Glioblastoma Data Set. *Radiology* **2013**, *267*, 560–569. [[CrossRef](#)] [[PubMed](#)]
31. Lasocki, A.; Gaillard, F.; Gorelik, A.; Gonzales, M. MRI Features Can Predict 1p/19q Status in Intracranial Gliomas. *Am. J. Neuroradiol.* **2018**, *39*, 687–692. [[CrossRef](#)] [[PubMed](#)]
32. Lu, C.-F.; Hsu, F.-T.; Hsieh, K.L.-C.; Kao, Y.-C.J.; Cheng, S.-J.; Hsu, J.B.-K.; Tsai, P.-H.; Chen, R.-J.; Huang, C.-C.; Yen, Y.; et al. Machine Learning–Based Radiomics for Molecular Subtyping of Gliomas. *Clin. Cancer Res.* **2018**, *24*, 4429–4436. [[CrossRef](#)] [[PubMed](#)]
33. Paech, D.; Windschuh, J.; Oberhollenzer, J.; Dreher, C.; Sahm, F.; Meissner, J.E.; Goerke, S.; Schuenke, P.; Zaiss, M.; Regnery, S.; et al. Assessing the predictability of IDH mutation and MGMT methylation status in glioma patients using relaxa-tion-compensated multipool CEST MRI at 7.0 T. *Neuro-Oncology* **2018**, *20*, 1661–1671. [[CrossRef](#)]
34. Park, Y.; Han, K.; Ahn, S.; Choi, Y.; Chang, J.; Kim, S.; Kang, S.-G.; Kim, E.; Lee, S.-K. Whole-Tumor Histogram and Texture Analyses of DTI for Evaluation of IDH1-Mutation and 1p/19q-Codeletion Status in World Health Organization Grade II Gliomas. *Am. J. Neuroradiol.* **2018**, *39*, 693–698. [[CrossRef](#)]
35. Takahashi, S.; Takahashi, M.; Kinoshita, M.; Miyake, M.; Kawaguchi, R.; Shinjima, N.; Mukasa, A.; Saito, K.; Nagane, M.; Otani, R.; et al. Development of fi-ne-tuning method of MR images of gliomas to normalize image differences among facilities. *Cancers* **2021**, *13*, 1415. [[CrossRef](#)]
36. Tang, J.; He, D.; Yang, P.; He, J.; Zhang, Y. Genome-wide expression profiling of glioblastoma using a large combined cohort. *Sci. Rep.* **2018**, *8*, 15104. [[CrossRef](#)] [[PubMed](#)]
37. 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 Determi-nation of IDH Status in Low- and High-Grade Gliomas from MR Imaging. *Clin. Cancer Res.* **2018**, *24*, 1073. [[CrossRef](#)]
38. Zhang, L.; Tanno, R.; Bronik, K.; Jin, C.; Nachev, P.; Barkhof, F.; Ciccarelli, O.; Alexander, D.C. Learning to segment when experts disagree. In *Lecture Notes in Computer Science*; Springer: Basel, Switzerland, 2020; Volume 12261, pp. 179–190.
39. Bae, S.; An, C.; Ahn, S.S.; Kim, H.; Han, K.; Kim, S.W.; Park, J.E.; Kim, H.S.; Lee, S.-K. Robust performance of deep learning for distinguishing glioblastoma from single brain metastasis using radiomic features: Model development and validation. *Sci. Rep.* **2020**, *10*, 1–10. [[CrossRef](#)] [[PubMed](#)]
40. Guo, C.; Ferreira, D.; Fink, K.; Westman, E.; Granberg, T. Repeatability and reproducibility of FreeSurfer, FSL-SIENAX and SPM brain volumetric measurements and the effect of lesion filling in multiple sclerosis. *Eur. Radiol.* **2019**, *29*, 1355–1364. [[CrossRef](#)]
41. Kazemi, K.; Noorizadeh, N. Quantitative Comparison of SPM, FSL, and Brainsuite for Brain MR Image Segmentation. *J. Biomed. Phys. Eng.* **2014**, *4*, 13–26.
42. Meyer, S.; Mueller, K.; Stuke, K.; Bisenius, S.; Diehl-Schmid, J.; Jessen, F.; Kassubek, J.; Kornhuber, J.; Ludolph, A.C.; Prudlo, J.; et al. Predicting behavioral variant frontotemporal dementia with pattern classification in multi-center structural MRI data. *NeuroImage Clin.* **2017**, *14*, 656–662. [[CrossRef](#)] [[PubMed](#)]
43. Tustison, N.J.; Cook, P.A.; Klein, A.; Song, G.; Das, S.R.; Duda, J.T.; Kandel, B.M.; van Strien, N.; Stone, J.R.; Gee, J.C.; et al. Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements. *NeuroImage* **2014**, *99*, 166–179. [[CrossRef](#)]
44. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [[CrossRef](#)]
45. Ceccarelli, M.; Barthel, F.P.; Malta, T.M.; Sabedot, T.S.; Salama, S.R.; Murray, B.A.; Morozova, O.; Newton, Y.; Radenbaugh, A.; Pagnotta, S.M.; et al. Molecular Profiling Reveals Biologi-cally Discrete Subsets and Pathways of Progression in Diffuse Glioma. *Cell* **2016**, *164*, 550–563. [[CrossRef](#)]
46. Schindelin, J.; Arganda-Carreras, I.; Frise, E.; Kaynig, V.; Longair, M.; Pietzsch, T.; Preibisch, S.; Rueden, C.; Saalfeld, S.; Schmid, B.; et al. Fiji: An open-source platform for biological-image analysis. *Nat. Methods* **2012**, *9*, 676–682. [[CrossRef](#)]
47. Friston, K.J. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*; Elsevier: Amsterdam, The Netherlands, 2011.
48. Jenkinson, M.; Beckmann, C.F.; Behrens, T.E.; Woolrich, M.W.; Smith, S. FSL. *NeuroImage* **2012**, *62*, 782–790. [[CrossRef](#)]

49. Avants, B.B.; Tustison, N.; Song, G. Advanced Normalization Tools (ANTs): V1.0. *Insight J.* **2009**, *2*, 1–35.
50. van der Maaten, L.J.P.; Hinton, G.E. Visualizing Data Using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
51. Nitta, M.; Muragaki, Y.; Maruyama, T.; Iseki, H.; Ikuta, S.; Konishi, Y.; Saito, T.; Tamura, M.; Chernov, M.; Watanabe, A.; et al. Updated Therapeutic Strategy for Adult Low-Grade Glioma Stratified by Resection and Tumor Subtype. *Neurol. Med.-Chir.* **2013**, *53*, 447–454. [[CrossRef](#)] [[PubMed](#)]
52. Zanfardino, M.; Pane, K.; Mirabelli, P.; Salvatore, M.; Franzese, M. TCGA-TCIA Impact on Radiogenomics Cancer Research: A Systematic Review. *Int. J. Mol. Sci.* **2019**, *20*, 6033. [[CrossRef](#)] [[PubMed](#)]
53. Hamamoto, R.; Komatsu, M.; Takasawa, K.; Asada, K.; Kaneko, S. Epigenetics Analysis and Integrated Analysis of Multi-omics Data, Including Epigenetic Data, Using Artificial Intelligence in the Era of Precision Medicine. *Biomolecules* **2019**, *10*, 62. [[CrossRef](#)] [[PubMed](#)]
54. Sanai, N.; Berger, M.S. Glioma Extent of Resection and Its Impact on Patient Outcome. *Neurosurgery* **2008**, *62*, 753–766. [[CrossRef](#)] [[PubMed](#)]
55. Alattar, A.A.; Brandel, M.G.; Hirshman, B.R.; Dong, X.; Carroll, K.T.; Ali, M.A.; Carter, B.S.; Chen, C.C. Oligodendroglioma resection: A Surveillance, Epidemiology, and End Results (SEER) analysis. *J. Neurosurg.* **2018**, *128*, 1076–1083. [[CrossRef](#)] [[PubMed](#)]
56. Chow, D.; Chang, P.; Weinberg, B.D.; Bota, D.A.; Grinband, J.; Filippi, C.G. Imaging Genetic Heterogeneity in Glioblastoma and Other Glial Tumors: Review of Current Methods and Future Directions. *Am. J. Roentgenol.* **2018**, *210*, 30–38. [[CrossRef](#)] [[PubMed](#)]