
MR Intensity Normalization Methods Impact Sequence Specific Radiomics Prognostic Model Performance in Primary and Recurrent High-Grade Glioma

[Patrick Salome](#)*, Francesco Sforazzini, Gianluca Grugnara, [Andreas Kudak](#), [Matthias Dostal](#),
[Christel Herold-Mende](#), [Sabine Heiland](#), Jürgen Debus, [Amir Abdollahi](#), [Maximilian Knoll](#)*

Posted Date: 10 January 2023

doi: 10.20944/preprints202301.0165.v1

Keywords: Multiparametric MRI; image preprocessing; intensity harmonization; intensity standardization;
high-grade glioma; radiomics signatures



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

MR Intensity Normalization Methods Impact Sequence Specific Radiomics Prognostic Model Performance in Primary and Recurrent High-Grade Glioma

Patrick Salome^{1,2,3,6,*}, Francesco Sforazzini^{1,2,3}, Gianluca Grugnara⁴, Andreas Kudak^{5,6,7}, Matthias Dostal^{5,6,7}, Christel Herold-Mende^{8,9}, Sabine Heiland⁴, Jürgen Debus^{3,5,6}, Amir Abdollahi^{1,3,5,6} and Maximilian Knoll^{1,3,5,6,*}

¹ Clinical Cooperation Unit Radiation Oncology, German Cancer Research Center, Heidelberg, Germany

² Heidelberg Medical Faculty, Heidelberg University, Heidelberg, Germany

³ German Cancer Consortium (DKTK) Core Center Heidelberg, Heidelberg, Germany

⁴ Department of Neuroradiology, Heidelberg University Hospital, Heidelberg, Germany

⁵ Department of Radiation Oncology, Heidelberg University Hospital, INF 400, 69120, Heidelberg, Germany

⁶ Heidelberg Ion-Beam Therapy Center (HIT), Im Neuenheimer Feld 450, 69120, Heidelberg, Germany

⁷ CCU Radiation Therapy, German Cancer Research Center (DKFZ), INF 280, 69120, Heidelberg, Germany

⁸ Brain Tumour Group, European Organization for Research and Treatment of Cancer, Brussels, Belgium

⁹ Division of Neurosurgical Research, Department of Neurosurgery, University of Heidelberg, Germany

* Correspondence: p.salome@dkfz.de (P.S.); m.knoll@dkfz.de (M.K.)

Simple Summary: As magnetic resonance (MR) intensities are acquired in arbitrary units, scans from different scanners are not directly comparable, thus intensity normalization is essential. In this study, we assess the impact of normalization methods on prognostic radiomics models in primary and recurrent high-grade glioma on different MR sequences. Furthermore, we present a methodology which allows for the handling of radiomics performance discrepancy due to MR intensity normalization.

Abstract: Purpose: This study investigates the impact of different intensity normalization (IN) methods on the overall survival (OS) radiomics models' performance of MR sequences in primary (pHGG) and recurrent high-grade glioma (rHGG). **Methods:** MR scans acquired before radiotherapy were retrieved from two independent cohorts (rHGG C1: 197, pHGG C2: 141) from multiple scanners (15, 14). The sequences are T1 weighted (w), contrast-enhanced T1w (T1wce), T2w, and T2w-FLAIR. Sequence-specific significant features (SF) associated with OS, extracted from the tumour volume, were derived after applying 15 different IN methods. Survival analyses were conducted using Cox proportional hazard (CPH) and Poisson regression (POI) models. A ranking score was assigned based on the 10-fold cross-validated (CV) concordance index (C-I), mean square error (MSE), and the Akaike information criterion (AICs), to evaluate the methods' performance.

Results: Scatter plots of the 10-CV C-I and MSE against the AIC showed an impact on the survival predictions between the IN methods and MR sequences (C1/C2 C-I range: 0.62-0.71/0.61-0.72, MSE range: 0.20-0.42/0.13-0.22). White stripe showed stable results for T1wce (C1/C2 C-I: 0.71/0.65, MSE: 0.21/0.14). Combat (0.68/0.62, 0.22/0.15) and histogram matching (HM, 0.67/0.64, 0.22/0.15) showed consistent prediction results for T2w models. They were also the top-performing methods for T1w in C2 (Combat: 0.67, 0.13; HM: 0.67, 0.13), however, only HM achieved high predictions in C1 (0.66, 0.22). After eliminating IN impacted SF using Spearman's rank-order correlation coefficient, a mean decrease in the C-I and MSE of 0.05 and 0.03 was observed in all four sequences. **Conclusion:** The IN method impacted the predictive power of survival models. Thus, performance is sequence-dependent.

Keywords: multiparametric MRI; image preprocessing; intensity harmonization; intensity standardization; high-grade glioma; radiomics signatures

1. Introduction

Radiomics, the extraction of features from medical images using data characterization algorithms, is an upcoming field of research expected to yield non-invasive surrogates for important molecular characteristics [1]. With the superior soft-tissue contrast of Magnetic Resonance Imaging (MRI), which allows for increased sensitivity and specificity in disease detection [2], several MR-based radiomics studies have been recently performed to determine image biomarkers that will help to assess prognosis and lead to better treatment decisions [3–6]. Robust radiomics models often require large amounts of data; therefore, MR images are usually collected from multiple centers, sites, and scanners [7]. MR intensities are acquired in scanner-dependent arbitrary units, which leads to scans coming from different scanners and subjects not directly comparable, even when the same scanning protocol is implemented [8,9]. While this intensity variation has no major effects on the clinical diagnosis, it impacts drastically the performance of subsequent MRI preprocessing, such as image registration, segmentation and, subsequently, radiomics features calculation [10–12]. Therefore, intensity normalization methods should be implemented to deal with the intra- and inter-scan image intensity variations. This task has nonetheless been proven to be non-trivial, as speed, accuracy, and applicability can vary depending on the given data set. Multiple intensity normalization algorithms have been developed over time [13–18]. However, even though the image biomarker standardization initiative (IBSI) has defined a more general standardized radiomics image processing workflow, no specific guidelines on the proper choice of intensity normalization methods are currently present [19].

In the context of MR-based radiomics and deep learning-based survival prediction in high-grade glioma patients, out of 23 publications, the methods identified are z-score transformation or similar (30%), histogram-matching [13] (27%), the MR brain-specific white strip [20] (18%), tissue-based [21] (3%), and feature-based [22] (1%) [23–47]. Intensity normalization was not reported in the remaining studies (Supplementary Table S1). Furthermore, a single intensity normalization method was applied to all MR sequences in study. As the problem of reproducibility becomes more evident with the increase of published radiomics studies [48,49], more effort has been put into the detailed investigations of the effects of different imaging preprocessing steps on the overall robustness and reproducibility of the radiomics models [50–52]. Most of these investigations have targeted grey-level discretization, i.e. the clustering of pixels based on intensity values to reduce feature calculation time and noise. However, as demonstrated by Carre et al., the impact of grey-level discretization on radiomics feature reproducibility is also correlated and thus influenced by choice of the normalization method [53]. Noting that radiomics features can be split into 4 different groups, i.e. shape-based, first-order, second-order [54] and higher-order statistics [55], the authors also report that z-score transformation of first and second-order radiomics features show robust results. In a systematic review on intensity normalization of MRI prior to radiomic feature extraction in glioma datasets, Fatania et al. report that intensity normalization is a key preprocessing step in the development of robust radiomics signatures and that few comparison studies of multiple methods exists [56]. In the context of radiomics survival model prediction, Um et al. demonstrated that the normalization of MR intensity through histogram matching before survival prediction modelling in primary glioblastoma (GBM) patient cohorts helped reduce radiomics feature variability in T1 weighted (w), T1w post-contrast agent (T1wce), and fluid-attenuated inversion recovery (FLAIR) and lead to better patient stratification. The radiomics signature consisted of a combination of all 3 radiomics feature groups [57]. Li et al. performed an in vitro and in vivo study by applying 7 different normalization methods on T1w images. They demonstrated that the feature-based harmonization method Combat [22] significantly removes scanner effects in brain MR-based radiomics studies [58].

Our study builds upon the studies mentioned above and aims to investigate the use of different normalization algorithms in multi-scanner brain MRI datasets, and more precisely, the performance of different methods on different sequences and their impact on the survival prediction model's performance, by analyzing the predictive power of the methods' respective normalized dataset in the prediction of overall survival (OS). Noting that the radiomics survival prediction signatures identified in previous work included all radiomics feature groups [23,25–27,29–47,59], our analyses were directly performed on the signature obtained through 2 multi scanners high-grade glioma (HGG) datasets, i.e. a recurrent HGG (rHGG) cohort of 197 patients (C1), and a primary HGG (pHGG) cohort of 141 patients (C2), after the application of stringent features significant selection pipeline. The MR sequences considered are T1w, T1wce, T2w, and FLAIR. Significant OS correlated features were first identified through multiple feature reduction and resampling techniques from MR images acquired pre-radiotherapy (RT) and normalized through 15 different normalization approaches. Sequence-specific survival radiomics prediction models were next trained using Cox proportional hazard and Poisson survival regressions models and applied on both cohorts. The performances of the intensity normalization algorithms were then compared based on the predictive power of their respective normalized dataset in the predictions of OS. Finally, features affected by the intensity normalization methods were further rejected, and comparisons with models trained with the remaining significant features were performed.

2. Materials and Methods

2.1. Datasets

This study analyses multiparametric pre-RT MR sequences from two independent data cohorts. Eligibility of patients was based on the availability of clinical information, of at least two MR sequences taken no longer than 30 days before RT, and of the RT-DICOM data specifically the DICOM structure set (SS). The first cohort (C1) consisted of 197 patients with pathologically confirmed rHGG collected retrospectively from 15 different MR scanners at the Heidelberg Ion-Beam Therapy Center (HIT) and University Clinic Heidelberg (UKHD) from 2009 to 2018. All 197 patients received carbon irradiation. The second cohort (C2) consists of 141 pHGG patients collected retrospectively from 14 different MR scanners at the UKHD from 2011 to 2016. All 141 patient received standard photon RT. Patients between both cohorts were matched for the frequency in gender, tumor grade and MR sequence. OS was calculated as the number of days between the start of the re-RT (C1) or RT(C2) and death. MR scans were acquired post-surgical tumor resection and prior to radiotherapy treatment (RT). Conventional multislice (2D) acquired in the axial, sagittal, or coronal plane and 3D scans are present. The MR sequences found in the cohorts are the widely used sequences for brain tumor imaging [60] in clinical routines and trials [61,62]. However, the classes considered in this study are T1w, T1wce, T2w, and FLAIR. The four considered sequences were identified at different rates in both cohorts. The in-plane resolution ranged from 0.45 x 0.45 to 1.40 x 1.40 mm in the discovery cohort, and 0.33 x 0.33 to 2 x 2 mm in the test set. Slice thickness ranged from 0.9 to 5 mm in all MR scans. A summary of both cohorts is shown in Table 1. Overview of the MR scanners and protocols found are reported in Supplementary Tables S2 and S3.

Table 1. Descriptions of the data cohorts C1 and C2 considered in this study.

	C1		C2	
	n	%	n	%
Patients	197	100	141	100
Gender				
Male	120	61	86	61
Female	77	39	55	39
Age				

<50	84	64	47	33
50-69	105	53	73	52
>=70	8	17	21	15
Tumor grade				
III	71	36	34	24
IV	126	64	65	46
MR sequence				
T1wce	197	100	141	100
T1w	186	94	135	96
T2w-FLAIR	168	85	118	83
T2w	141	71	100	71

2.2. MRI Preprocessing Workflow

Dicom dataset curation and MR image classification were performed using pyCuRT and MR-Class [63]. All images were first reoriented to a common orientation. T1w images were corrected for signal inhomogeneities using the N4 bias field correction algorithm [64]. Brain extraction with the HD-BET brain extraction tool was next performed [65]. When available, 3D MR sequences were mainly selected. Motion correction and volumetric image reconstruction were performed when 2D transversal, sagittal and coronal MR scans were present. Reconstruction of the low-resolution 2D slices to a high-resolution 3D MR was performed using NiftyMic [66]. Next, cross-sectional linear co-registrations with 6 degrees of freedom (DOF) of the present MR images were performed on the T1wce using advanced normalization tools (ANTs) [67]. Furthermore, cross-sectional linear co-registrations with 6 DOF of the T1wce was performed on the RT planning CT. This registration was solely to generate the MR to CT transformation matrix, used to bring the target volume (TV) segmentations extracted from the DICOM SS objects to the MR space. Next, intensity normalization was performed. The different intensity normalization methods implemented in the comparison study are described in the next section. All images and segmentations were then re-sampled to a matrix size of 2x2 mm and a slice thickness of 2 mm using a cubic spline and linear interpolation, respectively. As for image discretization in an attempt to neutralize the impact of grey-level discretization on the overall result, five different bin counts were implemented, resulting in five sets of features per normalization algorithm. A bin count discretization approach was implemented since it was more frequently seen in HGG radiomics survival prediction studies. The image preprocessing diagram is shown in Figure 1.

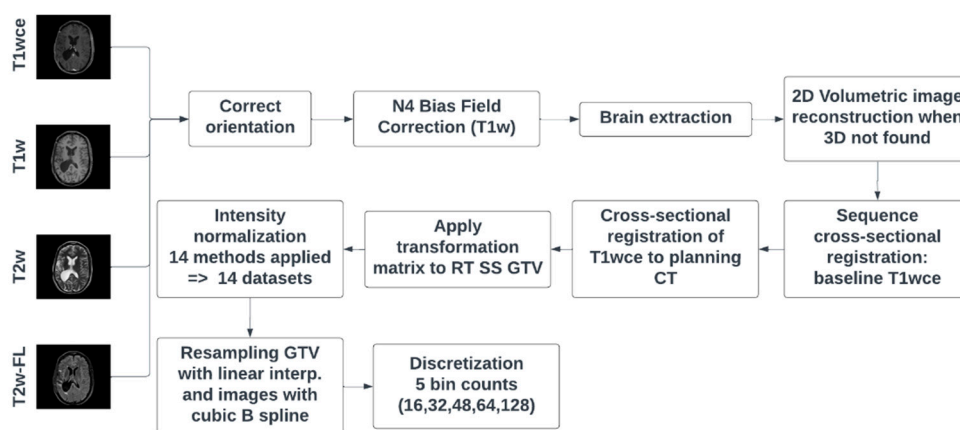


Figure 1. MR preprocessing diagram applied on both cohorts. RT SS GTV represents the gross tumor volume segmentation extracted from the DICOM RT structure set. T2w-FL: T2w-FLAIR.

2.3. Intensity Normalization Methods

Intensity normalization was performed with the help of the intensity normalization package by Reinhold et al. [68] and the FMRIB's Automated Segmentation Tool (FAST) [69]. The intensity-normalization methods considered are: Fuzzy C-Means (FCM) [21] (9 different masks combinations), kernel density estimation (KDE), Gaussian mixture models (GMM) [70], the Nyul's and Udupa's histogram matching-based abbreviated in this study as HM [13], white-strips (WS) [20], z-score normalization, and the feature-based batch adjustment method, i.e. Combat [22], resulting in 15 different MRI normalized datasets. A brief description of the methods is given in this section. For a broader description, we refer to the original normalization method papers as well as the for the manuscript by Reinhold et al. [68].

2.3.1. Standard Score

The standard score, also known as the z-score, represents the distance of a raw score from the mean measured in standard deviations. In MR brain image normalization, given that B is the brain mask in image I, the z-score calculates the mean μ and standard deviation σ of the intensities inside the brain image (excluding the background) as follows: $\mu = \frac{1}{|B|} * \sum_{b \in B} I(b)$, $\sigma = \sqrt{\frac{\sum_{b \in B} (I(b) - \mu)^2}{|B| - 1}}$ with the normalized image being $I_{norm}(x) = \frac{I(x) - \mu}{\sigma}$. A disadvantage of this method is that the high intensities in the images are usually attenuated, risking a loss of information.

2.3.2. Fuzzy Clustering

Clustering is a method for analysing data that aims to discover structures or groups in a data set. Fuzzy clustering allows a piece of data to be part of more than one cluster. In a fuzzy c-means algorithm, a data point is assigned a membership function with 0 being the farthest from a cluster's center and 1 being the closest to a cluster's center, with the data point theoretically being able to belong to all clusters. Used as a normalization technique in brain MRI, the fuzzy c-means algorithm uses the segmentation of specific brain tissue to normalize the image to the mean intensity of the tissue. If the tissue mean is $\mu = \frac{1}{|T|} * \sum_{t \in T} I(t)$, then the normalized image is $I_{norm}(x) = \frac{I(x)}{\mu}$, where x is the image voxels, and T is the tissue mask. The brain tissue masks, i.e. white matter (wm), gray matter (gm) and cerebrospinal fluid (csf), segmentations were performed using FSL's FAST. In conjunction with the most common intensity value (mode) in a particular image, nine different mask combinations were implemented to generate nine fuzzy c-means normalized datasets. The masks are: csf, gm, wm, csf-gm, wm-csf, wm-gm, csf-mode, wm-mode and gm-mode. The normalization with two brain tissue masks is performed as: With $\mu_1 = \frac{1}{|T1|} * \sum_{t \in T1} I(t)$ and $\mu_2 = \frac{1}{|T2|} * \sum_{t \in T2} I(t)$ the normalized image is derived as $I_{norm}(x) = \frac{I(x) - a}{b - a}$ with $a = \min(\mu_1, \mu_2)$ and $b = \max(\mu_1, \mu_2)$ The normalization with a brain tissue mask and the mode is performed as: as $I_{norm}(x) = \frac{I(x)}{diff}$ with $diff = \mu_T - mode(B)$ with T as the tissue mask and B as the brain mask.

2.3.3. Kernel Density Estimation

A density estimator aims to find a function for the probability distribution from which a dataset is generated. The kernel density estimation (KDE) is an empirical calculation in a parametrized form. The formula for calculating the KDE for the probability distribution function is $p(x) = \frac{1}{N * M * L * h} * \sum_{i=1}^{N * M * L} K(\frac{x - x_i}{h})$, where N, M, and L are the sizes of the images, K is the kernel (normalized to one), and h is the bandwidth parameter which scales the kernel. This method provides a smoother version of the histogram, making it easier to find the maxima π , which is used to normalize the entire image

as $I_{norm}(x) = c * \frac{I(x)}{\pi}$, where c is a positive, real constant. For the MR brain images, the KED finds the peak of the white matter histogram and translates it to a standard value.

2.3.4. Mixture Models

A mixture model assumes that a data set comprises of subsets whose individual distributions are the respective probability distributions in the overall data set. A specific mixture model is the Gaussian mixture model, where the subsets are considered to be generated from a finite number of Gaussian distributions with undefined parameters. The method used in our study fits three Gaussian distributions to the histogram of the brain mask and normalizes the white matter mean to a standard value.

2.3.5. Landmark Based Histogram Matching

The landmark-based histogram matching method by Nyúl et al. deform the input image intensity histogram to match a reference histogram. The reference histogram is commonly obtained by averaging histograms in a data set and setting the landmarks of interest. Each input image histogram is then matched to the reference through linear interpolation, based on the defined landmarks, which are usually quantiles.

2.3.6. White Stripe Normalization

The white stripe normalization approach by Shinohara et al. normalizes an image based on the normal-appearing white matter (NAWM) [20]. The NAWM values are obtained through a smoothening of the image histogram, followed by selecting the largest peak μ . The so-called white stripe contains intensity values up to 10% around μ . The white stripe can be defined as $\Omega_{\tau} = \{I(x) | F^{-1}(F(\mu) - \tau) < I(x) < F^{-1}(F(\mu) + \tau)\}$, where $F(x)$ is the cumulative distribution function of the image I and $\tau = 5\%$. If σ is the standard deviation in the white stripe, the normalized image is $I_{norm}(x) = \frac{I(x) - \mu}{\sigma}$.

2.3.7. Combat

Combat is a feature-based method originally developed for microarray expression data [22]. However, it has also been applied in imaging data and radiomics studies in recent years [71–73]. It eliminates batch effects through a known batch covariate by using parametric or non-parametric empirical Bayes frameworks. In this study, an empirical bayes Combat method was applied through the *sva* (v 3.20.0) to eliminate batch effects due to the MR scanner. Adjustment of the following covariates was performed i.e. age, tumor grade and gender.

2.4. Comparison Study Design

After MR image preprocessing, radiomics features were calculated automatically from the gross tumor volume (GTV) segmentations extracted from the DICOM RT structure set and the original image, as well as from derived images from each normalized/discretized dataset using Pyradiomics (v 3.0) [74]. The derived images were retrieved from first Wavelet filtering, which yielded 8 decompositions per level, each representing a combination of either a high or a low pass filter in each of the three dimensions, and then by applying a Laplacian of Gaussian filter with spatial scaling factors (SSFs) of 2,3 and 4 mm. The total yielded features were around 1200 per MR sequence. The different feature classes and corresponding feature numbers can be seen in Table 2.

Table 2. Number of shape, first and second-order statistics derived per sequence and calculated on both the original and derived images.

Class	No. features
First-order statistics	19
Shape-based (3D)	16

Second-order statistics	
Gray Level Co-occurrence Matrix	24
Gray Level Run Length Matrix	16
Gray Level Size Zone Matrix	16
Neighbouring Gray Tone Difference Matrix	5
Gray Level Dependence Matrix	14

A Spearman rank-order correlation coefficient was used on the total number of features to exclude redundant features ($r_s > 0.80$). Three feature selection methods, including a univariate analysis under Cox proportional hazard (CPH) models ($P < 0.05$), a random forest (RF) -based method and lasso regression, were applied on 1000 random subsamples of C1 and C2 (10% left out) separately to identify features correlated to OS. Sequence-specific significant features identified at least 950 times were selected and survival analyses were conducted using CPH [75] and Poisson survival regressions (POI) models [76]. A ranking score was next assigned to each normalization approach based on the converted standardized z score of the CPH averaged 10-fold cross-validated (CV) concordance index (C-I), the POI averaged 10-fold CV mean square error (mse), and the respective Akaike information criterion (AIC) of the OS prediction models. Lastly, after identifying the top-ranked methods for the different MR sequences, correlation heatmaps between the different normalization approaches for each significant feature forming the sequence-specific radiomics signature were plotted. Stable features that showed high correlation ($r_s > 0.80$) between at least 12 intensity normalization methods were further used to train again CPH and POI models and the effects on the model predictions were studied. Finally, the performance of the feature-based method Combat was assessed in combination with the top-ranked image-based normalization method for each sequence in both cohorts. A flowchart of the study design is shown in Figure 2.

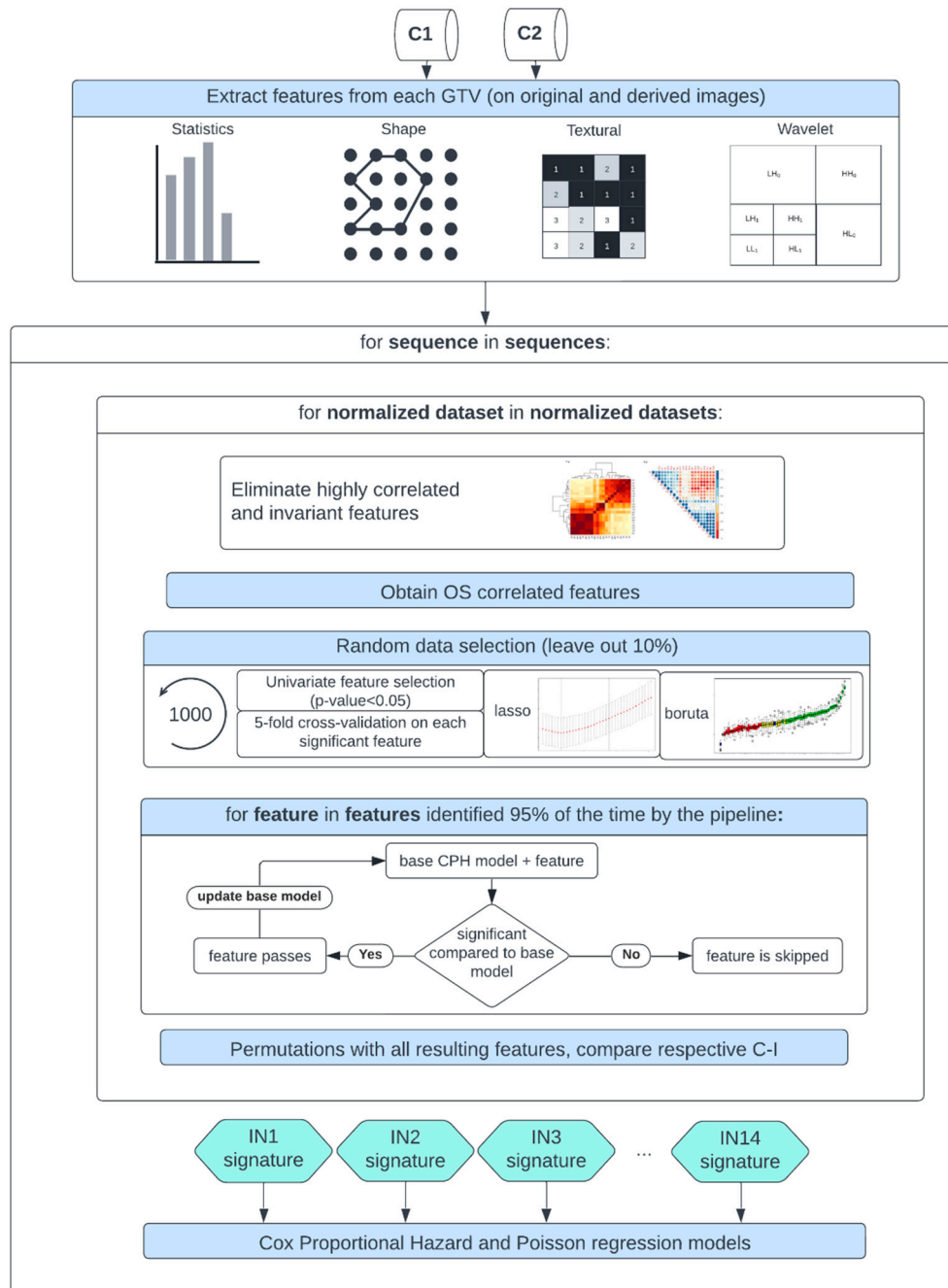


Figure 2. Study workflow – Following MR image preprocessing on cohorts C1 and C2, features were extracted from each normalized dataset, intensity normalization method-specific radiomics signatures were derived, and Cox proportional hazards and Poisson regression models were trained.

3. Results

3.1. Performance Assessment of the Intensity Normalization Method-Specific Survival Prediction Models for the Different MR Sequence

Scatter plots of the CPH averaged (over the five bin counts investigated) C-index and POI averaged mse, plotted against the respective AIC, for the 15 different intensity normalization specific OS models derived from cohort C1 and C2 are shown in Figures 3 and 4. The OS model derived by the non-normalized (nn) dataset is also plotted.

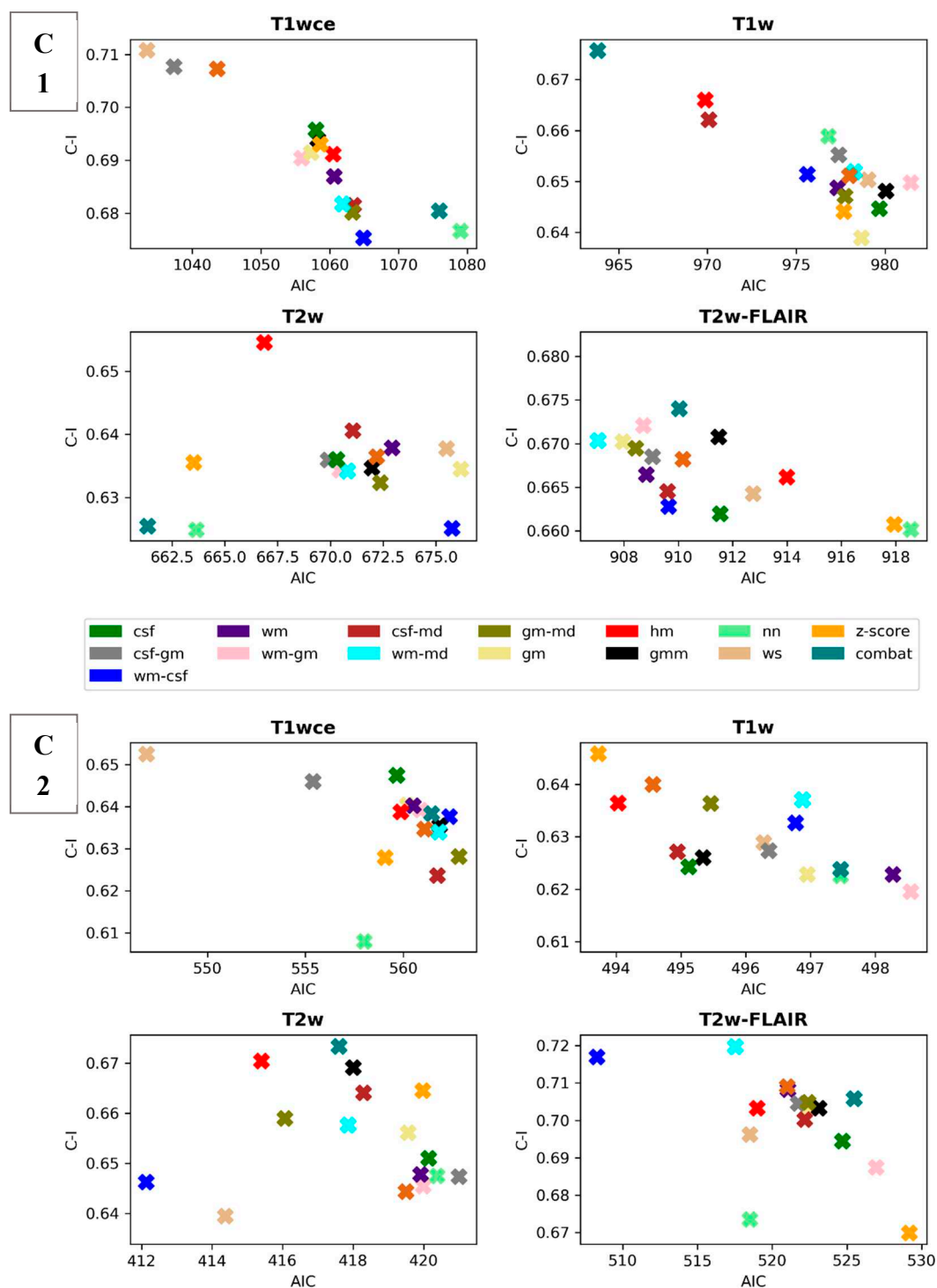


Figure 3. Scatter plots of the averaged (over the 5 bin counts considered) C-index vs AIC obtained by the CPH models for all 4 sequences in study. Upper panel: cohort C1, Lower panel: cohort C2. csf: cerebrospinal fluid, wm: white matter, gm: grey matter, md: mode, gmm: Gaussian mixture models, kde: kernel density estimation, hm: Nyúl/Udupa histogram matching, ws: white stripe, nn: no normalization.

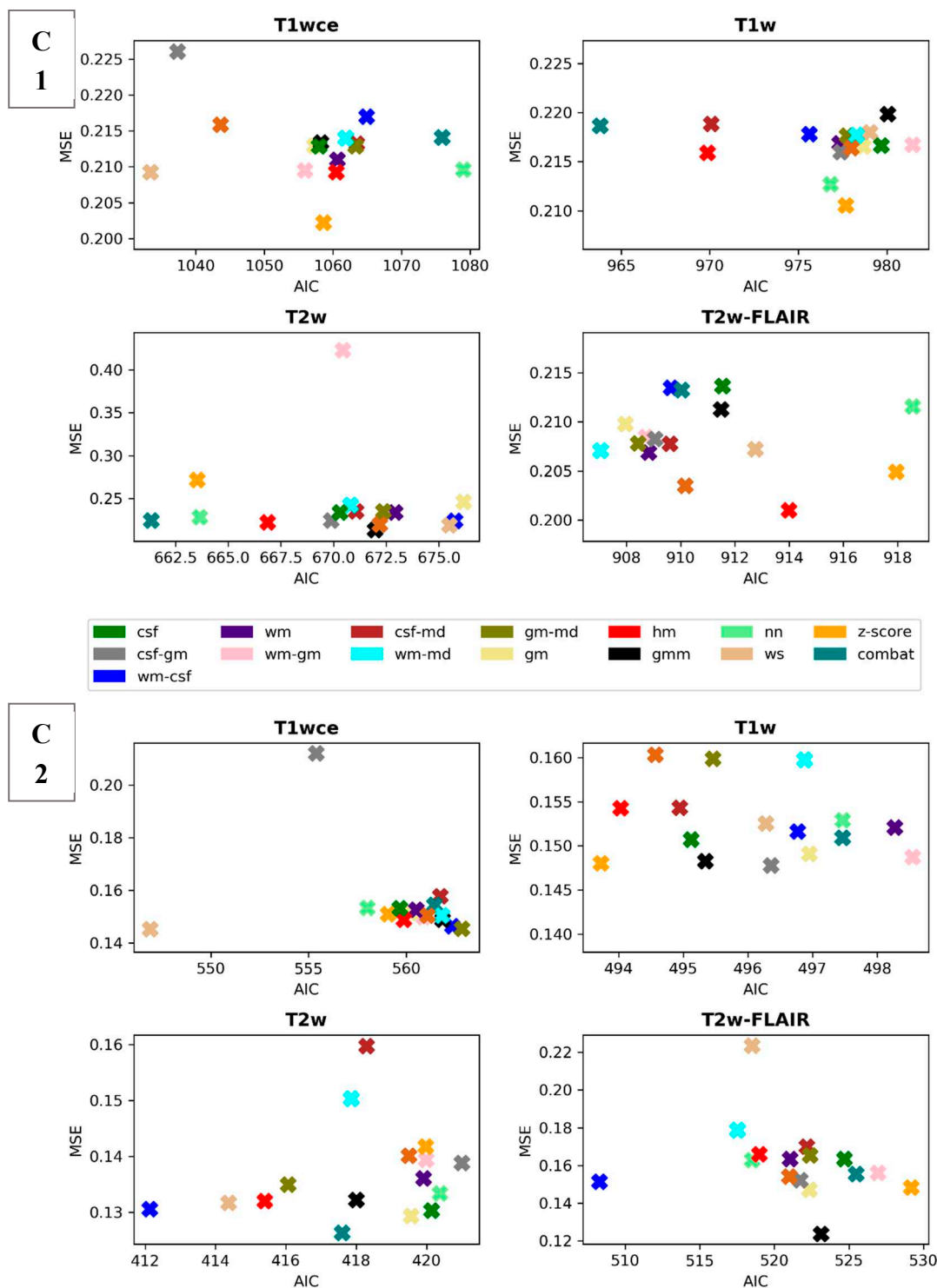


Figure 4. Scatter plots of the averaged (over the 5 bin counts considered) mse vs AIC obtained by the POI models for all 4 sequences in study. Upper panel: cohort C1, Lower panel: cohort C2. csf: cerebrospinal fluid, wm: white matter, gm: grey matter, md: mode, gmm: Gaussian mixture models, kde: kernel density estimation, hm: Nyúl/Udupa histogram matching, ws: white stripe, nn: no normalization.

Table 3 summarizes and ranks the performance scores of the intensity normalization methods for each of the four MR sequences considered in both cohorts, respectively.

Table 3. Ranking with scores of the intensity normalizations of the MR sequences for cohort C1 and C2. T2w-FL: T2w-FLAIR. Norm. method: Intensity normalization method.

C1	T1wce		T1w		T2w		T2w-FL	
	Norm. Method	Score	Norm. Method	Score	Norm. Method	Score	Norm. Method	Score
1	ws	0.71	combat	0.13	hm	0.27	wm-md	0.02
2	kde	-0.13	hm	-0.28	combat	-0.03	wm-gm	-0.11
3	csf-gm	-0.20	csf-md	-0.90	z-score	-0.28	kde	-0.13
4	z-score	-0.48	nn	-1.00	gmm	-0.38	gm-md	-0.23
5	wm-gm	-0.85	z-score	-1.14	csf-gm	-0.61	gm	-0.24
6	csf	-0.97	csf-gm	-1.58	kde	-0.71	wm	-0.42
7	hm	-1.04	wm-csf	-1.65	nn	-0.76	csf-gm	-0.46
8	gmm	-1.11	wm	-1.85	csf	-0.78	combat	-0.77
9	gm	-1.13	kde	-1.88	wm-md	-0.80	csf-md	-0.77
10	wm	-1.24	wm-md	-1.95	gm-md	-0.96	hm	-0.80
11	wm-md	-1.67	gm-md	-2.05	csf-md	-1.09	wm-csf	-1.01
12	csf-md	-1.71	ws	-2.15	ws	-1.18	gmm	-1.02
13	gm-md	-1.72	csf	-2.16	wm	-1.22	ws	-1.29
14	wm-csf	-2.16	gm	-2.23	wm-gm	-1.72	csf	-1.75
15	combat	-2.25	wm-gm	-2.37	gm	-1.79	z-score	-2.21
16	nn	-2.27	gmm	-2.48	wm-csf	-2.01	nn	-2.65
C2								
1	ws	1.00	z-score	0.64	combat	0.07	wm-csf	0.66
2	csf	-0.54	hm	-0.11	hm	-0.09	wm-md	-0.32
3	hm	-0.73	csf	-0.34	gm-md	-0.21	gmm	-0.56
4	z-score	-0.76	gmm	-0.35	wm-csf	-0.24	kde	-0.63
5	gm	-0.77	csf-md	-0.81	gmm	-0.41	csf-gm	-0.71
6	wm	-0.87	kde	-0.93	wm-md	-0.78	wm	-0.72
7	wm-gm	-0.87	gm-md	-0.97	gm	-1.00	gm	-0.76
8	csf-gm	-0.96	csf-gm	-0.97	csf-md	-1.12	hm	-0.81
9	kde	-0.98	ws	-1.04	ws	-1.13	gm-md	-0.90
10	wm-csf	-1.07	gm	-1.18	z-score	-1.21	csf-md	-1.05
11	gmm	-1.10	combat	-1.20	csf	-1.31	nn	-1.25
12	wm-md	-1.13	nn	-1.41	kde	-1.36	csf	-1.35
13	combat	-1.19	wm-csf	-1.43	wm	-1.52	combat	-1.42
14	gm-md	-1.28	wm-md	-1.64	nn	-1.60	ws	-1.50
15	csf-md	-1.39	wm-gm	-2.01	wm-gm	-1.69	wm-gm	-1.59
16	nn	-1.82	wm	-2.11	csf-gm	-1.81	z-score	-2.11

The white stripe method is ranked first for T1wce in both cohorts (C1/C2 10-fold CV C-I: 0.71/0.65, AIC: 1033/547, 10-CV mse: 0.21/0.14, AIC: 410/252). For T1w, the feature-based batch adjustment method, i.e. Combat had the best performance in C1 (0.68, 964, 0.22), while z-score

with a bin count of 32. Features with a high correlation ($r_s > 0.8$) between at least 12 methods were further selected for modelling. Wv: Wavelet filter transformation, or: original image, log 3: Laplacian of gaussian transformation with a sigma of 3 mm.

The 10-CV C-I and mse of the CPH and POI models with only the stable features that have a high correlation ($r_s > 0.8$) between at least 12 methods are reported in Table 4. Boxplots of the difference for each modality in both cohorts is shown in Figure 6.

Table 4. Performance of the of the top ranked image normalization method separate before and after the elimination of the intensity normalization impacted significant feature for cohort C1 and C2 for each MR sequence. The average (across all bin counts) 10-CV C-I/MSE with the 95% confidence intervals are reported.

	C1		C2	
	Before	After	Before	After
T1wce	0.71 [0.69 0.74] / 0.21 [0.19 0.23]	0.65 [0.63 0.69] / 0.23 [0.21 0.25]	0.65 [0.62 0.67] / 0.15 [0.13 0.17]	0.62 [0.60 0.65] / 0.19 [0.17 0.21]
T1w	0.68 [0.64 0.70] / 0.22 [0.20 0.25]	0.63 [0.61 0.67] / 0.24 [0.22 0.26]	0.65 [0.61 0.69] / 0.15 [0.12 0.18]	0.62 [0.58 0.65] / 0.18 [0.15 0.20]
T2w	0.65 [0.62 0.67] / 0.22 [0.19 0.25]	0.63 [0.60 0.67] / 0.25 [0.22 0.28]	0.67 [0.64 0.69] / 0.13 [0.11 0.17]	0.60 [0.58 0.65] / 0.16 [0.14 0.20]
T2w-FL	0.67 [0.64 0.69] / 0.20 [0.18 0.23]	0.62 [0.59 0.67] / 0.23 [0.21 0.25]	0.72 [0.65 0.76] / 0.18 [0.15 0.21]	0.66 [0.64 0.69] / 0.20 [0.17 0.22]

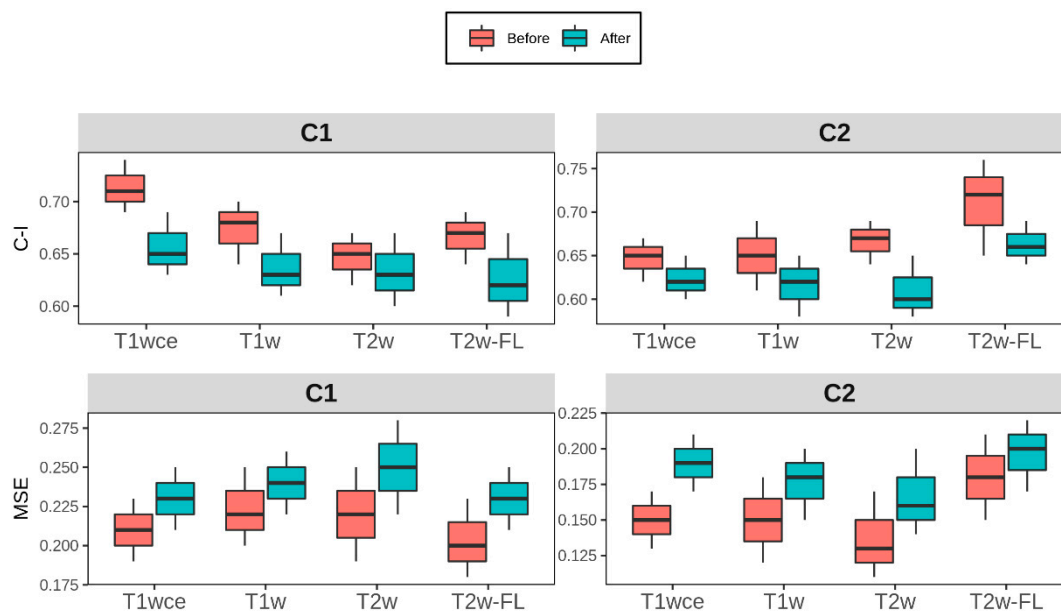


Figure 6. Box plots of the top-ranked image normalization method evaluation metrics C-I and MSE before and after the elimination of the intensity normalization impacted significant features for cohorts C1 and C2 for each MR sequence. The average (across all bin counts) 10-CV C-I/MSE with the 95% confidence intervals are plotted. Performance of both models was similarly affected after the elimination of the intensity normalization impacted significant features, with a mean decrease in the 10-CV C-I and 10-CV MSE of 0.05 and 0.03 in all four sequences across both cohorts C-I: Concordance-index, mse: mean squared error, T2w-FL:T2w-FLAIR.

3.3. Performance Comparison of the Feature-Based and Top Ranked Image-Based Normalisation Methods

Table 5 summarizes the performance of the top-ranked image normalization method separate and in combination with the feature-based method Combat for cohorts C1 and C2. Since Combat ranked first for the T1w models from C1 and T2w models from C2, the second-ranked method, i.e. the HM method, was the image-based intensity normalization method for these two datasets.

Table 5. Performance of the of the top ranked image normalization method separate and in combination with the feature-based method Combat for cohort C1 and C2 for each MR sequence. The average (across all bin counts) 10-CV C-I/MSE with the 95% confidence intervals are reported.

	C1			C2		
	Combat	I. norm.	Combined	Combat	I. norm.	Combined
T1wce	0.68 [0.66 0.70] /	0.71 [0.690.74] /	0.68 [0.66 0.69] /	0.64 [0.62 0.68] /	0.65 [0.62 0.67] /	0.63 [0.61 0.66] /
	0.21 [0.19 0.23]	0.21 [0.19 0.23]	0.21 [0.19 0.23]	0.15 [0.13 0.17]	0.15 [0.13 0.17]	0.17 [0.15 0.19]
T1w	0.68 [0.64 0.70] /	0.66 [0.64 0.68] /	0.62 [0.59 0.64] /	0.62 [0.60 0.66] /	0.65 [0.61 0.69] /	0.62 [0.59 0.65] /
	0.22 [0.20 0.24]	0.22 [0.19 0.24]	0.23 [0.20 0.26]	0.15 [0.12 0.17]	0.15 [0.12 0.18]	0.15 [0.11 0.16]
T2w	0.62 [0.59 0.64] /	0.65 [0.62 0.67] /	0.61 [0.58 0.63] /	0.67 [0.64 0.69] /	0.67 [0.64 0.69] /	0.62 [0.59 0.65] /
	0.23 [0.21 0.23]	0.22 [0.19 0.25]	0.25 [0.23 0.27]	0.13 [0.11 0.17]	0.13 [0.11 0.15]	0.15 [0.13 0.19]
T2w-FL	0.67 [0.64 0.69] /	0.67 [0.64 0.69] /	0.64 [0.61 0.66] /	0.70 [0.67 0.72] /	0.72 [0.65 0.76] /	0.68 [0.65 0.70] /
	0.21 [0.19 0.24]	0.20 [0.18 0.23]	0.24 [0.22 0.26]	0.16 [0.14 0.19]	0.14 [0.12 0.17]	0.17 [0.15 0.21]

4. Discussion

This study evaluated the impact of MRI intensity normalization algorithms on MR-based radiomics survival prediction models in primary and recurrent high-grade glioma. The sequences considered are T1w, T1wce, T2w, and T2w-FLAIR. Performance assessment of the intensity normalization method-specific CPH and POI survival prediction models showed an impact on the survival predictions between the different intensity normalization methods and the different MR sequences. Therefore, it can be concluded that the MR intensity normalization approach directly impacts the overall power of the radiomics based MR predictive models. Moreover, considering the variability of the acquired results for the different MR sequences, it can be seen that the intensity normalization algorithm performance is correlated with the MR sequence and that the problem cannot be simplified to one intensity normalization method.

Due to these variations and for a better interpretation of the results, the ranking score was developed, from which it was observed that the WS method showed promising results in T1wce models as it was ranked first in two independent multi-scanner datasets. Combat and the HM method showed consistent prediction results between the two cohorts for T2w models. These two methods were also the two top-performing methods for T1w in C1, however only HM achieved high predictions in C2 and not Combat. Though, this might be due to the higher number of batches and the number of images per batch, as 22% of T1w images in C2 were missing, which renders batch effect removal more challenging. As for T2w-FLAIR, the FCM showed favorable results in both cohorts; however, with different mask combinations, including the wm and csf or wm and mode. A tighter intensity range is observed in T2w-FLAIR than the other sequences, as csf signals are attenuated. These results might indicate that a mask-based normalization approach might be more favorable when dealing with images with tighter intensity ranges. The application of both an image-based and feature-based normalisation method had little impact on the performance of the CPH and POI models. Exceptions were observed in the dataset where combat was ranked first, i.e T1w in C1 and T2w in C2.

As CPH models were part of the radiomics signature building pipeline, POI models were also trained to assess whether model performances were biased to CPH models. Comparably to CPH models, the impact of the intensity normalization methods was also observed in the POI models. Furthermore, the performance of both models was similarly affected after the elimination of the intensity normalization impacted significant features. A mean decrease in the 10-CV C-I and 10-CV MSE of 0.05 and 0.03 was observed in all four sequences across both cohorts. The use of a correlation

coefficient between different normalization methods as a feature robustness check results in a tradeoff between model stability and the risk of eliminating important imaging biomarkers.

This study included 2 independent HGG cohorts collected from a single university hospital, UKHD. However, since the data cohorts included data between 2008 and 2019, 19 different scanners from 3 vendors with a 0.5 to 3.0-Tesla range were identified. Noting that the OS models derived from the non-normalized datasets generally ranked low in both cohorts across the sequences, the application of intensity normalization has indeed improved the OS prediction in radiomics survival models, demonstrating that the need for intensity normalization is based on the number of scanners and image protocols identified in the cohorts and not necessarily only the number of centers. However, an exception is seen in the T1w dataset in C1. This might be because a high number of images in the T1w dataset from C1 were reconstructed using NiftyMic (as mostly 2D MR scans were present) and therefore preprocessed prior to applying the intensity normalization methods [66].

Since multiple MR scanners were found in both cohorts, where some have been withdrawn from clinical practice, the application of phantoms to assess the actual impact of the IN methods could not be performed. Therefore, the hard endpoint OS was used in this study as a possible appropriate surrogate.

In literature, multiple intensity normalization methods have been reported in HGG radiomics studies where all implemented the same method across all MR sequences [56]. However, as demonstrated in this study, the performance of the different methods varies. This study shows that the variations are big and that if reproducibility of the radiomics model is to be possible, the method of intensity normalization should be reported. Another way is to eliminate features impacted by the different normalization methods. When unstable features are impacted, the performance of the individual MR sequence prediction models is reduced, a necessary trade-off for stable radiomics models. However, combining multiple stable radiomics signatures from multiple MR sequences or modalities might mitigate that reduction and lead to high survival prediction models.

The following limitations exist in this work. The application of different preprocessing methods makes it generally hard to assess the impact of different normalization methods seamlessly. The changes in the radiomics values are as much affected by other preprocessing methods as image discretization or delineating the region of interest. This suggests that the application of intensity normalization alone may not be enough. In this study, we attempted to limit the effect of intensity discretization by applying 5 different bin counts and reporting the average score. Yet, as demonstrated by many radiomics robustness studies, the overall performance and reproducibility of the radiomics models are indeed affected by the choice of discretization approach [50–52]. Nevertheless, similarly to the use of correlation coefficient heatmaps between the different normalization methods to determine stable radiomics features, the same can be implemented across different bin counts or bin widths.

Moreover, differences in the performance of the different IN methods across both cohorts can be possibly due to the alterations in the structure of intra-tumor heterogeneity which differ between pHGG and rHGG, as well as the difference in the treatment of rHGG in comparison to pHGG, since the treatment of rHGG is not standardized as for pHGG, i.e incorporating surgery, adjuvant postoperative RT and adjuvant chemotherapy [77]. In addition, Heterogeneity of cohorts such as in MGMT methylation, IDH1/2 mutation and 1p/19q deletion can also lead to survival prediction differences [78]. More detailed studies are required to assess the impact in more stringently defined cohorts.

Furthermore, all GTVs were segmented following institutional guidelines for RT treatment. Nonetheless, delineation variabilities are known to impact radiomics features, and the impact of intensity normalization and ROI segmentation should also be evaluated. As automatic tumor segmentation networks become more robust and popular, these inter-observer variabilities would be reduced and thus eliminate another layer of variability. However, as segmentation networks are also impacted by the intensity normalization method, future work will evaluate the performance of different normalization methods on automatic segmentation networks.

5. Conclusions

Variations in the results for the different MR sequences showed that the intensity normalization method performance is sequence-dependent and that it directly impacts the predictive power of glioma survival models. Therefore, the documentation of the adapted normalization approach is necessary to enable the reproducibility of the MRI-based radiomics model. The methodology presented in this study can be further implemented to different entities to determine the stable radiomics features for signature building. Future work includes the study of additional sequences and anatomy sites.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org, Table S1: Intensity normalization algorithms applied in MR-based radiomics and deep learning-based survival prediction studies in high-grade glioma patients. Table S2: MR scanner models found in the cohorts; Table S3: MR image protocols found in the cohorts; Table S4: Model performance metrics for each MR sequence and normalization method for cohort 1; Table S5: Model performance metrics for each MR sequence and normalization method for cohort 2; Figure S1: Correlation heatmaps between the 16 different normalization methods considered and the reference non-normalized dataset for each MR sequence in cohorts C1 and C2 discretized with bin counts of 16, 32, 48, 64 and 128.

Author Contributions: Conceptualization, Patrick Salome, Francesco Sforazzini, Amir Abdollahi and Maximilian Knoll; Data curation, Patrick Salome and Francesco Sforazzini; Formal analysis, Patrick Salome, Francesco Sforazzini and Maximilian Knoll; Funding acquisition, Jürgen Debus and Amir Abdollahi; Investigation, Gianluca Grugnara, Christel Herold-Mende and Sabine Heiland; Methodology, Patrick Salome and Maximilian Knoll; Resources, Andreas Kudak and Matthias Dostal; Software, Patrick Salome and Francesco Sforazzini; Supervision, Jürgen Debus, Amir Abdollahi and Maximilian Knoll; Visualization, Patrick Salome; Writing – original draft, Patrick Salome; Writing – review & editing, Francesco Sforazzini, Gianluca Grugnara, Christel Herold-Mende, Sabine Heiland, Amir Abdollahi and Maximilian Knoll.

Funding: This study was found by the H2020 MSCA-ITN PREDICT project, Grant Number 766276 and intramural funds of the National Center for Tumor Diseases (NCT) and German Cancer Consortium (DKTK) Radiation Oncology programs.

Institutional Review Board Statement: Ethical approval: Statement: Ethical approval 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 Medical Faculty of Heidelberg University (approval number S-540/2010, date of last updated approval: 20 July 2020).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the S-540 study.

Data Availability Statement: C1 and C2 are available from the corresponding author on reasonable request.

Conflicts of Interest: P.S. No relevant relationships. F.S. No relevant relationships. A.K. No relevant relationships. N.B. No relevant relationships. J.D. Grants/contracts from/with Viewray, CRI – The Clinical Research Institute, Accuray International Sarl, RaySearch Laboratories, Vision RT, Merck Serono, Astellas Pharma, AstraZeneca, Siemens Healthcare, Solution Akademie, Ergomed PLC Surrey Research Park, Quintiles, Pharmaceutical Research Associates, Boehringer Ingelheim Pharma & CoKG, PTW-Freiburg Dr. Pychlau, Nanobiotix, Accuray, Varian; participation on a data safety monitoring board or advisory board for Merck Serono. A.A. Predict MarieCurie innovative training network (ITN), in frame of Horizon 2020 from the European Union, Marie Skłodowska-Curie grant agreement No 766276. M.K. No relevant relationships.

References

1. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. 2012;48:441–6.
2. Liang ZP, Lauterbur PC. Principles of magnetic resonance imaging: a signal processing perspective. SPIE Optical Engineering Press; 2000.
3. Kickingeder P, Bonekamp D, Nowosielski M, Kratz A, Sill M, Burth S, et al. Radiogenomics of glioblastoma: machine learning-based classification of molecular characteristics by using multiparametric and multiregional MR imaging features. *Radiology*. 2016;281[3]:907–18.
4. Bonekamp D, Kohl S, Wiesenfarth M, Schelb P, Radtke JP, Götz M, et al. Radiomic machine learning for characterization of prostate lesions with MRI: comparison to ADC values. *Radiology*. 2018;289[1]:128–37.

5. Vallières M, Freeman CR, Skamene SR, El Naqa I. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. *Physics in Medicine & Biology*. 2015;60[14]:5471.
6. Tian Q, Yan LF, Zhang X, Zhang X, Hu YC, Han Y, et al. Radiomics strategy for glioma grading using texture features from multiparametric MRI. *Journal of Magnetic Resonance Imaging*. 2018;48[6]:1518–28.
7. L'heureux A, Grolinger K, Elyamany HF, Capretz MA. Machine learning with big data: Challenges and approaches. *IEEE Access*. 2017;5:7776–97.
8. Stonnington CM, Tan G, Klöppel S, Chu C, Draganski B, Jack Jr CR, et al. Interpreting scan data acquired from multiple scanners: a study with Alzheimer's disease. *Neuroimage*. 2008;39[3]:1180–5.
9. Preboske GM, Gunter JL, Ward CP, Jack Jr CR. Common MRI acquisition non-idealities significantly impact the output of the boundary shift integral method of measuring brain atrophy on serial MRI. *Neuroimage*. 2006;30[4]:1196–202.
10. Collewet G, Strzelecki M, Mariette F. Influence of MRI acquisition protocols and image intensity normalization methods on texture classification. *Magnetic resonance imaging*. 2004;22[1]:81–91.
11. Alam F, Sami U, Aziz U, Fawad Q. Medical image registration: Classification, applications and issues. *Journal of Postgraduate Medical Institute*. 2018;32:300–3007.
12. Chu R, Hurwitz S, Tauhid S, Bakshi R. Automated segmentation of cerebral deep gray matter from MRI scans: effect of field strength on sensitivity and reliability. *BMC neurology*. 2017;17[1]:172.
13. Nyúl LG, Udupa JK. On standardizing the MR image intensity scale. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*. 1999;42[6]:1072–81.
14. Shah M, Xiao Y, Subbanna N, Francis S, Arnold DL, Collins DL, et al. Evaluating intensity normalization on MRIs of human brain with multiple sclerosis. *Medical image analysis*. 2011;15[2]:267–82.
15. Jäger F, Deuring-Zheng Y, Frericks B, Wacker F, Hornegger J. A new method for MRI intensity standardization with application to lesion detection in the brain. In: *Vision modeling and visualization*. Citeseer; 2006. p. 296–276.
16. Hellier P. Consistent intensity correction of MR images. In: *Proceedings 2003 International Conference on Image Processing (Cat No 03CH37429)*. IEEE; 2003. p. I–1109.
17. Loizou CP, Pantziaris M, Seimenis I, Pattichis CS. Brain MR image normalization in texture analysis of multiple sclerosis. In: *2009 9th International Conference on Information Technology and Applications in Biomedicine*. IEEE; 2009. p. 1–5.
18. Bergeest JP, Jäger F. A comparison of five methods for signal intensity standardization in MRI. In: *Bildverarbeitung für die Medizin 2008*. Springer; 2008. p. 36–40.
19. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJ, Andrearczyk V, Apte A, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*. 2020;295[2]:328–38.
20. Shinohara RT, Sweeney EM, Goldsmith J, Shiee N, Mateen FJ, Calabresi PA, et al. Statistical normalization techniques for magnetic resonance imaging. *NeuroImage: Clinical*. 2014;6:9–19.
21. Bezdek JC, Ehrlich R, Full W. FCM: The fuzzy c-means clustering algorithm. *Computers & Geosciences*. 1984;10[2–3]:191–203.
22. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 2007 Jan;8[1]:118–27.
23. Ruan Z, Mei N, Lu Y, Xiong J, Li X, Zheng W, et al. A Comparative and Summative Study of Radiomics-based Overall Survival Prediction in Glioblastoma Patients. *Journal of Computer Assisted Tomography*. 2022 Jun;46[3]:470–9.
24. Lao J, Chen Y, Li ZC, Li Q, Zhang J, Liu J, et al. A Deep Learning-Based Radiomics Model for Prediction of Survival in Glioblastoma Multiforme. *Sci Rep*. 2017 Sep 4;7[1]:10353.
25. Li Q, Bai H, Chen Y, Sun Q, Liu L, Zhou S, et al. A Fully-Automatic Multiparametric Radiomics Model: Towards Reproducible and Prognostic Imaging Signature for Prediction of Overall Survival in Glioblastoma Multiforme. *Sci Rep*. 2017 Oct 30;7[1]:14331.
26. Liu X, Li Y, Qian Z, Sun Z, Xu K, Wang K, et al. A radiomic signature as a non-invasive predictor of progression-free survival in patients with lower-grade gliomas. *NeuroImage: Clinical*. 2018 Jan 1;20:1070–7.
27. Zhang X, Lu H, Tian Q, Feng N, Yin L, Xu X, et al. A radiomics nomogram based on multiparametric MRI might stratify glioblastoma patients according to survival. *Eur Radiol*. 2019 Oct 1;29[10]:5528–38.
28. Yang Y, Han Y, Hu X, Wang W, Cui G, Guo L, et al. An Improvement of Survival Stratification in Glioblastoma Patients via Combining Subregional Radiomics Signatures. *Frontiers in Neuroscience* [Internet]. 2021 [cited 2022 May 18];15. Available from: <https://www.frontiersin.org/article/10.3389/fnins.2021.683452>
29. Li G, Li L, Li Y, Qian Z, Wu F, He Y, et al. An MRI radiomics approach to predict survival and tumour-infiltrating macrophages in gliomas. *Brain*. 2022 Mar 1;145[3]:1151–61.

30. Wang J, Zheng X, Zhang J, Xue H, Wang L, Jing R, et al. An MRI-based radiomics signature as a pretreatment noninvasive predictor of overall survival and chemotherapeutic benefits in lower-grade gliomas. *Eur Radiol.* 2021 Apr 1;31[4]:1785–94.
31. Li Z, Zhuo L, Liu P, Fei A, Ting T, Yang H, Chao Z, Zhang W, Wang J, Xu W. Construction of a prognostic immune signature for lower grade glioma that can be recognized by MRI radiomics features to predict survival in LGG patients. *Translational Oncology.* 2021 Jun 1;14[6]:101065.
32. Chaddad A, Daniel P, Zhang M, Rathore S, Sargos P, Desrosiers C, et al. Deep radiomic signature with immune cell markers predicts the survival of glioma patients. *Neurocomputing.* 2022 Jan 16;469:366–75.
33. Han W, Qin L, Bay C, Chen X, Yu KH, Miskin N, et al. Deep Transfer Learning and Radiomics Feature Prediction of Survival of Patients with High-Grade Gliomas. *American Journal of Neuroradiology.* 2020 Jan 1;41[1]:40–8.
34. Shboul ZA, Alam M, Vidyaratne L, Pei L, Elbakary MI, Iftekharuddin KM. Feature-Guided Deep Radiomics for Glioblastoma Patient Survival Prediction. *Frontiers in Neuroscience* [Internet]. 2019 [cited 2022 May 18];13. Available from: <https://www.frontiersin.org/article/10.3389/fnins.2019.00966>
35. Tan Y, Mu W, Wang X, Chun W, Yang G, Qiang G, Gillies RJ, Zhang H. Improving survival prediction of high-grade glioma via machine learning techniques based on MRI radiomic, genetic and clinical risk factors. *European Journal of Radiology.* 2019 Nov 1;120:108609.
36. Choi YS, Ahn SS, Chang JH, Kang SG, Kim EH, Kim SH, et al. Machine learning and radiomic phenotyping of lower grade gliomas: improving survival prediction. *Eur Radiol.* 2020 Jul 1;30[7]:3834–42.
37. Chaddad A, Daniel P, Desrosiers C, Toews M, Abdulkarim B. Novel Radiomic Features Based on Joint Intensity Matrices for Predicting Glioblastoma Patient Survival Time. *IEEE Journal of Biomedical and Health Informatics.* 2019 Mar;23[2]:795–804.
38. Bakas S, Shukla G, Akbari H, Erus G, Sotiras A, Rathore S, et al. Overall survival prediction in glioblastoma patients using structural magnetic resonance imaging (MRI): advanced radiomic features may compensate for lack of advanced MRI modalities. *JMI.* 2020 Jun;7[3]:031505.
39. Baid U, Rane SU, Talbar S, Gupta S, Thakur MH, Moiyadi A, et al. Overall Survival Prediction in Glioblastoma With Radiomic Features Using Machine Learning. *Frontiers in Computational Neuroscience* [Internet]. 2020 [cited 2022 May 18];14. Available from: <https://www.frontiersin.org/article/10.3389/fncom.2020.00061>
40. Chaddad A, Sabri S, Niazi T, Abdulkarim B. Prediction of survival with multi-scale radiomic analysis in glioblastoma patients. *Med Biol Eng Comput.* 2018 Dec 1;56[12]:2287–300.
41. Tixier F, Um H, Bermudez D, Iyer A, Apte A, Graham MS, et al. Preoperative MRI-radiomics features improve prediction of survival in glioblastoma patients over MGMT methylation status alone. *Oncotarget.* 2019;10[6]:660.
42. Han K, Ren M, Wick W, Abrey L, Das A, Jin J, et al. Progression-free survival as a surrogate endpoint for overall survival in glioblastoma: a literature-based meta-analysis from 91 trials. *Neuro-oncology.* 2014;16[5]:696–706.
43. Yan J, Zhang B, Zhang S, Cheng J, Liu X, Wang W, et al. Quantitative MRI-based radiomics for noninvasively predicting molecular subtypes and survival in glioma patients. *npj Precis Onc.* 2021 Jul 26;5[1]:1–9.
44. Prasanna P, Patel J, Partovi S, Madabhushi A, Tiwari P. Radiomic features from the peritumoral brain parenchyma on treatment-naïve multi-parametric MR imaging predict long versus short-term survival in glioblastoma multiforme: Preliminary findings. *Eur Radiol.* 2017 Oct 1;27[10]:4188–97.
45. Bae S, Choi YS, Ahn SS, Chang JH, Kang SG, Kim EH, et al. Radiomic MRI phenotyping of glioblastoma: improving survival prediction. *Radiology.* 2018;289[3]:797–806.
46. Kickingreder P, Burth S, Wick A, Götz M, Eidel O, Schlemmer HP, et al. Radiomic profiling of glioblastoma: identifying an imaging predictor of patient survival with improved performance over established clinical and radiologic risk models. *Radiology.* 2016;280[3]:880–9.
47. Choi Y, Nam Y, Jang J, Shin NY, Lee YS, Ahn KJ, et al. Radiomics may increase the prognostic value for survival in glioblastoma patients when combined with conventional clinical and genetic prognostic models. *Eur Radiol.* 2021 Apr 1;31[4]:2084–93.
48. Ibrahim A, Primakov S, Beuque M, Woodruff H, Halilaj I, Wu G, et al. Radiomics for precision medicine: current challenges, future prospects, and the proposal of a new framework. *Methods.* 2020;
49. Rizzo S, Botta F, Raimondi S, Origi D, Fanciullo C, Morganti AG, et al. Radiomics: the facts and the challenges of image analysis. *European radiology experimental.* 2018;2[1]:1–8.
50. Molina D, Pérez-Beteta J, Martínez-González A, Martino J, Velásquez C, Arana E, et al. Influence of gray level and space discretization on brain tumor heterogeneity measures obtained from magnetic resonance images. *Computers in biology and medicine.* 2016;78:49–57.
51. Bologna M, Corino V, Mainardi L. Virtual phantom analyses for preprocessing evaluation and detection of a robust feature set for MRI-radiomics of the brain. *Medical physics.* 2019;46[11]:5116–23.

52. Duron L, Balvay D, Vande Perre S, Bouchouicha A, Savatovsky J, Sadik JC, et al. Gray-level discretization impacts reproducible MRI radiomics texture features. *PloS one*. 2019;14[3]:e0213459.
53. Carré A, Klausner G, Edjlali M, Lerousseau M, Briend-Diop J, Sun R, et al. Standardization of brain MR images across machines and protocols: bridging the gap for MRI-based radiomics. *Scientific reports*. 2020;10[1]:1–15.
54. Haralick RM, Shanmugam K, Dinstein I. Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*. 1973 Nov;SMC-3[6]:610–21.
55. Scapicchio C, Gabelloni M, Barucci A, Cioni D, Saba L, Neri E. A deep look into radiomics. *Radiol med*. 2021 Oct 1;126[10]:1296–311.
56. Fatania K, Mohamud F, Clark A, Nix M, Short SC, O'Connor J, et al. Intensity standardization of MRI prior to radiomic feature extraction for artificial intelligence research in glioma—a systematic review. *Eur Radiol* [Internet]. 2022 Apr 29 [cited 2022 May 23]; Available from: <https://doi.org/10.1007/s00330-022-08807-2>
57. Um H, Tixier F, Bermudez D, Deasy JO, Young RJ, Veeraraghavan H. Impact of image preprocessing on the scanner dependence of multi-parametric MRI radiomic features and covariate shift in multi-institutional glioblastoma datasets. *Phys Med Biol*. 2019 Aug;64[16]:165011.
58. Li Y, Ammari S, Balleyguier C, Lassau N, Chouzenoux E. Impact of Preprocessing and Harmonization Methods on the Removal of Scanner Effects in Brain MRI Radiomic Features. *Cancers*. 2021 Jan;13[12]:3000.
59. Tixier F, Um H, Bermudez D, Iyer A, Apte A, Graham MS, et al. Preoperative MRI-radiomics features improve prediction of survival in glioblastoma patients over MGMT methylation status alone. *Oncotarget*. 2019 Jan 18;10[6]:660–72.
60. Ellingson BM, Bendszus M, Boxerman J, Barboriak D, Erickson BJ, Smits M, et al. Consensus recommendations for a standardized Brain Tumor Imaging Protocol in clinical trials. *Neuro-Oncology*. 2015 Sep 1;17[9]:1188–98.
61. Combs SE, Burkholder I, Edler L, Rieken S, Habermehl D, Jäkel O, et al. Randomised phase I/II study to evaluate carbon ion radiotherapy versus fractionated stereotactic radiotherapy in patients with recurrent or progressive gliomas: The CINDERELLA trial. *BMC Cancer*. 2010 Oct 6;10[1]:533.
62. Combs SE, Kieser M, Rieken S, Habermehl D, Jäkel O, Haberer T, et al. Randomized phase II study evaluating a carbon ion boost applied after combined radiochemotherapy with temozolomide versus a proton boost after radiochemotherapy with temozolomide in patients with primary glioblastoma: The CLEOPATRA Trial. *BMC Cancer*. 2010 Sep 6;10[1]:478.
63. Sforazzini F, Salome P, Kudak A, Ulrich M, Bougatf N, Debus J, et al. pyCuRT: An Automated Data Curation Workflow for Radiotherapy Big Data Analysis using Python's NyPipe. *International Journal of Radiation Oncology, Biology, Physics*. 2020 Nov 1;108[3]:e772.
64. Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, et al. N4ITK: improved N3 bias correction. *IEEE transactions on medical imaging*. 2010;29[6]:1310–20.
65. Isensee F, Schell M, Pflueger I, Brugnara G, Bonekamp D, Neuberger U, et al. Automated brain extraction of multisequence MRI using artificial neural networks. *Human brain mapping*. 2019;40[17]:4952–64.
66. Ebner M, Wang G, Li W, Aertsen M, Patel PA, Aghwane R, et al. An automated framework for localization, segmentation and super-resolution reconstruction of fetal brain MRI. *NeuroImage*. 2020;206:116324.
67. Avants BB, Tustison N, Song G. Advanced normalization tools (ANTS). *Insight j*. 2009;2[365]:1–35.
68. Reinhold JC, Dewey BE, Carass A, Prince JL. Evaluating the impact of intensity normalization on MR image synthesis. In: *Medical Imaging 2019: Image Processing*. International Society for Optics and Photonics; 2019. p. 109493H.
69. Zhang Y, Brady M, Smith S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE transactions on medical imaging*. 2001;20[1]:45–57.
70. Reynolds DA. Gaussian Mixture Models. *Encyclopedia of biometrics*. 2009;741.
71. Beer JC, Tustison NJ, Cook PA, Davatzikos C, Sheline YI, Shinohara RT, et al. Longitudinal ComBat: A method for harmonizing longitudinal multi-scanner imaging data. *Neuroimage*. 2020 Oct 15;220:117129.
72. Da-Ano R, Masson I, Lucia F, Doré M, Robin P, Alfieri J, et al. Performance comparison of modified ComBat for harmonization of radiomic features for multicenter studies. *Sci Rep*. 2020 Jun 24;10[1]:10248.
73. Orhac F, Lecler A, Savatovski J, Goya-Outi J, Nioche C, Charbonneau F, et al. How can we combat multicenter variability in MR radiomics? Validation of a correction procedure. *Eur Radiol*. 2021 Apr;31[4]:2272–80.
74. Van Griethuysen JJ, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer research*. 2017;77[21]:e104–7.
75. Lin DY, Wei LJ. The robust inference for the Cox proportional hazards model. *Journal of the American statistical Association*. 1989;84[408]:1074–8.
76. Frome EL. The analysis of rates using Poisson regression models. *Biometrics*. 1983;665–74.
77. Campos B, Olsen LR, Urup T, Poulsen HS. A comprehensive profile of recurrent glioblastoma. *Oncogene*. 2016 Nov;35[45]:5819–25.

78. Boots-Sprenger SHE, Sijben A, Rijntjes J, Tops BBJ, Idema AJ, Rivera AL, et al. Significance of complete 1p/19q co-deletion, IDH1 mutation and MGMT promoter methylation in gliomas: use with caution. *Mod Pathol.* 2013 Jul;26[7]:922–9.

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.