

Article

# Comprehensive Analysis of Tumour Sub-Volumes for Radiomic Risk Modelling in Locally Advanced HNSCC

Stefan Leger <sup>1,2,3,\*</sup>, Alex Zwanenburg <sup>1,2,3</sup>, Karoline Leger <sup>1,2,3,4</sup>, Fabian Lohaus <sup>1,2,3,4</sup>, Annett Linge <sup>1,2,3,4</sup>, Andreas Schreiber <sup>5</sup>, Goda Kalinauskaitė <sup>6,7</sup>, Inge Tinhofer <sup>6,7</sup>, Nika Guberina <sup>8,9</sup>, Maja Guberina <sup>8,9</sup>, Panagiotis Balermipas <sup>10,11</sup>, Jens von der Grün <sup>10,11</sup>, Ute Ganswindt <sup>12,13,14,15</sup>, Claus Belka <sup>12,13,14</sup>, Jan C. Peeken <sup>12,16,17</sup>, Stephanie E. Combs <sup>12,16,17</sup>, Simon Boeke <sup>18,19</sup>, Daniel Zips <sup>18,19</sup>, Christian Richter <sup>1,2,4,20</sup>, Mechthild Krause <sup>1,2,3,4,20</sup>, Michael Baumann <sup>1,2,3,4,20,21</sup>, Esther G.C. Troost <sup>1,2,3,4,20,†</sup> and Steffen Löck <sup>1,2,3,†</sup>

- <sup>1</sup> OncoRay—National Center for Radiation Research in Oncology, Faculty of Medicine and University Hospital Carl Gustav Carus, Technische Universität Dresden, Helmholtz-Zentrum Dresden—Rossendorf, 01307 Dresden, Germany; Alex.Zwanenburg@nct-dresden.de (A.Z.); Karoline.Leger@uniklinikum-dresden.de (K.L.); Fabian.Lohaus@uniklinikum-dresden.de (F.L.); Annett.Linge@uniklinikum-dresden.de (A.L.); christian.richter@oncoray.de (C.R.); Mechthild.Krause@uniklinikum-dresden.de (M.K.); michael.baumann@dkfz-heidelberg.de (M.B.); esther.troost@uniklinikum-dresden.de (E.G.C.T.); steffen.loeck@oncoray.de (S.L.)
- <sup>2</sup> German Cancer Research Center (DKFZ), Heidelberg and German Cancer Consortium (DKTK) Partner Site, 01307 Dresden, Germany
- <sup>3</sup> National Center for Tumor Diseases (NCT), Partner Site Dresden of the German Cancer Research Center (DKFZ), Faculty of Medicine and University Hospital Carl Gustav Carus and Technische Universität Dresden, 01307 Dresden, Germany
- <sup>4</sup> Department of Radiotherapy and Radiation Oncology, Faculty of Medicine and University Hospital Carl Gustav Carus, Technische Universität Dresden, 01307 Dresden, Germany
- <sup>5</sup> Department of Radiotherapy, Hospital Dresden-Friedrichstadt, 01067 Dresden, Germany; schreiber-an@khdf.de
- <sup>6</sup> German Cancer Research Center (DKFZ), Heidelberg and German Cancer Consortium (DKTK) Partner Site, 10117 Berlin, Germany; goda.kalinauskaitė@charite.de (G.K.); ingeborg.tinhofer@charite.de (I.T.)
- <sup>7</sup> Department of Radiooncology and Radiotherapy, Charité University Hospital, 10117 Berlin, Germany
- <sup>8</sup> German Cancer Research Center (DKFZ), Heidelberg and German Cancer Consortium (DKTK) Partner Site, 45147 Essen, Germany; nika.guberina@uk-essen.de (N.G.); maja.guberina@uk-essen.de (M.G.)
- <sup>9</sup> Department of Radiotherapy, University Hospital Essen, Medical Faculty, University of Duisburg-Essen, 45147 Essen, Germany
- <sup>10</sup> German Cancer Research Center (DKFZ), Heidelberg and German Cancer Consortium (DKTK) Partner Site, 60596 Frankfurt, Germany; Panagiotis.Balermipas@usz.ch (P.B.); Jens.VonderGruen@kgu.de (J.v.d.G.)
- <sup>11</sup> Department of Radiotherapy and Oncology, Goethe-University Frankfurt, 60596 Frankfurt, Germany
- <sup>12</sup> German Cancer Research Center (DKFZ), Heidelberg and German Cancer Consortium (DKTK) Partner Site, 81377 Munich, Germany; ute.ganswindt@i-med.ac.at (U.G.); Claus.Belka@med.uni-muenchen.de (C.B.); jan.peeken@tum.de (J.C.P.); Stephanie.Combs@tum.de (S.E.C.)
- <sup>13</sup> Department of Radiation Oncology, Ludwig-Maximilians-Universität, 81377 Munich, Germany
- <sup>14</sup> Clinical Cooperation Group, Personalized Radiotherapy in Head and Neck Cancer, Helmholtz Zentrum, 81377 Munich, Germany
- <sup>15</sup> Department of Radiation Oncology, Medical University of Innsbruck, Anichstraße 35, A-6020 Innsbruck, Austria
- <sup>16</sup> Department of Radiation Oncology, Technische Universität München, 81675 Munich, Germany
- <sup>17</sup> Institute of Radiation Medicine (IRM), Helmholtz Zentrum München, 85764 Neuherberg, Germany
- <sup>18</sup> German Cancer Research Center (DKFZ), Heidelberg and German Cancer Consortium (DKTK) Partner Site, 72076 Tübingen, Germany; simon.boeke@med.uni-tuebingen.de (S.B.); Daniel.Zips@med.uni-tuebingen.de (D.Z.)

<sup>19</sup> Department of Radiation Oncology, Faculty of Medicine and University Hospital Tübingen, Eberhard Karls Universität Tübingen, 72076 Tübingen, Germany

<sup>20</sup> Institute of Radiooncology—OncoRay, Helmholtz-Zentrum Dresden—Rossendorf, 01328 Dresden, Germany

<sup>21</sup> German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany

\* Correspondence: Stefan.Leger@nct-dresden.de

† These authors share senior authorship.

Received: 28 August 2020; Accepted: 13 October 2020; Published: 19 October 2020



**Simple Summary:** Radiomic risk models are usually based on imaging features, which are extracted from the entire gross tumour volume (GTV<sub>entire</sub>). This approach does not explicitly consider the complex biological structure of the tumours. Therefore, in this retrospective study, we investigated the prognostic value of radiomic analyses based on different tumour sub-volumes using computed tomography imaging of patients with locally advanced head and neck squamous cell carcinoma who were treated with primary radio-chemotherapy. The GTV<sub>entire</sub> was cropped by different margins to define the rim and corresponding core sub-volumes of the tumour. Furthermore, the best performing tumour rim sub-volume was extended into surrounding tissue with different margins. As a result, the models based on the 5 mm tumour rim and on the 3 mm extended rim sub-volume showed an improved performance compared to models based on the corresponding tumour core. This indicates that the consideration of tumour sub-volumes may help to improve radiomic risk models.

**Abstract:** Imaging features for radiomic analyses are commonly calculated from the entire gross tumour volume (GTV<sub>entire</sub>). However, tumours are biologically complex and the consideration of different tumour regions in radiomic models may lead to an improved outcome prediction. Therefore, we investigated the prognostic value of radiomic analyses based on different tumour sub-volumes using computed tomography imaging of patients with locally advanced head and neck squamous cell carcinoma. The GTV<sub>entire</sub> was cropped by different margins to define the rim and the corresponding core sub-volumes of the tumour. Subsequently, the best performing tumour rim sub-volume was extended into surrounding tissue with different margins. Radiomic risk models were developed and validated using a retrospective cohort consisting of 291 patients in one of the six Partner Sites of the German Cancer Consortium Radiation Oncology Group treated between 2005 and 2013. The validation concordance index (C-index) averaged over all applied learning algorithms and feature selection methods using the GTV<sub>entire</sub> achieved a moderate prognostic performance for loco-regional tumour control (C-index:  $0.61 \pm 0.04$  (mean  $\pm$  std)). The models based on the 5 mm tumour rim and on the 3 mm extended rim sub-volume showed higher median performances (C-index:  $0.65 \pm 0.02$  and  $0.64 \pm 0.05$ , respectively), while models based on the corresponding tumour core volumes performed less (C-index:  $0.59 \pm 0.01$ ). The difference in C-index between the 5 mm tumour rim and the corresponding core volume showed a statistical trend ( $p = 0.10$ ). After additional prospective validation, the consideration of tumour sub-volumes may be a promising way to improve prognostic radiomic risk models.

**Keywords:** radiomic; image-based risk modelling; machine learning; personalised therapy; radiation oncology

---

## 1. Introduction

The individualisation of radiation oncology is a major objective in modern cancer therapy [1]. Radiomics aims to characterise the tumour phenotype using advanced image features to predict patient-specific outcome. Commonly, radiomic features are computed and extracted using the entire gross tumour volume (GTV<sub>entire</sub>) [2–5]. Such an approach assumes that the individual tumour is

either homogeneous or heterogeneous, but uniformly distributed over the entire tumour volume. However, tumours are biologically complex and exhibit substantial spatial variation, e.g., in gene expression and in microscopic structure [6]. Such spatial variation may be caused by, e.g., hypoxia and necrosis which may appear in the tumour core and high cell proliferation and infiltrating tumour cell growth, which may occur at the tumour periphery [7]. Some of these regional tumour variations are apparent in imaging data, e.g., necrosis or tumour vascularisation detected by magnetic resonance imaging (MRI) or tumour hypoxia measured by  $^{18}\text{F}$ -fluoromisonidazole positron emission tomography (FMISO-PET) [8–12]. Furthermore, different regions within an individual tumour may differ in radio-sensitivity, which may depend on the distribution of cancer stem cells and localised genetic or molecular alterations [13,14]. In the case of head and neck squamous cell carcinoma (HNSCC), several studies have shown that the tumour micro-environment plays a major role for cancer development and progression [15]. Alshafi et al. [16] showed that the poor response to therapy and the aggressive nature of HNSCC are not only caused by the complex alterations in intracellular signalling pathways, but are also influenced by the behaviour of the extracellular micro-environment. As consequences, such spatial variations may affect the performance of image-based risk models.

Thus far, only few studies have investigated and analysed specific tumour sub-volumes for radiomic risk modelling. Recently, Algohary et al. [17] showed that the combination of peri-tumoural and intra-tumoural radiomic features derived from prostate bi-parametric MR images leads to an improved risk assessment of prostate cancer patients. Furthermore, Grove et al. [18] showed that the expressions of 2-dimensional radiomic features computed on the rim of the tumour differed from those calculated on the tumour core. The ratio of tumour rim and core features led to an improved prediction of overall survival (OS) in non-small cell lung cancer patients. Wu et al. [6] identified clinically relevant tumour sub-volumes to characterise the regional heterogeneity of tumours in breast cancer patients based on dynamic contrast enhanced magnetic resonance imaging. The resulting risk models based on the identified sub-volumes also showed an improved outcome prediction compared to models based on the  $\text{GTV}_{\text{entire}}$ . In a further study, Wu et al. [19] identified different tumour sub-volumes using computed tomography (CT) and  $^{18}\text{F}$ -fluorodeoxyglucose PET (FDG-PET) imaging of lung cancer patients. It was shown that spatially distinct sub-volumes are linked to higher risk of recurrence compared to the  $\text{GTV}_{\text{entire}}$ , resulting in an improved model prediction of OS.

Aside from these initial findings, in most of the previously described studies, only individual clinical parameters or radiomic features (e.g., tumour volume) were investigated. Therefore, systematic investigations of the potential of radiomic risk models based on different tumour sub-volumes are still sparse.

In the present study, we systematically compared radiomic models based on two different sub-volumes of the  $\text{GTV}_{\text{entire}}$ , the outer tumour rim and the complementary tumour core, using pre-treatment CT imaging [20,21]. A multi-centre cohort of 291 patients with locally advanced HNSCC treated by primary radio-chemotherapy was considered. For the prediction of loco-regional tumour control (LRC), risk models were developed and independently validated. Patients were stratified into groups at low and high risk of loco-regional recurrence. Furthermore, we investigated the prognostic performance of the developed models for sub-groups of small and large tumours and extended the tumour rim beyond the  $\text{GTV}_{\text{entire}}$  to account for potential sub-microscopic spread, leading to the clinical target volume [22].

## 2. Materials and Methods

### 2.1. Characteristics of Patient Cohorts

A retrospective multi-centre cohort consisting of 291 patients with histologically confirmed loco-regionally advanced HNSCC was used. All patients received primary radio-chemotherapy (RCT) and underwent a CT scan with or without contrast-enhancement for treatment-planning purpose. The multi-centre cohort was divided into an exploratory and a validation cohort by an approximate

ratio of 2:1. In the exploratory cohort, 149 of the 206 patients were treated in one of the six Partner Sites of the German Cancer Consortium Radiation Oncology Group (DKTK-ROG) between 2005 and 2011 [23]. The remaining 57 patients were treated at the University Hospital Dresden (UKD) between 1999 and 2006. The validation cohort consisted of 85 patients from which 51 patients received their treatment within a prospective clinical trial (ClinicalTrials.gov Identifier: NCT00180180) at the UKD between 2006 and 2012 [9,12]. The remaining 34 patients were treated at the UKD or the Radiotherapy Centre Dresden-Friedrichstadt between 2005 and 2009 as well as at the University Hospital Tübingen between 2008 and 2013. Patient characteristics for the exploratory and the independent validation cohort are summarised in Table 1.

Radiomic risk models were developed to predict the primary clinical endpoint LRC, which was defined as the time from the first day of RCT to the date of loco-regional recurrence (event) or to the end of follow-up (censoring). Ethical approval for the multi-centre retrospective analyses of clinical and imaging data was obtained from the Ethics Committee at the Technische Universität Dresden (EK177042017, May 2017). All analyses were carried out in accordance with the relevant guidelines and regulations. Informed consent was obtained from all patients.

**Table 1.** Patient characteristics of the exploratory and the independent validation cohort.

Clinical Variable	Exploratory Cohort	Validation Cohort	p-Value
Number of patients	206	85	-
Gender			
male	174	74	0.70 <sup>2</sup>
female	32	11	
Age in years			
median	59.0	55.0	0.023 <sup>3</sup>
range	39.2–84.5	37.0–76.0	-
cTN staging			
T stage 1/2/3/4	2/23/51/130	2/9/30/44	0.21 <sup>1</sup>
N stage 0/1/2/3/missing	30/7/154/15/0	9/8/64/3/1	0.097 <sup>1</sup>
UICC stage 2010			
I/II/III/IV	0/0/15/191	1/2/9/73	0.039 <sup>1</sup>
GTV (cm <sup>3</sup> )			
median	29.1	40.6	0.067 <sup>3</sup>
range	4.5–321.7	2.7–239.1	-
Tumour site			
oropharynx/oral cavity/ hypopharynx/larynx	93/51/62/0	29/23/28/5	0.003 <sup>3</sup>
p16 status			
negative/positive/missing	148/28/30	52/5/28	0.26 <sup>1</sup>
Loco-regional tumour recurrence	84 (41%)	28 (33%)	0.26 <sup>3</sup>
Follow up time (months)			
median	21.2	24.3	-
range	1.2–131.9	1.3–107.2	0.64 <sup>3</sup>

Abbreviations: T, clinical tumour stage; N, clinical nodal stage; UICC, Union internationale contre le cancer; Gy, Gray; DNA, deoxyribonucleic acid; GTV, primary gross tumour volume; <sup>1</sup>  $\chi^2$  test; <sup>2</sup> exact Fisher test; <sup>3</sup> Wilcoxon–Mann–Whitney test.

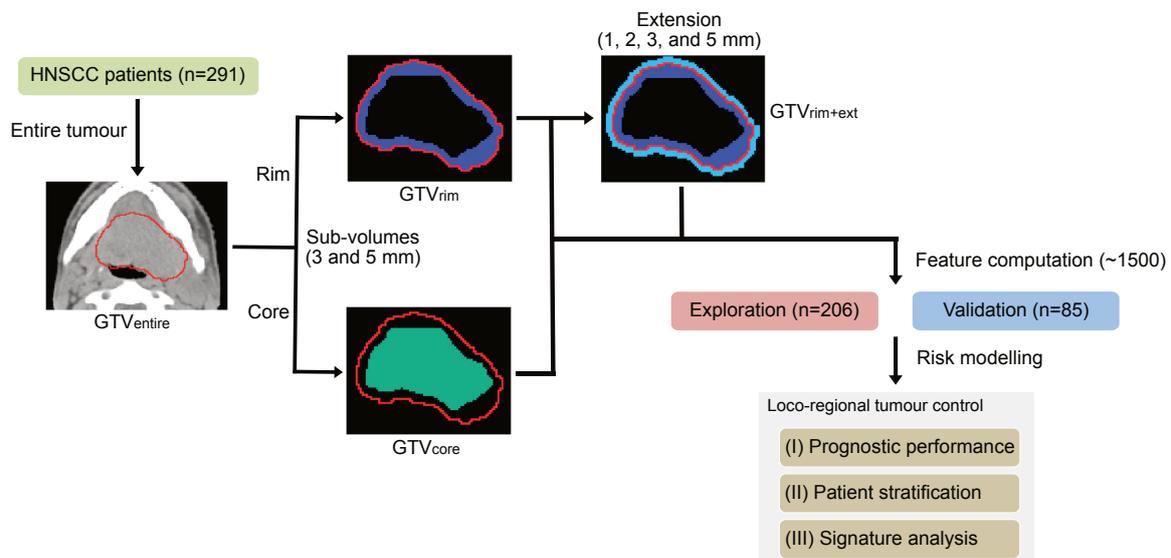
## 2.2. Tumour Sub-Volume Definition and Feature Computation

The analysis was divided into two subsequent steps, which are shown in Figure 1. The GTV<sub>entire</sub>, i.e., the primary gross tumour volume, was manually delineated on each planning CT scan by a radiation oncologist using the CT image information only. Subsequently, the image voxel spacing was

resampled using cubic spline image interpolation to an isotropic voxel size of  $1.0 \times 1.0 \times 1.0 \text{ mm}^3$  to correct for differences in voxel spacing and slice thickness between the cohorts [2,24].

Based on the delineated  $\text{GTV}_{\text{entire}}$ , two distinct sub-volumes were generated. The outer contour of the  $\text{GTV}_{\text{entire}}$  was cropped by different margins (3 and 5 mm) to define the rim of the tumour ( $\text{GTV}_{3\text{mm-rim}}$  and  $\text{GTV}_{5\text{mm-rim}}$ , respectively). The corresponding remaining sub-volumes were defined as tumour core ( $\text{GTV}_{3\text{mm-core}}$  and  $\text{GTV}_{5\text{mm-core}}$ , respectively). The minimum core volume was restricted to 40% of the entire tumour volume to avoid disappearance of the core sub-volume in small tumours. Furthermore, the best performing tumour rim sub-volume was extended ( $\text{GTV}_{\text{rim+ext}}$ ) into surrounding tissue with different distances (1, 2, 3 and 5 mm) to assess the prognostic performance of the microscopic tumour extension.

Nine additional images were created by applying spatial filtering to the base image to emphasise image characteristics, such as edges and blobs. Eight additional images were created by applying a stationary coiflet-1 wavelet high-/low-pass filter along each of the three spatial dimensions. One further image was created by applying a Laplacian of Gaussian (LoG) filter consisting of five different filter kernel widths (1.0, 2.0, 3.0, 5.0 and 6.0 mm). Subsequently, the tumour mask was re-segmented to include only soft tissue voxels between  $-150$  and  $180$  Hounsfield units, thereby removing voxels containing air or bone, which may affect feature expression. Features were implemented in compliance with the Image Biomarker Standardisation Initiative [25]. A total of 1538 features were computed and extracted from each sub-volume. A total of 18 statistical, 38 histogram-based and 95 texture features were calculated on the base image and on the nine transformed images. Moreover, 28 morphological features were determined on the base image only. The configuration settings for the image feature computation and extraction is summarised in Table A2.



**Figure 1.** Experimental design. A multi-centre cohort of 291 patients with loco regionally advanced head and neck squamous cell carcinoma (HNSCC) was used to generate different sub-volumes based on the delineated entire tumour. In particular, the outer contour of the entire primary cross tumour volume ( $\text{GTV}_{\text{entire}}$ ) was cropped by different margins (3 and 5 mm) to define the rim of the tumour ( $\text{GTV}_{\text{rim}}$ ) and the corresponding core ( $\text{GTV}_{\text{core}}$ ). Furthermore, the best performing tumour rim sub-volume was extended ( $\text{GTV}_{\text{rim+ext}}$ ) into surrounding tissue with different distances (1, 2, 3, and 5 mm) to assess the prognostic performance of the microscopic tumour extension. The entire cohort was split into an exploratory and an independent validation cohort for risk modelling. Prognostic model performance and patient risk group stratification were assessed on the validation cohort. Selected features within the developed signatures were analysed in terms of their univariate association with loco-regional tumour control using the entire cohort.

### 2.3. Radiomic Risk Modelling

Radiomic risk models were developed using an end-to-end modelling framework, which consists of five steps: (I) feature pre-processing; (II) feature selection; (III) hyper-parameter optimisation; (IV) model development; and (V) model validation. The risk models were generated as previously described [4]. Briefly, after feature normalisation and clustering, feature selection was performed multiple times using 1000 bootstrap samples of the exploratory cohort. Subsequently, model training was conducted on 1000 bootstrap samples of the exploratory cohort, using the highest ranked features as well as the optimised hyper-parameter set. Finally, an ensemble prediction was made by averaging the predicted risk scores of each model for both the exploratory and the independent validation cohort separately.

Combinations of five feature selection methods and six learning algorithms were used for model development to reduce the risk of incidental findings, based on the recommendation in Leger et al. [4]. The following feature selection methods were used: Spearman correlation, mutual information maximisation (MIM), mutual information feature selection (MIFS), minimum redundancy maximum relevance (MRMR) and random forest variable importance (RFVI). The six learning algorithms comprised: Cox proportional hazard model (Cox), boosted tree-Cox (BT-Cox), boosted gradient linear model-Cox (BGLM-Cox), random survival forest (RSF) and maximally selected rank statistics random forest (MSR-RF) as well as the full-parametric BT-Weibull model. Table A3 summarises the definition of the hyper-parameters of the feature selection methods and of the machine learning algorithms, which were used during the hyper-parameter optimisation.

### 2.4. Performance Assessments

(I) The prognostic performance of the radiomic models was assessed on the exploratory and on the independent validation cohort using the concordance index (C-index) [26,27]. The C-index is a generalisation of the area under the curve for continuous time-to-event survival data and a C-index of 0.5 describes a random prediction, whereas a perfectly predicting model has C-index of 1.0. Risk models were developed based on  $GTV_{entire}$ ,  $GTV_{3mm-rim}$  and  $GTV_{5mm-rim}$ , as well as the corresponding core volumes  $GTV_{3mm-core}$  and  $GTV_{5mm-core}$ .

The median C-indices over all combination of feature selection methods and machine learning algorithms were determined based on the exploratory and the validation cohort for each tumour sub-volume. For the considered feature selection methods and machine learning algorithms, the model performance was statistically compared between the  $GTV_{3mm-rim}$  and  $GTV_{5mm-rim}$  and their corresponding core volumes using a multi-level model approach (MLA), which is described in Appendix A.1 [5]. Subsequently, representative model combinations for each sub-volume were selected, consisting of one feature selection method and one learning algorithm. To choose this representative model, the median performances of every feature selection method over all learning algorithms and vice versa were determined. The model generated by the feature selection method with a C-index closest to the median feature selection performances and the learning algorithm with a C-index closest to the median learning algorithm performances was then selected. For the further analyses, the representative models based on the sub-volume of (a) the  $GTV_{entire}$ ; (b) the tumour rim; (c) the corresponding core and (d) the extended rim were investigated in more detail. In addition, we assigned the patients of the validation cohort into two sub-groups according their initial tumour volume using  $20 \text{ cm}^3$  as a threshold value, which corresponds to a tumour radius of approximately 1.5 cm in the case of a spherical tumour. Subsequently, we investigated the prognostic performance of the developed models using the resulting sub-groups individually.

(II) Risk-based patient stratification into groups at low and high risk of loco-regional recurrence was performed for each tumour sub-volume and for all model combinations. The results for the selected models (a)–(d) were analysed in more detail. Patients were stratified into a low and high risk group based on the predicted risk of the radiomics models. The cut-off value used for stratification was based on the median predicted risk ( $median_{risk}$ ) determined on the exploratory cohort. This cut-off value

was directly applied to the validation cohort. Survival curves were estimated using the Kaplan–Meier method and the stratification was compared using log-rank tests. Log-rank test  $p$ -values  $< 0.05$  were considered to be statistically significant.

(III) Radiomic signatures were analysed in detail for the models trained on the different selected tumour sub-volumes (a)–(d). Features included in the signatures and their expression values were depicted as heatmaps for the exploratory and the validation cohort. For this purpose, all patients were sorted according to their predicted risk and to their risk group stratification. To quantify the overall importance of the identified features, univariate significance of the individual radiomic features included in the signatures were tested by the Cox model on the entire patient cohort.

### 3. Results

The number of loco-regional recurrences was 84 for the exploratory and 28 for the independent validation cohort, respectively. The primary endpoint LRC showed no significant difference between both cohorts ( $p = 0.26$ ). The median follow-up time was 21.2 months (range: 1.2–131.9 months) for the exploratory and 24.3 months (range: 1.3–107.2 months) for the validation cohort ( $p = 0.64$ ). The median GTV<sub>entire</sub> was 29.2 cm<sup>3</sup> (range: 4.4–322.2 cm<sup>3</sup>) in the exploratory cohort and 40.1 cm<sup>3</sup> (range: 2.7–239.0 cm<sup>3</sup>) in the validation cohort ( $p = 0.067$ , Table 1). For further analyses, 3 mm and 5 mm margins were subtracted from the GTV<sub>entire</sub>, respectively. The median volume fractions of the resulting tumour rim sub-volumes were 47.7% (range: 26.1–59.5%) for the GTV<sub>3mm-rim</sub> and 52.8% (range: 33.1–59.9%) for the GTV<sub>5mm-rim</sub> sub-volumes in the exploratory cohort and 46.4% (range: 26.1–59.5%) and 52.1% (range: 33.1–59.9%) in the validation cohort ( $p = 0.066$  and  $p = 0.081$ , respectively).

#### 3.1. Prognostic Performance

Radiomic models were developed and validated based on the different (sub-)volumes of the tumour. Their performance for the prognosis of LRC is summarised in Table 2 for both cohorts. The median C-index on the exploratory cohort was between 0.72 and 0.76 for the considered sub-volumes. For the validation cohort, models based on the GTV<sub>entire</sub> achieved a median prognostic performance of  $0.61 \pm 0.04$  (median  $\pm$  standard deviation (SD)), while the models based on the tumour rim sub-volumes showed a slightly better median performance on the validation cohort (C-index: GTV<sub>3mm-rim</sub>:  $0.63 \pm 0.03$  and GTV<sub>5mm-rim</sub>:  $0.65 \pm 0.02$ , respectively). The core-based risk models revealed the lowest prognostic performance on the validation cohort (C-index: GTV<sub>3mm-core</sub>:  $0.60 \pm 0.02$  and GTV<sub>5mm-core</sub>:  $0.59 \pm 0.01$ , respectively). The difference in C-index between GTV<sub>5mm-rim</sub> and GTV<sub>5mm-core</sub> showed a statistical trend (MLA:  $p = 0.10$ ), while the difference between GTV<sub>3mm-rim</sub> and GTV<sub>3mm-core</sub> was not statistically significant (MLA:  $p = 0.50$ ). The median performances of the sub-group analyses showed similar results for small GTV<sub>entire</sub> ( $\leq 20$  cm<sup>3</sup>) between the tumour rim- and core-based models (3 mm: 0.62 vs. 0.63 and 5 mm: 0.67 vs. 0.69, respectively), whereas, the differences between rim and core were larger for larger GTV<sub>entire</sub> (3 mm: 0.61 vs. 0.57 and 5 mm: 0.61 vs. 0.57) in the validation cohort. Furthermore, overall performance was higher for the sub-group of smaller tumours.

The C-indices of the representative models for each tumour sub-volume are shown in Table 3. Among all GTV<sub>entire</sub>-based risk models, the RSF algorithm in combination with the RFVI feature selection method was selected as representative model for further analysis (C-index: 0.75, 95% confidence interval [0.71–0.81]). On the validation cohort, this model achieved a C-index of 0.63 ([0.49–0.67]). The RSF–MIM model trained on the GTV<sub>5mm-rim</sub> was selected as representative model compared to all other rim-based models on the exploratory cohort (C-index: 0.77, [0.72–0.82]). This model attained an improved performance on the validation cohort (C-index: 0.66, [0.52–0.69]), which was slightly higher compared to the GTV<sub>entire</sub>-based model. The corresponding GTV<sub>5mm-core</sub>-based model (RSF–MIM) showed a lower prognostic performance on the validation cohort (C-index: 0.61, [0.49–0.69]) compared to the GTV<sub>5mm-rim</sub> model. Figures A1 and A2 show the prognostic performance for the considered feature selection methods and learning algorithms based

on the  $GTV_{entire}$ ,  $GTV_{3mm-rim}$  and  $GTV_{5mm-rim}$ , as well as the corresponding core sub-volumes on the exploratory and the validation cohorts.

**Table 2.** Median concordance indices (C-index) of radiomic models using the entire gross tumour volume ( $GTV_{entire}$ ) and the different sub-volumes, i.e., tumour rim ( $GTV_{rim}$ ), extended tumour rim ( $GTV_{rim+ext}$ ) and tumour core ( $GTV_{core}$ ) for the endpoint loco-regional tumour control. Results are presented for the exploratory and the validation cohort. Median results over all feature selection methods and learning algorithms are shown (top) as well as C-indices of the representative model combinations and the  $p$ -values of the log-rank tests of stratified patient groups (bottom).

Tumour Sub-Volume	Exploratory Cohort		Validation Cohort	
	All ( $n = 206$ )	All ( $n = 85$ )	$GTV \leq 20 \text{ cm}^3$ ( $n = 20$ )	$GTV > 20 \text{ cm}^3$ ( $n = 65$ )
$GTV_{entire}$	0.75 ± 0.05	0.61 ± 0.04	0.61 ± 0.07	0.59 ± 0.02
$GTV_{3mm-rim}$	0.76 ± 0.06	0.63 ± 0.03	0.62 ± 1.00	0.61 ± 0.02
$GTV_{3mm-core}$	0.74 ± 0.06	0.60 ± 0.02	0.63 ± 0.05	0.57 ± 0.02
$GTV_{5mm-rim}$	0.76 ± 0.06	0.65 ± 0.02	0.67 ± 0.07	0.61 ± 0.01
$GTV_{5mm-core}$	0.72 ± 0.04	0.59 ± 0.01	0.69 ± 0.07	0.57 ± 0.04
$GTV_{5mm-rim+1mm-ext}$	0.76 ± 0.06	0.62 ± 0.03	0.58 ± 0.07	0.65 ± 0.03
$GTV_{5mm-rim+2mm-ext}$	0.76 ± 0.07	0.63 ± 0.03	0.67 ± 0.04	0.66 ± 0.04
$GTV_{5mm-rim+3mm-ext}$	0.75 ± 0.08	0.64 ± 0.05	0.61 ± 0.05	0.65 ± 0.05
$GTV_{5mm-rim+5mm-ext}$	0.75 ± 0.07	0.63 ± 0.04	0.65 ± 0.06	0.62 ± 0.05

GTV: primary gross tumour volume; sd: standard deviation; CI: confidence interval.

**Table 3.** Concordance indices (C-index) of the representative radiomic combinations and the  $p$ -values of the log-rank tests of stratified patient based on the entire gross tumour volume ( $GTV_{entire}$ ) and the different sub-volumes, i.e., tumour rim ( $GTV_{rim}$ ), extended tumour rim ( $GTV_{rim+ext}$ ) and tumour core ( $GTV_{core}$ ) for the endpoint loco-regional tumour control. Results are presented for the exploratory and the validation cohort.

Representative Model	Exploratory Cohort ( $n = 206$ )			Validation Cohort ( $n = 85$ )		
	C-Index	95% CI	$p$ -Value	C-Index	95% CI	$p$ -Value
$GTV_{entire}$ RSF-RFVI	0.75	[0.71–0.81]	<0.001	0.63	[0.49–0.67]	0.012
$GTV_{3mm-rim}$ BT-Cox-MRMR	0.76	[0.71–0.82]	<0.001	0.63	[0.52–0.70]	0.005
$GTV_{3mm-core}$ BT-Cox-MIM	0.75	[0.70–0.80]	<0.001	0.63	[0.50–0.70]	0.069
$GTV_{5mm-rim}$ RSF-MIM	0.77	[0.72–0.82]	<0.001	0.66	[0.52–0.69]	0.006
$GTV_{5mm-core}$ RSF-MIM	0.71	[0.66–0.77]	<0.001	0.61	[0.49–0.69]	0.11
$GTV_{5mm-rim+3mm-ext}$ RSF-RFVI	0.77	[0.69–0.80]	<0.001	0.67	[0.60–0.77]	<0.001

GTV: primary gross tumour volume; sd: standard deviation; CI: confidence interval; RSF: random survival forest; BT-Cox: boosted tree-Cox proportional hazard model; RFVI: random forest variable importance; MRMR: minimum redundancy maximum relevance; MIM: mutual information maximisation.

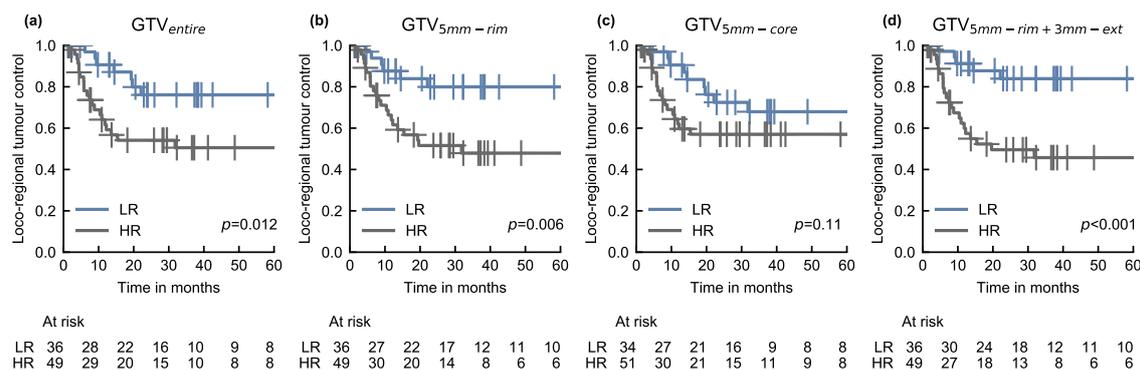
The  $GTV_{5mm-rim}$  sub-volume, which achieved the highest prognostic performance among all rim-based models was subsequently extended by different margins beyond the originally delineated tumour into surrounding tissue. The tumour extensions  $GTV_{5mm-rim+2mm-ext}$  and  $GTV_{5mm-rim+3mm-ext}$  showed the highest median performances on the validation cohort (C-indices:  $0.63 \pm 0.03$  and  $0.64 \pm 0.05$ , respectively). The C-index of the representative models trained on the different tumour extensions are shown in Table 2. The representative model trained on the different tumour extensions are shown in Table 2. The representative model trained on the  $GTV_{5mm-rim+3mm-ext}$  (RSF-RFVI) achieved a slightly better performance in validation compared to the model based on the  $GTV_{entire}$  and on the  $GTV_{5mm-rim}$  (C-index: 0.67, [0.60–0.77]). The resulting C-indices for all

extended sub-volumes and developed radiomic risk models on the exploratory and validation cohort are summarised in Figures A3 and A4, respectively. The hyper-parameters including the optimised values for the representative models are given in Table A4.

### 3.2. Risk-Based Patient Stratification

Patients were stratified into groups at low and high risk for loco-regional recurrence based on the model prediction of the exploratory cohort. Table 3 shows the  $p$ -values of the log-rank test for LRC for all representative models on the validation cohort based on the  $\text{median}_{\text{risk}}$  cut-off. The RSF–RFVI model trained on  $\text{GTV}_{\text{entire}}$  was able to stratify patients into low and high risk groups with a significant difference in LRC ( $p = 0.012$ ). A slightly improved stratification could be achieved by the  $\text{GTV}_{3\text{mm-rim}}$ - and  $\text{GTV}_{5\text{mm-rim}}$ -based models ( $p = 0.005$  and  $p = 0.006$ , respectively) as well as by the extended volume  $\text{GTV}_{5\text{mm-rim}+3\text{mm-ext}}$  ( $p < 0.001$ ). Stratification based on the predicted risk of the corresponding  $\text{GTV}_{5\text{mm-core}}$  model (RSF–MIM) did not lead to significant differences in LRC between both groups ( $p = 0.11$ ).

Figure 2 shows the Kaplan–Meier curves using the  $\text{median}_{\text{risk}}$  cut-off for the representative models based on (a)  $\text{GTV}_{\text{entire}}$ , (b)  $\text{GTV}_{5\text{mm-rim}}$ , (c)  $\text{GTV}_{5\text{mm-core}}$  and (d)  $\text{GTV}_{5\text{mm-rim}+3\text{mm-ext}}$ , respectively, for the validation cohort. The resulting  $p$ -values for all considered sub-volumes and developed radiomic risk models on the validation cohort are summarised in Figures A5 and A6.



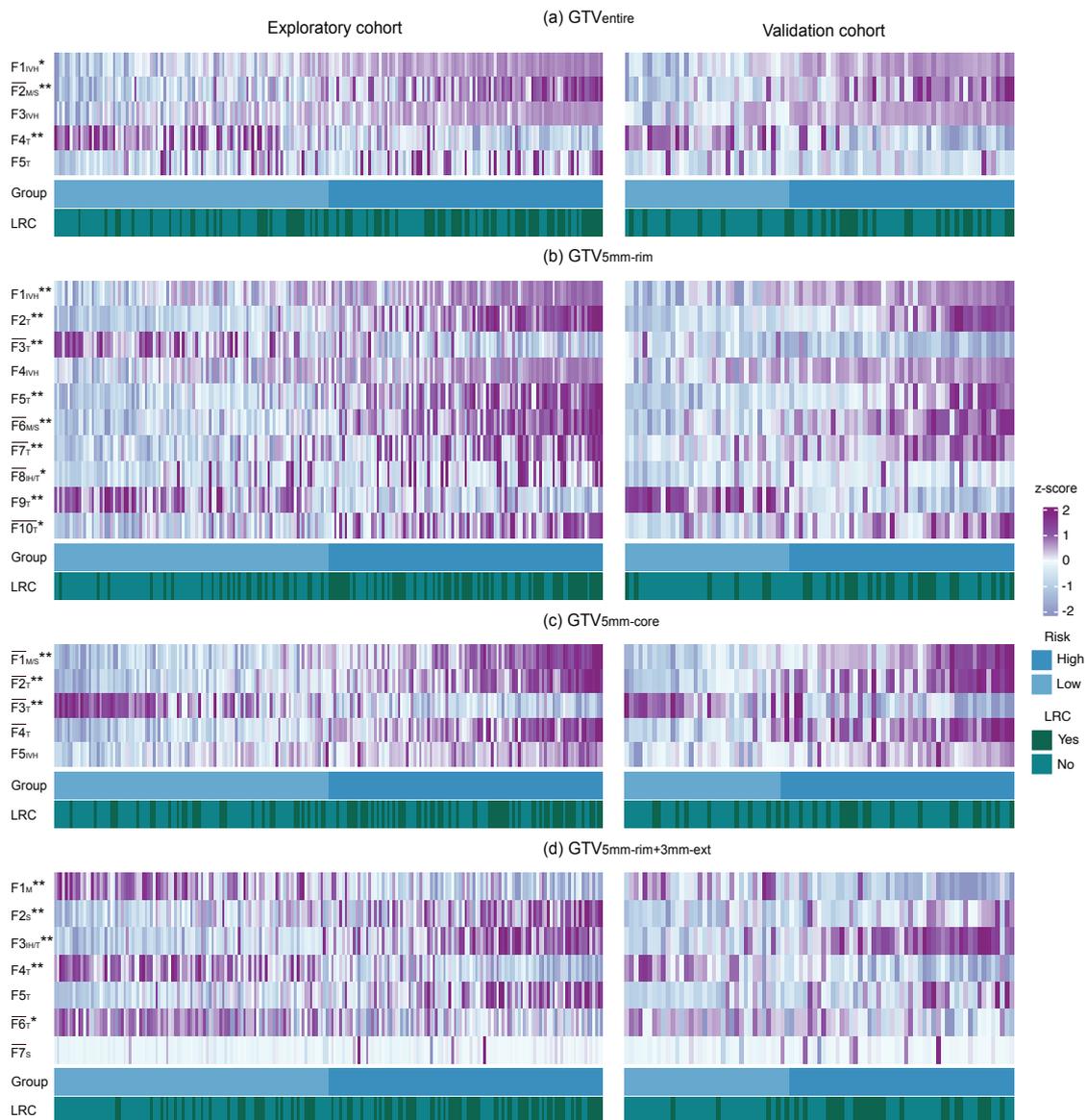
**Figure 2.** Kaplan–Meier curves for the prediction of loco-regional tumour control of the representative models based on the (a) entire primary gross tumour volume ( $\text{GTV}_{\text{entire}}$ ); (b) 5 mm rim of the tumour ( $\text{GTV}_{5\text{mm-rim}}$ ); (c) corresponding tumour core ( $\text{GTV}_{5\text{mm-core}}$ ) and (d) 3 mm extension of the 5 mm tumour rim ( $\text{GTV}_{5\text{mm-rim}+3\text{mm-ext}}$ ) sub-volumes for patients of the validation cohort. Patients were stratified into low (LR) and high (HR) risk groups based on the median risk of loco-regional recurrence determined on the exploratory cohort.

### 3.3. Radiomic Signature Analysis

Radiomics signatures were investigated for the representative models based on  $\text{GTV}_{\text{entire}}$ ,  $\text{GTV}_{5\text{mm-rim}}$ ,  $\text{GTV}_{5\text{mm-core}}$  and  $\text{GTV}_{5\text{mm-rim}+3\text{mm-ext}}$ . Figure 3 shows the feature expressions of the developed signatures for each patient in a heatmap. Image features within the signatures are listed in Table A5.

The developed signatures for the different models (a)–(d) consist of two to ten imaging features extracted from the original and wavelet transformed images. The selected features typically comprise first-order statistical or texture-based features. For instance, the ‘statistics energy’ feature, which describes the overall density of the tumour volume, appears in all four signatures as a single or as a feature within a cluster [2]. Furthermore, the signatures of the trained models (a) and (b) consist of the same intensity-volume histogram feature (i.e., ‘ivh\_diff\_v10\_v90’) computed and extracted from the wavelet transformed images. This feature describes the difference between the largest volume fractions at two different intensity values of at least 10% and 90% [25,28]. The selected features

for (a) and (b) were mostly based on the low-pass wavelet transformed images, which may contain reduced noise. Features within the signatures (c) and (d) were mostly computed on the high-pass wavelet transformed images, which may characterise edges and blobs within the considered regions. For all developed signatures (a)–(d), almost all features were significantly associated with LRC based on univariate Cox analyses using the entire patient cohort.



**Figure 3.** Heatmaps showing different expression patterns of the radiomic features of the developed signatures for the representative models based on the (a) entire primary gross tumour volume (GTV<sub>entire</sub>); (b) 5 mm rim of the tumour (GTV<sub>5mm-rim</sub>); (c) corresponding tumour core (GTV<sub>5mm-core</sub>); and (d) 3 mm extension of the 5 mm tumour rim (GTV<sub>5mm-rim+3mm-ext</sub>) sub-volumes. Feature expression values are sorted according to the predicted risk and the risk group based on the determined median<sub>risk</sub> cut-off values. Loco-regional tumour control (LRC) during follow-up (yes, light; no, dark) and features with a significant association with LRC are shown (\*  $p < 0.05$  and \*\*  $p < 0.001$ ). A detailed description of the feature abbreviations can be found in Table A5. Abbreviations:  $\bar{F}$  cluster feature consisting of several features represented by the mean value as a new meta-feature,  $F_S$  first order statistical feature,  $F_M$  morphological,  $F_{IH}$  intensity histogram,  $F_{IVH}$  intensity volume histogram and  $F_T$  texture feature.

#### 4. Discussion

Tumours may contain biologically complex structures and exhibit substantial spatial variation. Thus, the main objective of this study was to compare radiomic models based on different sub-volumes of the tumour, i.e., on the tumour rim, the tumour core and the macroscopic tumour extensions, in order to identify potential regions containing the most relevant prognostic information for LRC. Using CT imaging of patients with locally advanced HNSCC revealed that radiomic risk models based on tumour rim sub-volumes achieved a slightly improved prognostic performance and better patient stratification compared to models based on the corresponding core regions. Furthermore, sub-group analyses showed that the differences in prognostic performance between rim and core regions were larger for large tumours compared to small tumours. In general, our analysis showed a good median performance and a better patient stratification for the models based on the  $GTV_{5mm-rim}$ , while the corresponding core-based models performed slightly less. This may indicate that the tumour rim contains more prognostic information. The statistical comparison between both sub volumes led to a borderline statistical trend (MLA:  $p = 0.10$ ), i.e., the presented findings require additional validation in the future.

These results are in-line with previously published data of other tumour entities [10,11,29]. For example, Dou et al. [30] showed that models based on CT-imaging features of the 3 mm rim of the GTV lead to an improved prediction of distant metastasis compared to the model based on the entire GTV for patients with locally advanced non-small cell lung cancer (NSCLC). Furthermore, Hosney et al. [31] developed a deep learning-based prediction model using a 3D convolutional neuronal network for the prediction of OS for NSCLC patients and observed that the network tended to focus on the interface between the tumour and stroma (parenchyma or pleura) regions in the CT images. In contrast to that, Keek et al. [32] concluded that the consideration of the tumour rim did not lead to an improved prediction of overall survival, loco-regional recurrence and distant metastases in stage III and IV HNSCC patients. However, for the prediction of loco-regional recurrence, a better prediction could be observed for the 5 mm rim-based model compared to the model using the  $GTV_{entire}$  in the exploratory and validation cohort (C-index: 0.86/0.59 and 0.81/0.52, respectively). In addition, Grove et al. [18] showed that tumour-rim-based radiomic features (i.e., entropy) were higher expressed compared to features extracted from corresponding tumour-core sub-volumes in NSCLC patients. While, the entropy feature and their ratios of core and rim regions were associated with overall survival in the exploratory cohort, but not in the independent validation cohort.

The biological characteristics of the tumour rim were already discussed by published data from the Danish Head and Neck Cancer (DAHANCA) group [33]. Based on experience from pathological examination of surgical resections, the DAHANCA group concluded that for primary tumours, the risk of sub-clinical microscopic spread was around 50% of which more than 99% was within 5 mm and 95% within 4 mm of the rim of the primary tumour. Furthermore, Apolle et al. [22] showed that most solid tumours exhibit microscopic tumour extension in particular for head and neck cancer. Our findings suggest that biological processes such as microscopic spread capacity are associated with macroscopic CT imaging.

Defining the precise extent of the macroscopic tumour prior to and during RCT is difficult, especially using CT imaging without contrast enhancement [34]. Slight extensions of the delineated tumour volume into normal tissue did not reduce the performance of the radiomic risk models, which indicates that these regions may also contain prognostic information. In addition, slight extensions of the tumour may be useful for assessing feature stability, simulating different tumour delineations of different observers [35]. Furthermore, uncertainties in the delineation of the  $GTV_{entire}$  may affect radiomic features and in turn the results of radiomic analyses. In the current study, the  $GTV_{entire}$  was manually delineated by one expert radiation oncologist. The consideration of multiple tumour delineations of different experts or the usage of semi-automatic segmentation algorithms as well as contour randomisation techniques may help to increase the robustness of the radiomics features and improve the corresponding risk model performance, which should be investigated in the future [35–37].

The presented study is motivated from the assumption that hypoxic or necrotic regions preferentially appear in the tumour core due to inadequate vascular supply, and that proliferating cancer cells mainly occur in the tumour periphery [38]. Our retrospective patient cohort contains tumours with a wide range of different volumes. While necrotic and hypoxic regions will be minimal in small tumours they may be substantial in larger tumours, i.e., the prognostic value of tumour core and rim may change depending on the tumour volume. Therefore, we performed a subgroup analysis considering patients with small and large tumours separately. We found larger differences in prognostic performance between rim and core regions for larger tumours in validation, supporting this hypothesis. Still, the inclusion of patients with small and large tumours in our main analyses may affect the difference in performance between the rim- and core-based risk models. Moreover, necrotic/hypoxic regions may be heterogeneously distributed in the tumour and not be sufficiently captured by our simple approach of defining the tumour core and rim, which in addition does not consider other complex spatial and temporal variations in the tumour micro-environment. This may lead to smaller observable differences in the performance between the rim- and core-based models [14,39]. The identification and incorporation of tumour specific regional variations by more sophisticated image analysis techniques may help to overcome this gap. For instance, differential information from multi-modal imaging data, such as PET-CT or functional MRI may be used. Moreover, super-voxel algorithms can be applied to group voxels into super-voxel segments based on their grey value, e.g., using the FDG uptake value [40]. Subsequently, the resulting super-voxel segments can be further merged to generate tumour sub-volumes, e.g., by hierarchical or fuzzy c-means clustering algorithms across the entire patient cohort. Wu et al. [19] proposed such a two-stage clustering process, for the identification and determination of sub-volumes based on CT imaging combined with FDG-PET scans in lung cancer patients. Furthermore, the consideration of regions with temporal changes, e.g., due to RCT-induced tumour shrinkage or re-oxygenation using in-treatment CT images in combination with functional imaging may offer the potential to enhance radiomic risk models in future [5,39]. However, due to missing functional imaging, it was not possible to use such imaging data in this study.

In addition to radiomic features, clinical parameters may be relevant for the prediction of treatment outcome. On our cohort, from the parameters shown in Table 1, only the primary tumour volume and the derived tumour sub-volumes were significantly related to LRC ( $p < 0.01$ ). These parameters revealed C-indices between 0.62 and 0.63 in the validation cohort using univariable Cox regression model. This was slightly lower than observed for the presented radiomic models based on the  $GTV_{5mm-rim}$  and the  $GTV_{5mm-rim+3mm-ext}$  (C-index: 0.65 and 0.67, respectively). While the radiomic signature based on the  $GTV_{5mm-rim}$  contained two CT features with a strong Spearman correlation ( $\rho$ ) to the tumour volume ( $\rho > 0.85$ ), the features of the signature based on the  $GTV_{5mm-rim+3mm-ext}$  showed only moderate correlations to the tumour volume ( $\rho$  range:  $[-0.61-0.40]$ ). This indicates that additional imaging features, which are not related to the tumour volume, may improve the risk model performance.

One limitation of this retrospective study is the different distribution of the clinical characteristics between the exploratory and validation cohort, e.g., in tumour site and UICC stage (Table 1). Despite these differences, the validation of the presented radiomic models was successful, and due to the definition of both cohorts that was based on independent clinical trials, the presented results should be more robust compared for example to a random split of the data. Furthermore, other factors related to the retrospective nature of our study may have implications on the presented results [41]. For instance, the variety of different CT imaging acquisition and reconstruction parameters may affect the feature robustness and thereby the results of risk modelling (Table A1). Therefore, open and standardised protocols for image acquisition, reconstruction, and analysis may help to increase the robustness of radiomic risk model [42–44]. In addition, the biological meaning of the selected imaging features within the developed signatures and the differences between the features of the tumour rim and core remains still unclear. Therefore, these open questions should be investigated systematically in the future for a better understanding of the underlying mechanisms.

## 5. Conclusions

In the present study, we showed that radiomic models based on the rim of locally advanced HNSCC achieved a slightly higher prognostic performance for LRC after primary radio-chemotherapy compared to models using the tumour core. This supports our initial hypothesis that the tumour rim is biologically more diverse and important treatment-related processes occur primarily in this region, which may be visible in clinical imaging data. Therefore, after additional prospective validation the consideration of tumour sub-volumes may be a promising way to improve prognostic radiomic risk models.

**Author Contributions:** Conceptualisation, S.L. (Stefan Leger), S.L. (Steffen Löck), M.B., M.K., C.R. and E.G.C.T.; methodology, S.L. (Stefan Leger), S.L. (Steffen Löck) and A.Z.; software, S.L. (Stefan Leger) and A.Z.; validation, S.L. (Stefan Leger), S.L. (Steffen Löck) and E.G.C.T.; investigation, S.L. (Stefan Leger); resources, S.L. (Steffen Löck); data curation, K.L., E.G.C.T., F.L., A.L., A.S., G.K., I.T., N.G., M.G., P.B., J.v.d.G., U.G., C.B., J.C.P., S.E.C., S.B. and D.Z.; writing—original draft preparation, S.L. (Stefan Leger); writing—review and editing, S.L. (Steffen Löck), A.Z., E.G.C.T., A.L., A.S., G.K., I.T., N.G., M.G., P.B., J.v.d.G., U.G., C.B., J.C.P., S.E.C., S.B., D.Z. and C.R.; visualisation, S.L. (Stefan Leger); supervision, S.L. (Steffen Löck) and E.G.C.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** The author S.L. was supported by the Federal Ministry of Education and Research (BMBF-13GW0211D).

**Conflicts of Interest:** In the past 5 years, Michael Baumann attended an advisory board meeting of MERCK KGaA (Darmstadt), for which the University of Dresden received a travel grant. He further received funding for his research projects and for educational grants to the University of Dresden by Teutopharma GmbH (2011–2015), IBA (2016), Bayer AG (2016–2018), Merck KGaA (2014–open), Medipan GmbH (2014–2018). He is on the supervisory board of HI-STEM gGmbH (Heidelberg) for the German Cancer Research Center (DKFZ, Heidelberg) and also member of the supervisory body of the Charité University Hospital, Berlin. As former chair of OncoRay (Dresden) and present CEO and Scientific Chair of the German Cancer Research Center (DKFZ, Heidelberg), he has been or is still responsible for collaborations with a multitude of companies and institutions, worldwide. In this capacity, he discussed potential projects with and has signed/signs contracts for his institute(s) and for the staff for research funding and/or collaborations with industry and academia, worldwide, including but not limited to pharmaceutical corporations like Bayer, Boehringer Ingelheim, Bosch, Roche and other corporations like Siemens, IBA, Varian, Elekta, Bruker and others. In this role, he was/is further responsible for commercial technology transfer activities of his institute(s), including the DKFZ-PSMA617 related patent portfolio [WO2015055318 (A1), ANTIGEN (PSMA)] and similar IP portfolios. Baumann confirms that, to the best of his knowledge, none of the above funding sources was involved in the preparation of this paper. In the past 5 years, Krause received funding for her research projects by IBA (2016), Merck KGaA (2014–2018 for preclinical study; 2018–2020 for clinical study), Medipan GmbH (2014–2018) and by the Gert and Susanna Mayer Foundation (2019–2022). She is involved in an ongoing publicly funded (German Federal Ministry of Education and Research) project with the companies Medipan, Attomol GmbH, GA Generic Assays GmbH, Gesellschaft für medizinische und wissenschaftliche genetische Analysen, Lipotype GmbH and PolyAn GmbH (2019–2021). For the present manuscript, Krause confirm that none of the above mentioned funding sources were involved. In the past 5 years, Troost received funding for her research projects by Merck KGaA (2018–2020 for clinical study). She is involved in an ongoing publicly funded (German Federal Ministry of Education and Research) project with the companies Medipan, Attomol GmbH, GA Generic Assays GmbH, Gesellschaft für medizinische und wissenschaftliche genetische Analysen, Lipotype GmbH and PolyAn GmbH (2019–2021). For the present manuscript, Troost confirm that none of the above mentioned funding sources were involved. In the past 5 years, Richter and his institution OncoRay received funding from Siemens Healthineers for research collaborations and as reference center as well as lecturer. For the present manuscript, Richter confirms that none of the above mentioned funding sources were involved. D.Z., S.B.: The Depart. of Radiation Oncology Tübingen receives support through research, travel and training grants from DFG, Elekta, Philips, Siemens, Therapanacea and Sennewald. Non of these sources are relevant to the presented project. Linge is involved in an ongoing publicly funded (German Federal Ministry of Education and Research) project with the companies Medipan, Attomol GmbH, GA Generic Assays GmbH, Gesellschaft für medizinische und wissenschaftliche genetische Analysen, Lipotype GmbH and PolyAn GmbH (2019–2021). For the present manuscript, Linge confirms that this above mentioned funding source was not involved. Tinhofer served as advisor and gave lectures for Merck-Serono. For the present manuscript, Tinhofer confirms that this above mentioned funding source was not involved in the preparation of this paper.

## Abbreviations

The following abbreviations are used in this manuscript:

BGLM-Cox	boosted gradient linear model-Cox
BT-Cox	boosted tree-Cox
C-index	concordance index

Cox	Cox proportional hazard model
CT	computed tomography
DKTK-ROG	German Cancer Consortium Radiation Oncology Group
FDG	<sup>18</sup> F-fluorodeoxyglucose
FMISO	<sup>18</sup> F-fluoromisonidazole
GTV	gross tumour volume
HNSCC	head and neck squamous cell carcinoma
MIFS	mutual information feature selection
MIM	mutual information maximisation
MLA	multi-level model approach
MRI	magnetic resonance imaging
MRMR	minimum redundancy maximum relevance
MSR-RF	maximally selected rank statistics random forest
OS	overall survival
PET	positron emission tomography
RCT	radio-chemotherapy
RFVI	random forest variable importance
RSF	random survival forest
UKD	University Hospital Dresden

## Appendix A

### Appendix A.1. Multi-Level Model

The multi-level model approach (MLA) was developed for assessing the differences in concordance index between the rim-based and the corresponding core-based models independent from the effects of the feature selection methods and learning algorithms, similar as in [5]. For this purpose, we defined a multi-level model consisting of three levels (L) representing the different factors:

L0 (volume):

$$y_{i,j} = \alpha_{\text{method},i,j} + \beta_{\text{vol}}x_{\text{vol}} + \varepsilon_{\text{vol}}$$

$$\varepsilon_{\text{vol}} \sim \mathcal{N}(0, \sigma_{\text{vol}}^2),$$

L1 (machine learning algorithms):

$$\alpha_{\text{method},i,j} = \alpha_{fs,i} + \beta_{\text{learner},j} + \varepsilon_{\text{learner},j}$$

$$\varepsilon_{\text{learner},j} \sim \mathcal{N}(0, \sigma_{\text{learner},j}^2),$$

L2 (feature selection methods):

$$\alpha_{fs,i} = \beta_{fs,i} + \varepsilon_{fs,i}$$

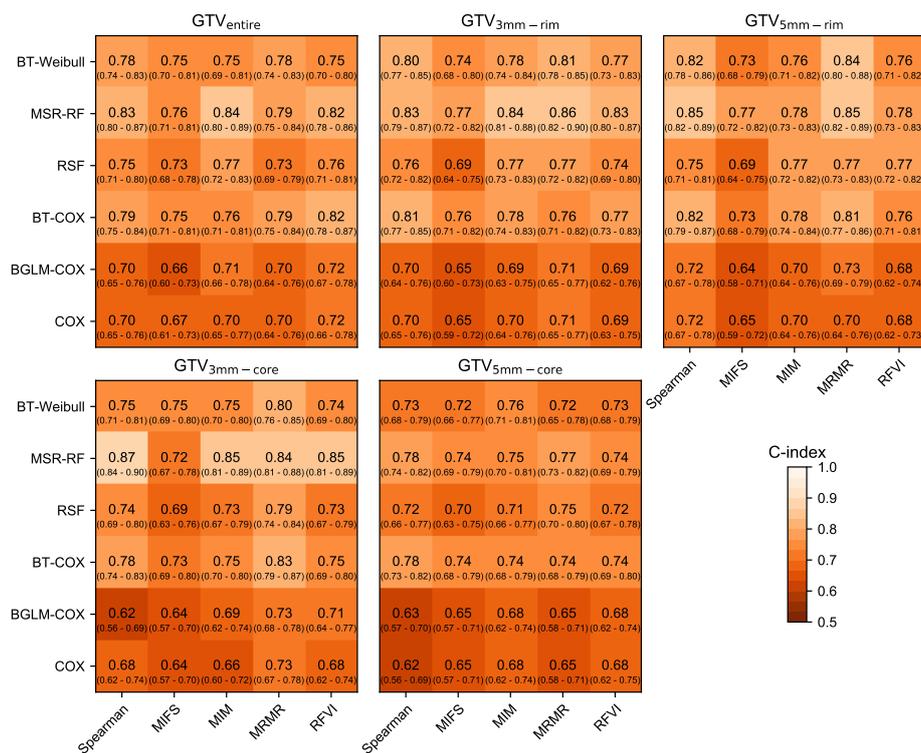
$$\varepsilon_{fs,i} \sim \mathcal{N}(0, \sigma_{fs,i}^2). \tag{A1}$$

In the multi-level model, the top level describes the effect of the tumour sub-volume.  $y_{ij}$  is the concordance index of a bootstrap sample using feature selection method  $i$  and learning algorithm  $j$ .  $\alpha_{\text{method},i,j}$  is an offset term, modelled separately in levels L1 and L2.  $x_{(\text{vol})}$  is a contrast variable, which has the values 0 or 1 for the two specific sub-volume that are to be compared, e.g., rim versus core.  $\beta_{(\text{vol})}$  is the effect of the specific sub-volume compared to the other sub-volume and has a weakly informative prior  $\mathcal{N}(0, 1)$  and is limited to the range  $[-1, 1]$ . The error term  $\varepsilon_{\text{vol}}$  is modelled with a normal distribution with mean 0 and standard deviation  $\sigma_{\text{vol}} \in [0, 1]$ . Level L1 models the effect of learner  $j$  with feature selection method  $i$ .  $\alpha_{fs,i}$  is an offset modelled separately in L2.  $\beta_{\text{learner},j}$  is the effect of learner  $j$ . It has a weakly informative prior  $\mathcal{N}(0, 1)$  and is limited to the range  $[-1, 1]$ . The error term  $\varepsilon_{\text{learner},j}$  is modelled with a normal distribution with mean 0 standard deviation  $\sigma_{\text{learner},j}$ , with  $\sigma_{\text{learner},j} \in [0, 1]$ . Level L2 models the effect of feature selection method  $i$ ,  $\beta_{fs,i}$ .  $\beta_{fs,i}$  has a weakly informative prior  $\mathcal{N}(0.5, 1)$ , as a concordance index takes the baseline value of 0.5 for random data,

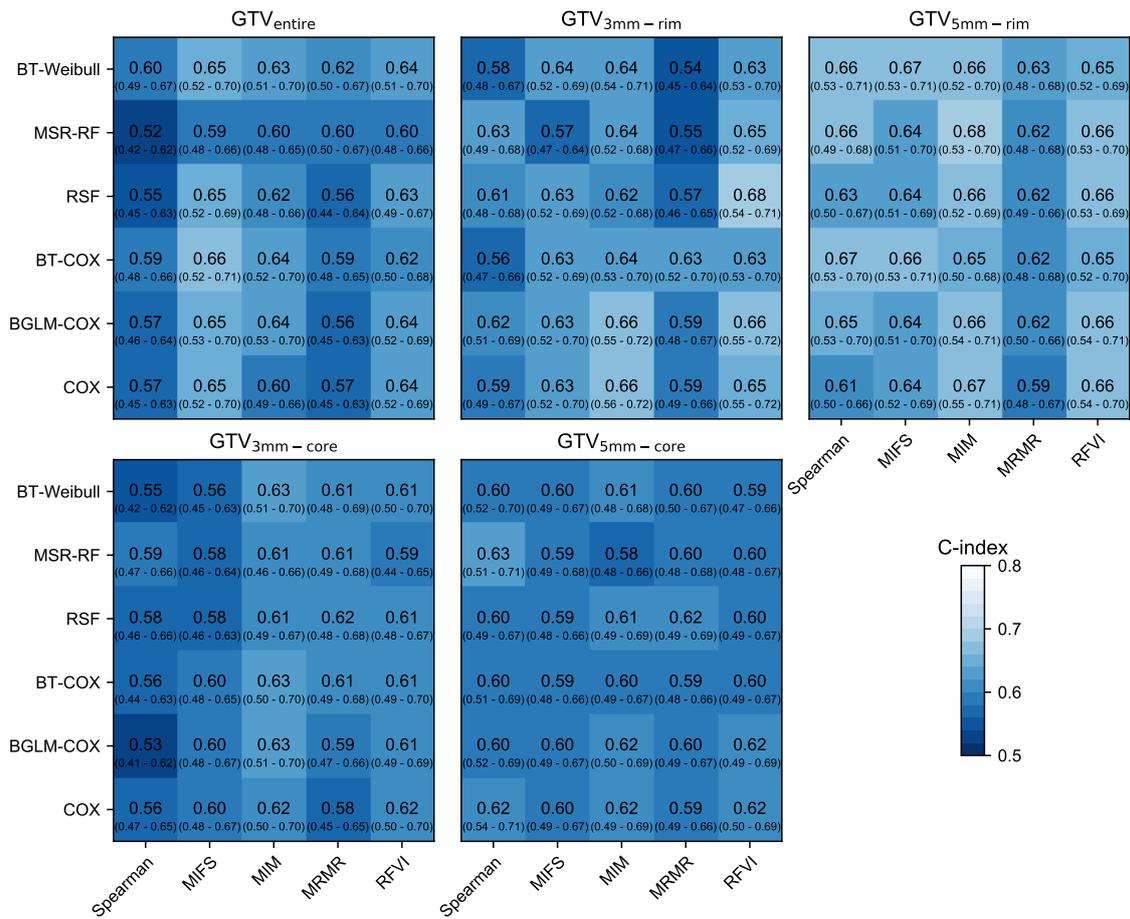
and is limited to the range  $[0, 1]$ . The error term  $\epsilon_{fs,i}$  is modelled with a normal distribution with mean 0 standard deviation  $\sigma_{fs,i}$  with  $\sigma_{fs,i} \in [0, 1]$ . The model was fitted using Markov chain Monte Carlo in STAN 2.21.2 (Stan Development Team (2020). RStan: the R interface to Stan. R package version 2.21.1. <http://mc-stan.org/>), using 7 chains with 500 warm-up iterations and 500 sample iterations each. Model convergence was checked using the R-hat statistic.

**Table A1.** Computed tomography acquisition and reconstruction settings for the exploratory and the validation cohort.

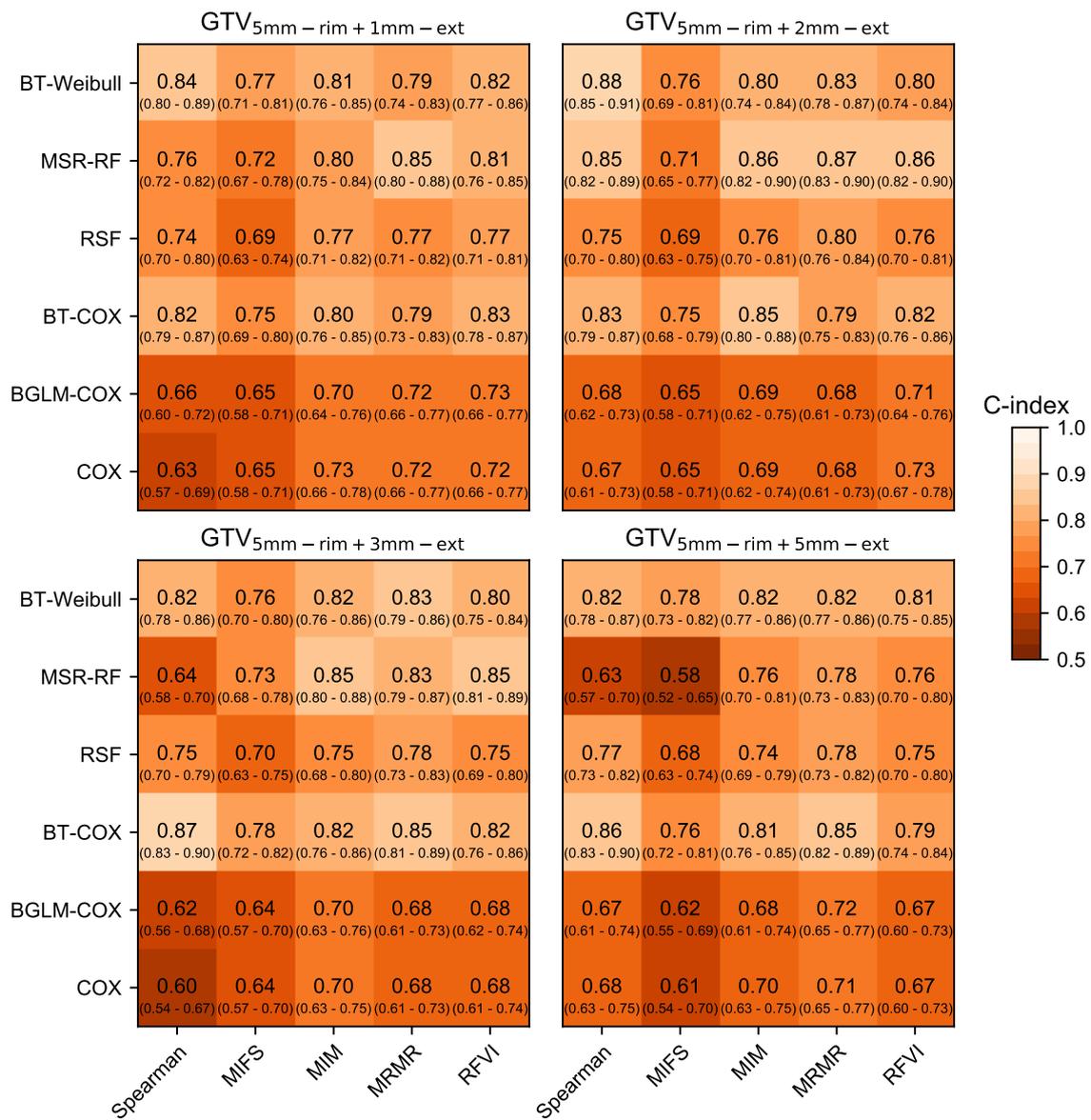
Image Acquisition Parameters	Exploratory Cohort (n = 206)	Validation Cohort (n = 85)
Voxel spacing in-plane (mm)		
0.85/0.87/0.88/0.90	1/2/1/1	0/0/0/0
0.92/0.93/0.94/0.96	1/1/3/2	0/0/0/0
0.97/0.98/1.17/1.27/1.36	3/113/21/26/29	0/13/0/14/58
Spacing in z-direction (mm)		
2.0/2.5/3.0/3.75/5.0	36/22/74/1/63	0/0/27/0/58
Reconstruction kernel		
B10s/B20f/s/B30f/s/B31f/s	20/3/1/2/29/19/16	1/51/1/0/0/12/0
B40f/s/B50s/Missing	1/1/9/12/93	0/0/0/0/20
Mean exposure (mA)	181.3 (Missing: 59)	76.8 (Missing: 14)
Mean exposure time (ms)	733.8 (Missing: 59)	508.8 (Missing: 14)
Tube voltage (kV)		
120/130/140/Missing	86/9/16/95	71/0/0/14



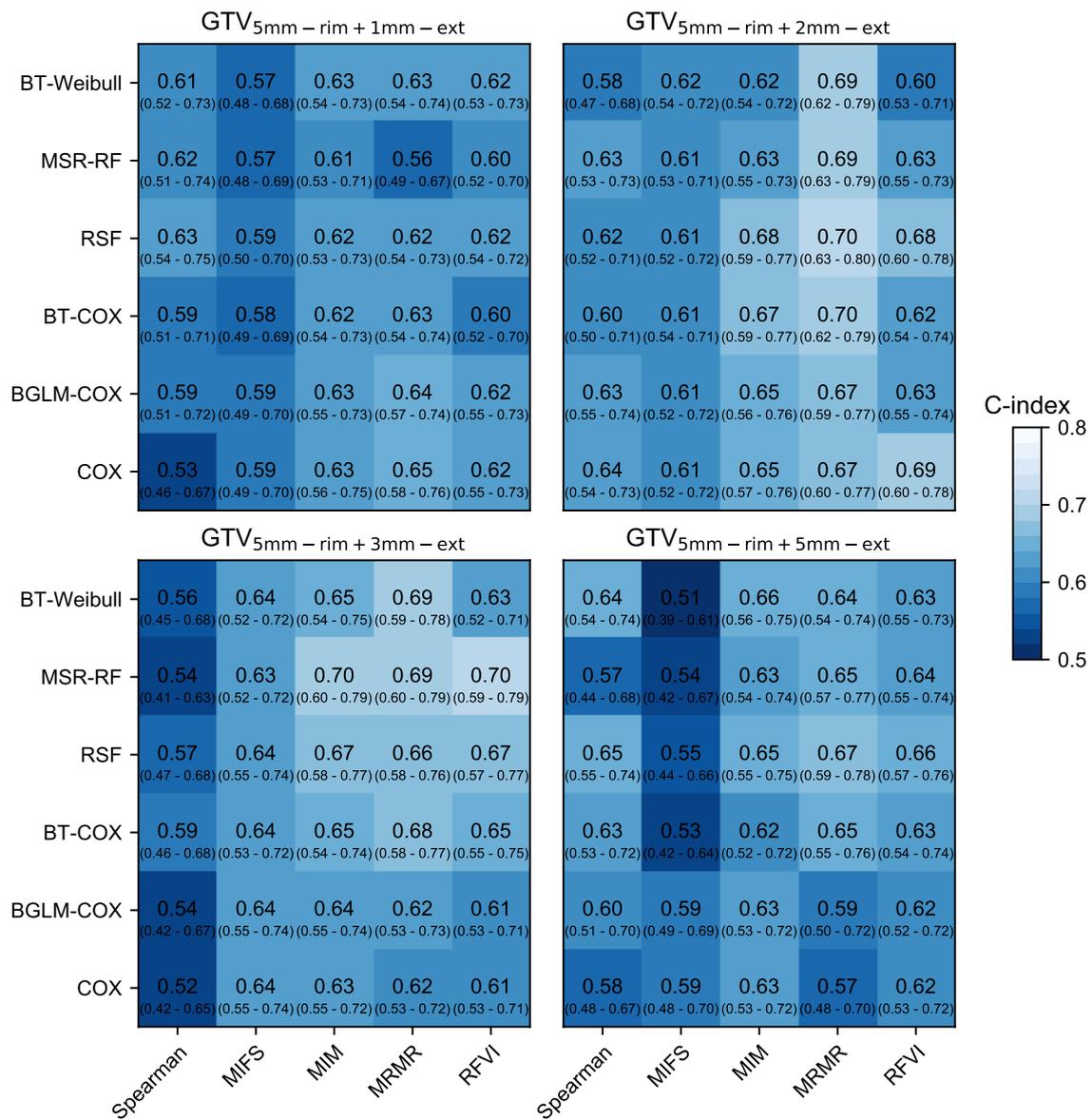
**Figure A1.** Concordance indices (C-index) for the feature selection methods (columns) and the learning algorithms (rows) trained on the  $GTV_{entire}$ , the  $GTV_{3mm-rim}$ , the  $GTV_{5mm-rim}$  and the corresponding core sub-volumes on the exploratory cohort. Furthermore, the 95% confidence intervals for each model combination are shown (in parentheses).



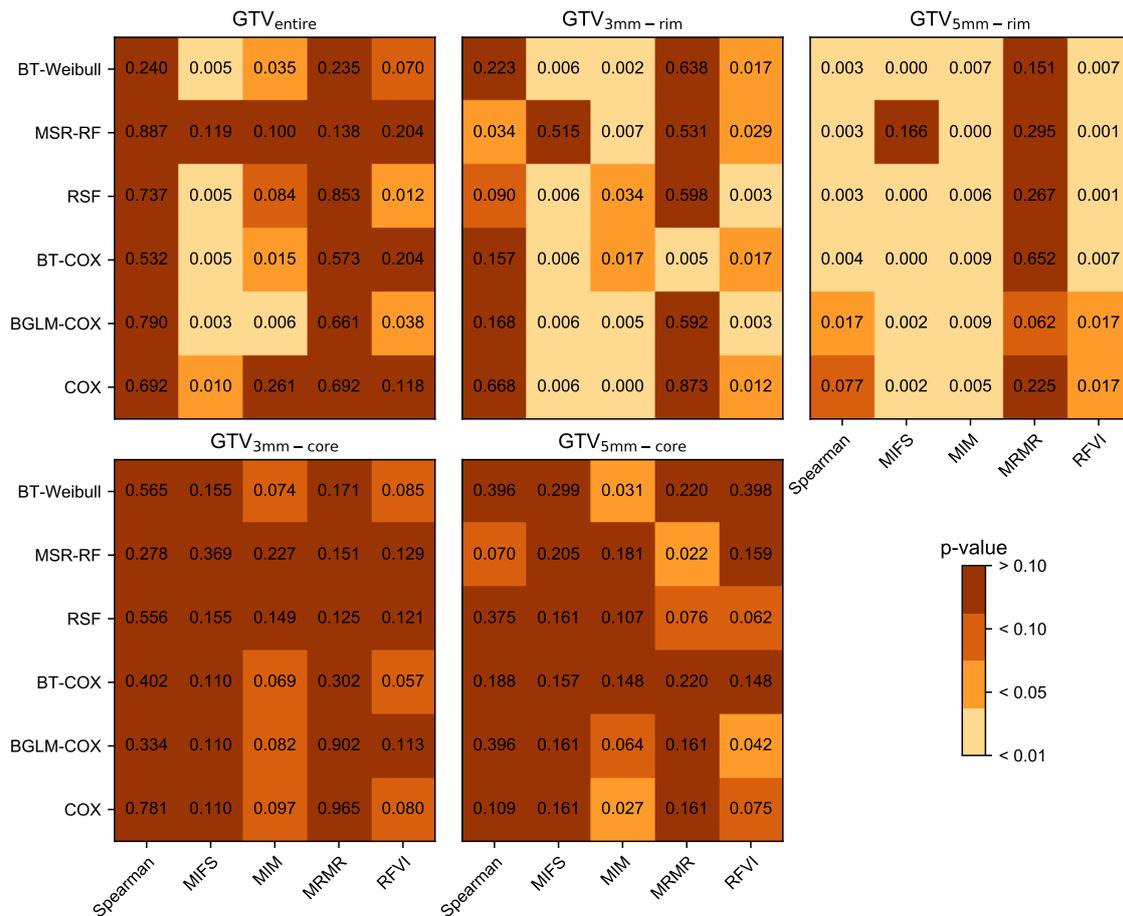
**Figure A2.** Concordance indices (C-index) for the feature selection methods (columns) and the learning algorithms (rows) trained on the GTV<sub>entire</sub>, the GTV<sub>3mm-rim</sub>, the GTV<sub>5mm-rim</sub> and the corresponding core sub-volumes on the validation cohort. Furthermore, the 95% confidence intervals for each model combination are shown (in parentheses).



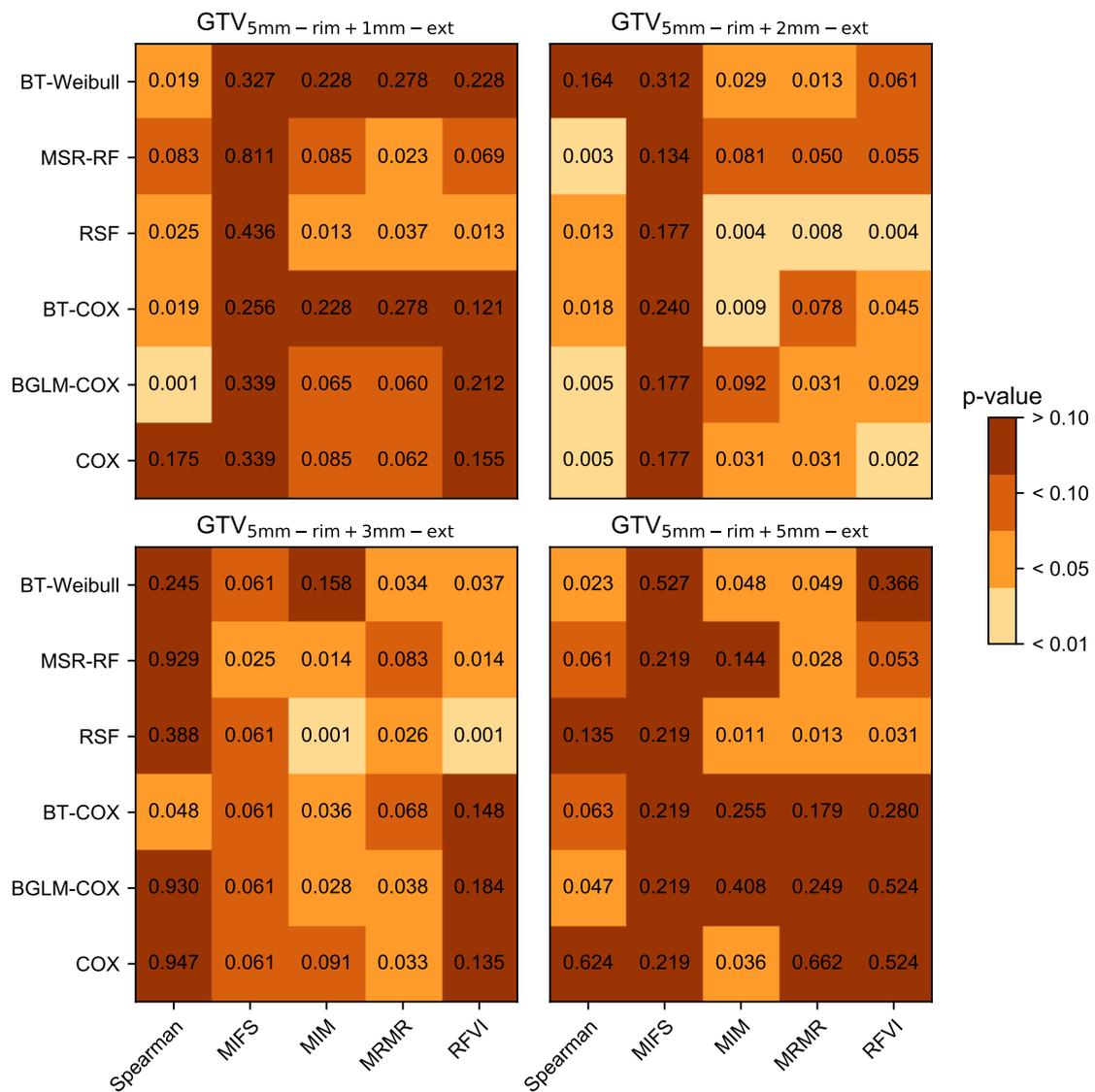
**Figure A3.** Concordance indices (C-index) for the feature selection methods (columns) and the learning algorithms (rows) trained on the tumour extension  $GTV_{5mm-rim+1mm-ext}$ ,  $GTV_{5mm-rim+2mm-ext}$ ,  $GTV_{5mm-rim+3mm-ext}$  and  $GTV_{5mm-rim+5mm-ext}$  sub-volumes for the exploratory cohort. Furthermore, the 95% confidence intervals for each model combination are shown (in parentheses).



**Figure A4.** Concordance indices (C-index) for the feature selection methods (columns) and the learning algorithms (rows) trained on the tumour extension  $GTV_{5mm-rim+1mm-ext}$ ,  $GTV_{5mm-rim+2mm-ext}$ ,  $GTV_{5mm-rim+3mm-ext}$  and  $GTV_{5mm-rim+5mm-ext}$  sub-volumes for the validation cohort. Furthermore, the 95% confidence intervals for each model combination are shown (in parentheses).



**Figure A5.** Resulting *p*-values of the log-rank tests for loco-regional tumour control in the validation cohort for the feature selection methods (columns) and the learning algorithms (rows) trained on the GTV<sub>entire</sub>, the GTV<sub>3mm-rim</sub>, the GTV<sub>5mm-rim</sub> and the corresponding core sub-volumes based on median<sub>risk</sub> cut-off values using the predicted risk values. The cut-off values used for stratification were determined on the exploratory cohort and applied to the validation cohort unchanged.



**Figure A6.** Resulting  $p$ -values of the log-rank tests for loco-regional tumour control on the validation cohort for the feature selection methods (columns) and the learning algorithms (rows) trained on the GTV<sub>5mm-rim+1mm-ext</sub>, the GTV<sub>5mm-rim+2mm</sub>, the GTV<sub>5mm-rim+3mm-ext</sub> and the GTV<sub>5mm-rim+5mm-ext</sub> based on median<sub>risk</sub> cut-off values using the predicted risk values. The cut-off values used for stratification were determined on the exploratory cohort and applied to the validation cohort unchanged.

**Table A2.** Configuration settings for the image feature computation and extraction.

Configuration Settings	Configuration Value(s)
Image interpolation	
Interpolation method	Cubic spline
Voxel dimensions	$1.0 \times 1.0 \times 1.0 \text{ mm}^3$
Anti-aliasing smoothing parameter $\beta$	0.98
RoI interpolation	
Interpolation method	Cubic spline
Inclusion threshold	0.5
Discretisation	
Discretisation method	Fixed Bin Number (FBN) of 32 bins
Intensity Volume Histogram discretisation method	Fixed Bin Number (FBN) of 1000 bins
Image transformation	
Wavelet	coiflet-1
Mean-Intensity Laplacian of Gaussian	kernel widths: 1.0, 2.0, 3.0, 5.0, 6.0 mm
Texture matrices	
Grey-level Run Length Matrix (GLRLM)	Calculation method: 3D Merge method: volume merge
Grey-level Size Zone Matrix (GLSZM)	Calculation method: 3D
Neighbourhood Grey Tone Difference Matrix (NGTDM)	Calculation method: 3D Distance for neighborhood: 1.8 voxels
Neighbourhood Grey Level Dependence Matrix (NGLDM)	Difference level: 0.0 Calculation method: 3D Distance for neighborhood: 1.0 voxels
Grey Level Co-occurrence Matrix (GLCM)	Calculation method: 3D Merge method: volume merge
Grey Level Distance Zone Matrix (GLDZM)	Calculation method: 3D

RoI: region of interest.

**Table A3.** Definition of the hyper-parameters of the feature selection methods and of the machine learning algorithms, which were used during the hyper-parameter optimisation. The hyper-parameters for the feature selection methods were kept fixed and not optimised during hyper-parameter optimisation.

Algorithm	Hyper-Parameter Name	Hyper-Parameter Value(s)
Cox proportional hazard model	Signature size	2, 3, 4, 5, 7, 10
Boosted tree and boosted gradient models	Signature size:	2, 3, 4, 5, 7, 10
	$\alpha$ :	0.001, 0.01, 0.05
	$\omega$ :	lambda.min <sup>1</sup> , lambda.1se <sup>2</sup>
	mStop:	200
Random survival forest	Signature size:	2, 3, 4, 5, 7, 10
	ntree:	2000, 5000
	mtry:	100
	node-Size:	25–50, step size 1
	maxDepth:	10, 15, 20, 25, 40
	nSplit:	1, 2, 100
	splitRule:	logrank, logrankscore
Maximally selected rank statistics random forest	Signature size:	2, 3, 4, 5, 7, 10
	ntree:	2000, 5000
	mtry:	100
	node-Size:	25–50, step size 1
	minprop:	0.1
	$\alpha$ :	0.1, 0.5
	splitRule:	C, maxstat

Table A3. Cont.

Algorithm	Hyper-Parameter Name	Hyper-Parameter Value(s)
Minimum redundancy maximum relevance	topFeatures	100
	RelativeImportanceThreshold	0
Mutual information feature selection	topFeatures	100
	RelativeImportanceThreshold	0.05
Random forest variable importance	topFeatures	20
	nTree	1000
	K	2
	nRepetition	50
	nSteps	2
	nSplits	1
	splitrule	logrank
	nodeSize	45
	mTry	500
nVariables	10	

<sup>1</sup> Minimum mean cross-validated error; <sup>2</sup> Error within one standard error of the minimum.

Table A4. Hyper-parameters and their optimised values for the representative models.

Sub-Volume	Hyper-Parameter Value(s)
GTV <sub>entire</sub> RSF-RFVI	Signature size: 5 ntree: 5000 mtry: 100 node-Size: 10 maxDepth: 25 nSplit: 1 splitRule: logrankscore
GTV <sub>5mm-rim</sub> RSF-MIM	Signature size: 10 ntree: 5000 mtry: 100 node-Size: 5 maxDepth: 40 nSplit: 1 splitRule: logrank
GTV <sub>5mm-core</sub> RSF-MIM	Signature size: 5 ntree: 5000 mtry: 100 nodeSize: 7 maxDepth: 40 nSplit: 1 splitRule: logrank
GTV <sub>5mm-rim+3mm-ext</sub> RSF-RFVI	Signature size: 7 ntree: 5000 mtry: 100 node-Size: 5 maxDepth: 15 nSplit: 1 splitRule: logrankscore

**Table A5.** Radiomic signatures for predicting loco-regional tumour control for the representative models based on the GTV<sub>entire</sub>, the GTV<sub>5mm-rim</sub> and the GTV<sub>5mm-core</sub> as well as the GTV<sub>5mm-rim+3mm-ext</sub> tumour sub-volumes. The mathematical description and the abbreviations of features can be found in Zwanenburg et al. [25].

Tumour Sub-Volume	Feature Name	Synonym	Image	Cluster
GTV <sub>entire</sub>	ivh_diff_v10_v90	F1 <sub>IVH</sub>	wav_coif1_llh	no
	morph_integ_int, stat_energy	$\overline{F2}_{M/S}$	base, wav_coif1_lll	yes
	ivh_v10	F3 <sub>IVH</sub>	wav_coif1_llh	no
	rlm_gl_var_3d_avg	F4 <sub>T</sub>	wav_coif1_lhl	no
	rlm_srhge_3d_avg	F5 <sub>T</sub>	base	no
GTV <sub>5mm-rim</sub>	ivh_diff_v10_v90	F1 <sub>IVH</sub>	wav_coif1_llh	no
	dzm_ldhge_3d	F2 <sub>T</sub>	base	no
	dzm_lgze_3d, dzm_sdlge_3d, szm_lgze_3d, szm_szlge_3d	$\overline{F3}_{T}$	base	yes
	ivh_v10	F4 <sub>IVH</sub>	wav_coif1_llh	no
	dzm_ldhge_3d	F5 <sub>T</sub>	wav_coif1_lll	no
	morph_integ_int, stat_energy	$\overline{F6}_{M/S}$	base, wav_coif1_lll	yes
	dzm_hgze_3d, szm_hgze_3d, szm_szhge_3d	F7 <sub>T</sub>	base	yes
	cm_auto_corr_d1_3d, cm_joint_avg_d1_3d, cm_sum_avg_d1_3d, ih_mean, ngl_hgce_d1_a0.0_3d, rlm_hgre_3d_avg	$\overline{F8}_{IH/T}$	base	yes
	cm_info_corr2_d1_3d_avg	F9 <sub>T</sub>	base	no
	ngt_strength_3d, rlm_rlnu_3d_avg	$\overline{F10}_{T}$	wav_coif1_llh	yes
GTV <sub>5mm-core</sub>	morph_integ_int, stat_energy	$\overline{F1}_{S}$	base, wav_coif1_lll	yes
	ngt_strength_3d, rlm_rlnu_3d_avg	$\overline{F2}_{T}$	wav_coif1_llh	yes
	dzm_gl_var_3d, szm_gl_var_3d	$\overline{F3}_{T}$	wav_coif1_hhl	yes
	dzm_glnu_3d, szm_glnu_3d	$\overline{F4}_{T}$	wav_coif1_llh	yes
	ivh_i50	F5 <sub>IVH</sub>	wav_coif1_lhl	no
GTV <sub>5mm-rim+3mm-ext</sub>	morph_moran_i	F1 <sub>M</sub>	base	no
	stat_energy	F2 <sub>S</sub>	wav_coif1_hhl	no
	ih_kurt,rlm_glnu_norm_3d_avg, stat_kurt	$\overline{F3}_{IH/T}$	wav_coif1_lhl	yes
	rlm_gl_var_3d_avg	F4 <sub>T</sub>	wav_coif1_lhh	no
	stat_max, stat_min, stat_range	$\overline{F5}_{S}$	wav_coif1_hhl	yes
	ngl_dc_var_d1_a0.0_3d	F6 <sub>T</sub>	wav_coif1_lhh	no
stat_cov	$\overline{F7}_{S}$	base, wav_coif1_lll	yes	

l: low pass, h: high pass, wav\_coif1: coiflet-1 wavelet high-/low-pass filter; M: morphological, S: first order statistic, T: texture-based feature; IH: intensity histogram, IVH: intensity volume histogram feature;  $\overline{F}$ : cluster of features represented by the mean value as a new meta-feature.

## References

- Baumann, M.; Krause, M.; Overgaard, J.; Debus, J.; Bentzen, S.M.; Daartz, J.; Richter, C.; Zips, D.; Bortfeld, T. Radiation oncology in the era of precision medicine. *Nat. Rev. Cancer* **2016**, *16*, 234. [[CrossRef](#)] [[PubMed](#)]
- Aerts, H.; Velazquez, E.; Leijenaar, R.; Parmar, C.; Grossmann, P.; Carvalho, S.; Bussink, J.; Monshouwer, R.; Haibe-Kains, B.; Rietveld, D.; et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat. Commun.* **2014**, *5*, 4006. [[CrossRef](#)]
- Vallièrès, M.; Kay-Rivest, E.; Perrin, L.; Liem, X.; Furstoss, C.; Aerts, H.; Khaouam, N.; Nguyen-Tan, P.; Wang, C.; Sultanem, K.; et al. Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer. *Sci. Rep.* **2017**, *7*, 10117. [[CrossRef](#)] [[PubMed](#)]
- Leger, S.; Zwanenburg, A.; Pilz, K.; Lohaus, F.; Linge, A.; Zöphel, K.; Kotzerke, J.; Schreiber, A.; Tinhofer, I.; Budach, C.; et al. A comparative study of machine learning methods for time-to-event survival data for radiomics risk modelling. *Sci. Rep.* **2017**, *7*, 13206. [[CrossRef](#)] [[PubMed](#)]

5. Leger, S.; Zwanenburg, A.; Pilz, K.; Zschaek, S.; Zöphel, K.; Kotzerke, J.; Schreiber, A.; Zips, D.; Krause, M.; Baumann, M.; et al. CT imaging during treatment improves radiomic models for patients with locally advanced head and neck cancer. *Radiother. Oncol.* **2019**, *130*, 10–17. [[CrossRef](#)] [[PubMed](#)]
6. Wu, J.; Gong, G.; Cui, Y.; Li, R. Intratumor partitioning and texture analysis of dynamic contrast-enhanced (DCE)-MRI identifies relevant tumor subregions to predict pathological response of breast cancer to neoadjuvant chemotherapy. *J. Magn. Reson. Imaging* **2016**, *44*, 1107–1115. [[CrossRef](#)]
7. Serganova, I.; Doubrovin, M.; Vider, J.; Ponomarev, V.; Soghomonyan, S.; Beresten, T.; Ageyeva, L.; Serganov, A.; Cai, S.; Balatoni, J.; et al. Molecular imaging of temporal dynamics and spatial heterogeneity of hypoxia-inducible factor-1 signal transduction activity in tumors in living mice. *Cancer Res.* **2004**, *64*, 6101–6108. [[CrossRef](#)]
8. Troost, E.G.; Laverman, P.; Philippens, M.E.; Lok, J.; van der Kogel, A.J.; Oyen, W.J.; Boerman, O.C.; Kaanders, J.H.; Bussink, J. Correlation of [18 F] FMISO autoradiography and pimonodazole immunohistochemistry in human head and neck carcinoma xenografts. *Eur. J. Nucl. Med. Mol. Imaging* **2008**, *35*, 1803–1811. [[CrossRef](#)]
9. Zips, D.; Zöphel, K.; Abolmaali, N.; Perrin, R.; Abramyuk, A.; Haase, R.; Appold, S.; Steinbach, J.; Kotzerke, J.; Baumann, M. Exploratory prospective trial of hypoxia-specific PET imaging during radiochemotherapy in patients with locally advanced head-and-neck cancer. *Radiother. Oncol.* **2012**, *105*, 21–28. [[CrossRef](#)]
10. Gatenby, R.; Grove, O.; Gillies, R. Quantitative imaging in cancer evolution and ecology. *Radiology* **2013**, *269*, 8–15. [[CrossRef](#)]
11. O'Connor, J.; Rose, C.; Waterton, J.; Carano, R.; Parker, G.; Jackson, A. Imaging intratumor heterogeneity: Role in therapy response, resistance, and clinical outcome. *Clin. Cancer Res.* **2015**, *21*, 249–257. [[CrossRef](#)] [[PubMed](#)]
12. Löck, S.; Perrin, R.; Seidlitz, A.; Bandurska-Luque, A.; Zschaek, S.; Zöphel, K.; Krause, M.; Steinbach, J.; Kotzerke, J.; Zips, D.; et al. Residual tumour hypoxia in head-and-neck cancer patients undergoing primary radiochemotherapy, final results of a prospective trial on repeat FMISO-PET imaging. *Radiother. Oncol.* **2017**, *124*, 533–540. [[CrossRef](#)]
13. Schütze, C.; Bergmann, R.; Yaromina, A.; Hessel, F.; Kotzerke, J.; Steinbach, J.; Baumann, M.; Beuthien-Baumann, B. Effect of increase of radiation dose on local control relates to pre-treatment FDG uptake in FaDu tumours in nude mice. *Radiother. Oncol.* **2007**, *83*, 311–315. [[CrossRef](#)]
14. Schütze, C.; Bergmann, R.; Brüchner, K.; Mosch, B.; Yaromina, A.; Zips, D.; Hessel, F.; Krause, M.; Thames, H.; Kotzerke, J.; et al. Effect of [18F] FMISO stratified dose-escalation on local control in FaDu hSCC in nude mice. *Radiother. Oncol.* **2014**, *111*, 81–87. [[CrossRef](#)]
15. Peltanova, B.; Raudenska, M.; Masarik, M. Effect of tumor microenvironment on pathogenesis of the head and neck squamous cell carcinoma: A systematic review. *Mol. Cancer* **2019**, *18*, 63. [[CrossRef](#)] [[PubMed](#)]
16. Alsahafi, E.; Begg, K.; Amelio, I.; Raulf, N.; Lucarelli, P.; Sauter, T.; Tavassoli, M. Clinical update on head and neck cancer: Molecular biology and ongoing challenges. *Cell Death Dis.* **2019**, *10*, 1–17. [[CrossRef](#)] [[PubMed](#)]
17. Alghary, A.; Shiradkar, R.; Pahwa, S.; Purysko, A.; Verma, S.; Moses, D.; Shnier, R.; Haynes, A.M.; Delprado, W.; Thompson, J.; et al. Combination of Peri-Tumoral and Intra-Tumoral Radiomic Features on Bi-Parametric MRI Accurately Stratifies Prostate Cancer Risk: A Multi-Site Study. *Cancers* **2020**, *12*, 2200. [[CrossRef](#)] [[PubMed](#)]
18. Grove, O.; Berglund, A.; Schabath, M.; Aerts, H.; Dekker, A.; Wang, H.; Rios Velazquez, E.; Lambin, P.; Gu, Y.; Balagurunathan, Y.; et al. Quantitative computed tomographic descriptors associate tumor shape complexity and intratumor heterogeneity with prognosis in lung adenocarcinoma. *PLoS ONE* **2015**, *10*, 1–14. [[CrossRef](#)] [[PubMed](#)]
19. Wu, J.; Gensheimer, M.; Dong, X.; Rubin, D.; Napel, S.; Diehn, M.; Loo, B.; Li, R. Robust Intratumor Partitioning to Identify High-Risk Subregions in Lung Cancer: A Pilot Study. *Int. J. Radiat. Oncol. Biol. Phys.* **2016**, *95*, 1504–1512. [[CrossRef](#)] [[PubMed](#)]
20. Leger, S. Radiomics Risk Modelling Using Machine Learning Algorithms for Personalised Radiation Oncology. Ph.D. Thesis, Faculty of Medicine and University Hospital Carl Gustav Carus, Technische Universität Dresden, 2018.
21. Leger, S.; Zwanenburg, A.; Pilz, K.; Lohaus, F.; Linge, A.; Zöphel, K.; Kotzerke, J.; Schreiber, A.; Tinhofer, I.; Budach, C.; et al. Identification of tumour sub-volumes for improved radiomic risk modelling in locally advanced HNSCC. *Radiother. Oncol.* **2018**, *127*, 263–264. [[CrossRef](#)]

22. Apolle, R.; Rehm, M.; Bortfeld, T.; Baumann, M.; Troost, E.G. The clinical target volume in lung, head-and-neck, and esophageal cancer: Lessons from pathological measurement and recurrence analysis. *Clin. Transl. Radiat. Oncol.* **2017**, *3*, 1–8. [[CrossRef](#)] [[PubMed](#)]
23. Linge, A.; Lohaus, F.; Löck, S.; Nowak, A.; Gudziol, V.; Valentini, C.; von Neubeck, C.; Jütz, M.; Tinhofer, I.; Budach, V.; et al. HPV status, cancer stem cell marker expression, hypoxia gene signatures and tumour volume identify good prognosis subgroups in patients with HNSCC after primary radiochemotherapy: A multicentre retrospective study of the German Cancer Consortium Radiation Oncology Group (DKTK-ROG). *Radiother. Oncol.* **2016**, *121*, 364–373. [[PubMed](#)]
24. Shafiq-UI-Hassan, G.G.; Latifi, K.; Ullah, G.; Hunt, D.; Balagurunathan, Y.; Abdalah, M.; Matthew, B.; Goldhof, D.; Mackin, D.; Court, L.; et al. Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. *Med. Phys.* **2017**, *44*, 1050–1062. [[CrossRef](#)] [[PubMed](#)]
25. Zwanenburg, A.; Vallières, M.; Abdalah, M.A.; Aerts, H.J.; Andrearczyk, V.; Apte, A.; Ashrafinia, S.; Bakas, S.; Beukinga, R.J.; Boellaard, R.; et al. The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* **2020**, *295*, 328–338. [[CrossRef](#)]
26. Harrell, F.E., Jr.; Lee, K.L.; Mark, D.B. Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat. Med.* **1996**, *15*, 361–387. [[CrossRef](#)]
27. Pencina, M.J.; D’Agostino, R.B. Overall C as a measure of discrimination in survival analysis: Model specific population value and confidence interval estimation. *Stat. Med.* **2004**, *23*, 2109–2123. [[CrossRef](#)]
28. El Naqa, I.; Grigsby, P.; Apte, A.; Kidd, E.; Donnelly, E.; Khullar, D.; Chaudhari, S.; Yang, D.; Schmitt, M.; Laforest, R.; et al. Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. *Pattern Recognit.* **2009**, *42*, 1162–1171. [[CrossRef](#)]
29. Dou, T.; Aerts, H.; Coroller, T.; Mak, R. Radiomic-Based Phenotyping of Tumor Core and Rim to Predict Survival in Non-small Cell Lung Cancer. *Int. J. Radiat. Oncol. Biol. Phys.* **2017**, *99*, S84. [[CrossRef](#)]
30. Dou, T.H.; Coroller, T.P.; van Griethuysen, J.J.; Mak, R.H.; Aerts, H.J. Peritumoral radiomics features predict distant metastasis in locally advanced NSCLC. *PLoS ONE* **2018**, *13*, e0206108. [[CrossRef](#)]
31. Hosny, A.; Parmar, C.; Coroller, T.P.; Grossmann, P.; Zeleznik, R.; Kumar, A.; Bussink, J.; Gillies, R.J.; Mak, R.H.; Aerts, H.J. Deep learning for lung cancer prognostication: A retrospective multi-cohort radiomics study. *PLoS Med.* **2018**, *15*, e1002711. [[CrossRef](#)]
32. Keek, S.; Sanduleanu, S.; Wesseling, F.; de Roest, R.; van den Brekel, M.; van der Heijden, M.; Vens, C.; Giuseppina, C.; Licitra, L.; Scheckenbach, K.; et al. Computed tomography-derived radiomic signature of head and neck squamous cell carcinoma (peri) tumoral tissue for the prediction of locoregional recurrence and distant metastasis after concurrent chemo-radiotherapy. *PLoS ONE* **2020**, *15*, e0232639.
33. Campbell, S.; Poon, I.; Markel, D.; Vena, D.; Higgins, K.; Enepekides, D.; Rapheal, S.; Wong, J.; Allo, G.; Morgen, E.; et al. Evaluation of microscopic disease in oral tongue cancer using whole-mount histopathologic techniques: Implications for the management of head-and-neck cancers. *Int. J. Radiat. Oncol. Biol. Phys.* **2012**, *82*, 574–581. [[CrossRef](#)] [[PubMed](#)]
34. Apolle, R.; Bijl, H.P.; Blanchard, P.; Laprie, A.; Madani, I.; Ruffier, A.; Van Elmpt, W.; Troost, E.G. Target volume delineation for adaptive treatment in HNSCC is highly variable among experts. *Radiother. Oncol.* **2019**, *133*, S655–S656. [[CrossRef](#)]
35. Zwanenburg, A.; Leger, S.; Agolli, L.; Pilz, K.; Troost, E.G.; Richter, C.; Löck, S. Assessing robustness of radiomic features by image perturbation. *Sci. Rep.* **2019**, *9*, 1–10. [[CrossRef](#)]
36. Haarbarger, C.; Müller-Franzes, G.; Weninger, L.; Kuhl, C.; Truhn, D.; Merhof, D. Radiomics feature reproducibility under inter-rater variability in segmentations of CT images. *Sci. Rep.* **2020**, *10*, 1–10. [[CrossRef](#)]
37. Pavic, M.; Bogowicz, M.; Würms, X.; Glatz, S.; Finazzi, T.; Riesterer, O.; Roesch, J.; Rudofsky, L.; Friess, M.; Veit-Haibach, P.; et al. Influence of inter-observer delineation variability on radiomics stability in different tumor sites. *Acta Oncol.* **2018**, *57*, 1070–1074. [[CrossRef](#)]
38. Vaupel, P.; Kallinowski, F.; Okunieff, P. Blood flow, oxygen and nutrient supply, and metabolic microenvironment of human tumors: A review. *Cancer Res.* **1989**, *49*, 6449–6465.
39. Ljungkvist, A.S.; Bussink, J.; Rijken, P.F.; Raleigh, J.A.; Denekamp, J.; Van Der Kogel, A.J. Changes in tumor hypoxia measured with a double hypoxic marker technique. *Int. J. Radiat. Oncol. Biol. Phys.* **2000**, *48*, 1529–1538. [[CrossRef](#)]

40. Kanungo, T.; Mount, D.; Netanyahu, N.; Piatko, C.; Silverman, R.; Wu, A. An efficient k-means clustering algorithm: Analysis and implementation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 881–892. [[CrossRef](#)]
41. Euser, A.M.; Zoccali, C.; Jager, K.J.; Dekker, F.W. Cohort studies: Prospective versus retrospective. *Nephron Clin. Pract.* **2009**, *113*, c214–c217. [[CrossRef](#)]
42. Clarke, L.P.; Nordstrom, R.J.; Zhang, H.; Tandon, P.; Zhang, Y.; Redmond, G.; Farahani, K.; Kelloff, G.; Henderson, L.; Shankar, L.; et al. The quantitative imaging network: NCI’s historical perspective and planned goals. *Transl. Oncol.* **2014**, *7*, 1. [[CrossRef](#)] [[PubMed](#)]
43. Buckler, A.J.; Bresolin, L.; Dunnick, N.R.; Sullivan, D.C.; Group. A collaborative enterprise for multi-stakeholder participation in the advancement of quantitative imaging. *Radiology* **2011**, *258*, 906–914. [[CrossRef](#)] [[PubMed](#)]
44. Buckler, A.J.; Bresolin, L.; Dunnick, N.R.; Sullivan, D.C.; Group. Quantitative imaging test approval and biomarker qualification: Interrelated but distinct activities. *Radiology* **2011**, *259*, 875–884. [[CrossRef](#)] [[PubMed](#)]

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



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).