

Improved outcome prediction of oropharyngeal cancer by combining clinical and MRI features in machine learning models

Citation for published version (APA):

Bos, P., van den Brekel, M. W. M., Gouw, Z. A. R., Al-Mamgani, A., Taghavi, M., Waktola, S., Aerts, H. J. W. L., Castelijns, J. A., Beets-Tan, R. G. H., & Jasperse, B. (2021). Improved outcome prediction of oropharyngeal cancer by combining clinical and MRI features in machine learning models. *European Journal of Radiology*, 139, Article 109701. <https://doi.org/10.1016/j.ejrad.2021.109701>

Document status and date:

Published: 01/06/2021

DOI:

[10.1016/j.ejrad.2021.109701](https://doi.org/10.1016/j.ejrad.2021.109701)

Document Version:

Publisher's PDF, also known as Version of record

Document license:

Taverne

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.



Improved outcome prediction of oropharyngeal cancer by combining clinical and MRI features in machine learning models

Paula Bos^{a,b,c,*}, Michiel W.M. van den Brekel^{b,d}, Zeno A.R. Gouw^e, Abraham Al-Mamgani^e, Marjaneh Taghavi^a, Selam Waktola^a, Hugo J.W.L. Aerts^{a,c,f}, Jonas A. Castelijns^a, Regina G. H. Beets-Tan^{a,c,g}, Bas Jasperse^{a,h}

^a Department of Radiology, The Netherlands Cancer Institute, Amsterdam, the Netherlands

^b Department of Head and Neck Oncology and Surgery, The Netherlands Cancer Institute, Amsterdam, the Netherlands

^c GROW School for Oncology and Developmental Biology, University of Maastricht, Maastricht, the Netherlands

^d Department of Oral and Maxillofacial Surgery, Amsterdam University Medical Center (AUMC), Amsterdam, the Netherlands

^e Department of Radiation Oncology, The Netherlands Cancer Institute, Amsterdam, the Netherlands

^f Artificial Intelligence in Medicine (AIM) Program, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States

^g Department of Regional Health Research, University of Southern Denmark, Denmark

^h Department of Radiology, Amsterdam University Medical Center, Amsterdam, the Netherlands

ARTICLE INFO

Keywords:

Machine learning
Head and neck neoplasms
Magnetic Resonance Imaging
Treatment outcome
Oropharyngeal neoplasms
Radiomics

ABSTRACT

Objectives: New markers are required to predict chemoradiation response in oropharyngeal squamous cell carcinoma (OPSCC) patients. This study evaluated the ability of magnetic resonance (MR) radiomics to predict locoregional control (LRC) and overall survival (OS) after chemoradiation and aimed to determine whether this has added value to traditional clinical outcome predictors.

Methods: 177 OPSCC patients were eligible for this study. Radiomic features were extracted from the primary tumor region in T1-weighted postcontrast MRI acquired before chemoradiation. Logistic regression models were created using either clinical variables (clinical model), radiomic features (radiomic model) or clinical and radiomic features combined (combined model) to predict LRC and OS 2-years posttreatment. Model performance was evaluated using area under the curve (AUC), 95 % confidence intervals were calculated using 500 iterations of bootstrap. All analyses were performed for the total population and the Human papillomavirus (HPV) negative tumor subgroup.

Results: A combined model predicted treatment outcome with a higher AUC (LRC: 0.745 [0.734–0.757], OS: 0.744 [0.735–0.753]) than the clinical model (LRC: 0.607 [0.594–0.620], OS: 0.708 [0.697–0.719]). Performance of the radiomic model was comparable to the combined model for LRC (AUC: 0.740 [0.729–0.750]), but not for OS prediction (AUC: 0.654 [0.646–0.662]). In HPV negative patients, the performance of all models was not sufficient with AUCs ranging from 0.587 to 0.660 for LRC and 0.559 to 0.600 for OS prediction.

Conclusion: Predictive models that include clinical variables and radiomic tumor features derived from MR images of OPSCC better predict LRC after chemoradiation than models based on only clinical variables. Predictive models that include clinical variables perform better than models based on only radiomic features for the prediction of OS.

1. Introduction

Oropharyngeal squamous cell carcinoma (OPSCC) is a frequent tumor of the upper aero-digestive tract, with an increasing incidence in

the last decades [1]. Although definitive chemo- and radiation therapy (chemoradiation (CRT)) is currently considered the standard of care for patients with locally advanced OPSCC, surgery, especially minimal invasive transoral robotic surgery (TORS), followed by CRT can be a

Abbreviations: CRT, chemoradiation therapy; HPV, Human Papilloma Virus; LRC, locoregional control within 2 year after treatment; OPSCC, oropharyngeal squamous cell carcinoma; OS, overall survival within 2 year after treatment.

* Corresponding author at: The Netherlands Cancer Institute, Department Radiology, N0.010, Plesmanlaan 121, 1066 CX Amsterdam, the Netherlands.

E-mail address: p.bos@nki.nl (P. Bos).

<https://doi.org/10.1016/j.ejrad.2021.109701>

Received 19 November 2020; Received in revised form 14 March 2021; Accepted 1 April 2021

Available online 8 April 2021

0720-048X/© 2021 Elsevier B.V. All rights reserved.

good alternative and might enable de-intensification of the post-operative CRT, depending on the disease stage and patients' or clinicians' preference [2,3]. Although CRT has a high rate of treatment response, a considerable number of OPSCC patients have recurrent or residual disease after CRT leading to significant morbidity, mortality, and deterioration of quality of life. HPV tumor status is the most important predictor of treatment success, generally showing better treatment outcomes for HPV positive and less favorable treatment outcomes for HPV negative tumors [4]. Additional markers to predict CRT response are needed especially for HPV negative patients, allowing these patients to undergo an alternative treatment strategy (e.g. neoadjuvant chemotherapy combined with TORS or induction immunotherapy) at an early stage of the treatment trajectory.

Over the past years, image analysis techniques have been developed to extract and quantify visually occult tumor properties from computer tomography (CT) and MR images, collectively called radiomic features. These radiomic features have been associated with gene expression, histological tissue properties, survival, and treatment outcome. Previous studies on this topic have found prognostic radiomic features from CT images. For instance, intratumor heterogeneity quantified on CT images proved to be predictive of survival [5]. Compared to CT, MRI may provide other insights in tissue properties due to fundamental differences in image acquisition [6]. Few studies have investigated prognostic radiomic features from MRI images of head and neck cancer. These studies mainly focused on outcome prediction in nasopharyngeal carcinoma using radiomics or deep learning [6–10]. MRI is the preferred modality for OPSCC patients in most centers, providing an unique chance to study the ability of MRI radiomics to predict treatment outcome.

This study aimed to predict CRT treatment outcome for OPSCC using radiomic features derived from pretreatment MR images, and to determine whether these MR-based radiomic features have added value to clinical predictors of treatment outcome.

2. Materials and methods

The institutional ethics review board approved the study. Informed consent was waived for this retrospective analysis of anonymous data.

2.1. Patients

A total of 240 consecutive OPSCC patients, treated with CRT between January 2010 and December 2015 at our institute, were considered for this study. Inclusion criteria were: 1) histologically proven primary OPSCC treated with CRT, 2) minimum of 2 years of follow-up after treatment and, 3) availability of relevant clinical parameters. Exclusion criteria were unavailable pretreatment MRI examination of the primary tumor ($n = 38$), poor image quality ($n = 7$), and small undetectable ($n = 17$) or double tumors ($n = 1$). A total of 177 patients were eligible for this study.

Age, gender, smoking status (non-smoker vs. smoker), date of tumor recurrence, occurrence of lymph node metastasis, and survival within 2-years after treatment were collected for all patients. TNM-stage (7th edition), subsite and HPV status based on immunohistochemistry p16 and DNA HPV polymerase chain reaction were collected for each tumor. Clinical variables age and TNM stage were dichotomized to create groups of patients younger or older than 60 years, low and high T-stage ($T1 + T2$, $T3 + T4$) and positive or negative nodal disease.

2.2. Treatment

Patients were treated by chemoradiotherapy using Image-guided Intensity-modulated radiation therapy (IMRT) or volumetric modulated arc therapy (VMAT). Prescribed dose was 70 Gy to the primary tumor and the involved nodes and 46 Gy electively to the low-risk regions. The radiation was given in a daily fraction of 2 Gy, 5 times a week

for 7 weeks. Set-up verification and correction of the patients was done using daily cone-beam CT. Patients received three cycles of cisplatin-based chemotherapy (100 gr/m^2), administered on day 1, 22 and 43 of their radiation treatment.

2.3. Outcome variables

The primary outcome was locoregional control (LRC), defined as the absence of a histopathological proven local recurrence and/or lymph node metastases within 2 years after initial complete response. Secondary outcome was overall survival (OS), defined as the proportion of patients surviving 2 years after treatment.

2.4. Imaging data

Pretreatment MRI was routinely performed as part of primary staging for patients with OPSCC. All MRI examinations were acquired at 1.5 Tesla ($n = 82$ patients) or 3.0 Tesla ($n = 95$ patients) on a Philips Medical System, see supplementary Table 1 for detailed acquisition information. The full imaging protocol included T1w, T2w, postcontrast 3D T1w and dynamic scans.

2.5. Tumor delineation

Primary tumors were manually delineated by one observer in training (PB, 1 year of head and neck experience, non-expert delineations), and, subsequently controlled and corrected by an experienced head and neck radiologist (BJ, >7 years of head and neck experience) on the postcontrast 3D T1w MRI using the freely available segmentation software 3D Slicer (version 4.8.0, www.slicer.org) (see Fig. 1). Average spatial agreement was good with a mean dice similarity coefficient (DSC) of 0.83. Dice similarity coefficient was between 0.9–1.0 in 53 % of the cases, between 0.8–0.9 in 22 %, between 0.7–0.8 in 9 %, and, below 0.7 in 16 % of the cases. Larger tumor volumes showed significantly better overlap compared to small tumor volumes ($p = 0.001$, independent t-test). Tumor volumes were delineated on every axial slice on the postcontrast 3D T1w MRI. Both observers were blinded to outcome data but were allowed to interpret other available pretreatment imaging data to optimize their delineations.

2.6. Feature extraction

Imaging features were extracted from tumor volumes using the open-source python package, Pyradiomics (version 2.2.0) [11]. All MRI examinations were normalized (centering at zero mean and one standard deviation) to obtain a homogeneous histogram of MR signal and resampled by B-spline interpolation to a pixel spacing of $1.0 \times 1.0 \times 1.0 \text{ mm}^3$. Gray values were discretized using a fixed bin width of five. Features were extracted from the image data three times: original image, with a wavelet image filter, including eight decompositions, and finally with a Laplacian of Gaussian (LoG) filter (four levels (0.5–2.0 mm)).

Features with zero variance (i.e. constant features), and therefore of no discriminatory value, were removed. Features were considered stable, if they had no significant difference between non-expert and expert tumor delineations (intraclass correlation coefficient >0.75) and between magnetic field strengths (Mann-Whitney U test $p \geq 0.05$). The remaining stable features were then tested for collinearity. Features that correlated with other features with a Pearson coefficient higher than 0.9 were removed. In this removal process, the feature that showed high correlation with the greatest number of other features was removed. This was repeated until only the diagonal elements of the correlation matrix exceeded the threshold of 0.9.

2.7. Machine learning analysis

Analysis was performed in all eligible patients ($n = 177$) and in a

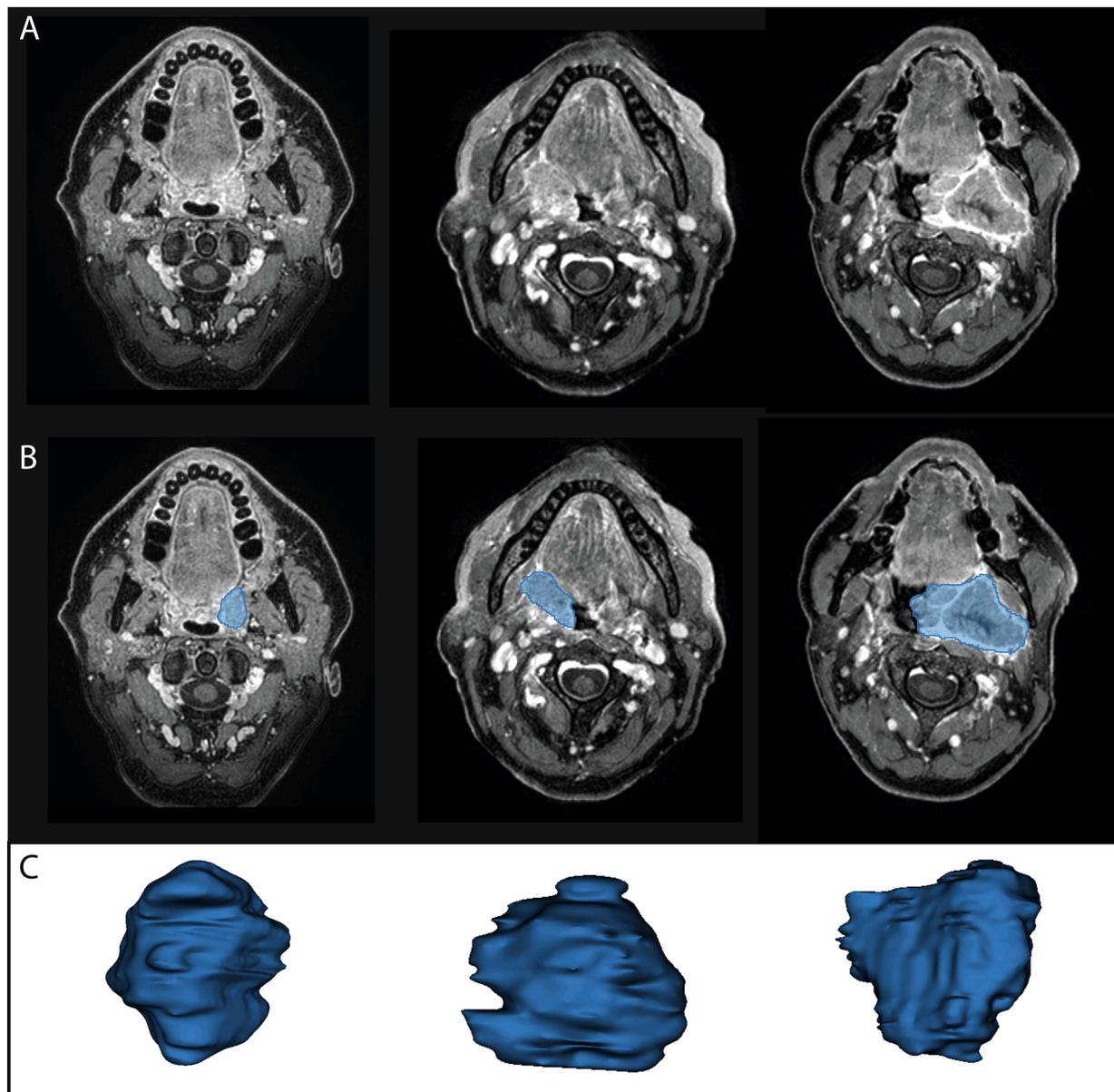


Fig. 1. Examples of tumor delineations on postcontrast 3D T1w MRIs. From top to bottom the MRI without manual delineation (A), MRI with manual delineation (B) and the reconstructed 3D tumor volumes (C) for three patients (left, middle, right).

subset of patients with HPV negative tumors ($n = 77$). Sub-analysis of patients with HPV positive tumors ($n = 76$) was considered of limited added value, as the majority had favorable outcomes for both LRC and OS (LRC: 68/76, OS: 67/76). In 24 patients, HPV status was unavailable. Patients were randomly split into a training (70 %) and test-set (30 %), see [Table 1](#), stratifying for treatment outcome and MRI field strength. HPV status was included as stratification factor for the total patient cohort.

Table 1
Detailed information of patient numbers in training and test set for the development of a prediction model.

	Total patient cohort	HPV negative subset
Total number of patients	177	77
Training set (70 %)	124	53
Cross validation: training (75 %)	93	40
Cross validation: validation (25 %)	31	13
Test set (30 %)	53	24

Three models were created for each of the outcome variables (LRC and OS) using only clinical variables (clinical model), only radiomic features (radiomics model) and a combination of clinical variables and radiomic features (combined model). Features were prepared for logistic regression analysis using the following steps: 1) Standardization of features to zero mean and unit variance, and 2) Reduction of the number of features by wrapper feature selection using a sequential backward feature selection method, which removed irrelevant features by iteratively removing the feature with the weakest feature importance score [12].

In the training phase, optimal model settings of the machine learning pipeline were found utilizing 1000 iterations of Bayesian hyperparameter optimization (Python library Hyperopt version 0.2 [13]), applying fourfold cross validation within the training set (see [Table 1](#) for patient numbers). The regularization parameter and the number of selected features in wrapper feature selection were tuned during Bayesian hyperparameter optimization (supplementary Table 2). Training performance of the predictive models was evaluated using median AUC and its 95 % confidence interval (95 %CI) from the

performance of the optimal hyperparameters in fourfold cross validation.

In the testing phase, the optimal hyperparameter combination obtained in the training step was applied to the unseen test dataset. Model test performance was evaluated by the median and 95 %CI of AUC, sensitivity, specificity and accuracy obtained using 500 iterations of bootstrap (with replacement).

2.8. Statistical analysis

Univariate Fishers' exact test was used to test differences in clinical features between groups with regard to outcome parameters (OS and LRC). P-values <0.05 were considered statistically significant (p = 0.004 after Bonferroni correction). Statistical differences between the predictive radiomic features of the models were tested using the Wald test (p-values < 0.05 were considered statistically significant). All analyses were implemented in python 3.5 and SPSS version 25.0 (SPSS Inc.). The radiomic workflow is visualized in Fig. 2.

3. Results

Detailed patient characteristics and oncologic outcomes are summarized in Table 2. Of the total patient group, approximately half had a high T-stage (T3-T4) and 80 % had node-positive disease. Considering only patients with known HPV status revealed an equal distribution between positive (n = 76) and negative (n = 77) HPV status. Patients with favorable outcomes for LRC and OS were more likely to have HPV positive tumors (LRC: p = 0.004, OS: p = 0.001).

3.1. Predictive performance of models for all patients

Out of 1184 radiomic features, 75 features were stable. Performance of the clinical, radiomics and combined model based on logistic regression for prediction of LRC and OS are summarized in Table 3.

3.1.1. Locoregional control

The predictive properties of the clinical model (Test AUC: 0.607, Sens: 0.57, Spec: 0.60, Acc: 0.57) are less favorable compared to the radiomic model (Test AUC: 0.740, Sens: 0.75, Spec: 0.60, Acc: 0.71) with regard to LRC. The combined model (Test AUC: 0.745, Sens: 0.73, Spec: 0.71, Acc: 0.71) shows a similar performance as the radiomic model.

Lower T-stage (r: 0.330), HPV positivity (r: 0.305), tumor not located at the posterior oropharyngeal wall (r: -0.174) and lower age (r: -0.166) were predictive determinants of LRC in the clinical model (supplementary Table 3). Four and five radiomic features were selected in the radiomic and combined model, respectively. Rounder and more homogeneous tumors were associated with disease control (supplementary Table 4). No clinical variables were selected in the combined model.

Table 2

Patient demographics. Baseline characteristics and outcome after CRT for all patients, and HPV negative tumors. Summaries are given as number of patients and % of the total group between parentheses. Median and interquartile range (IQR) are used to summarize continuous variables. Fisher exact test after Bonferroni correction *p = 0.004 and p = 0.001 for LRC and OS respectively. Clinical values were only significant for total patient cohort.

Patients, n	Total patient cohort (n = 177)	HPV negative tumors (n = 77)	HPV positive tumors (n = 76)
Age (>60years)	101 (57)	52 (67)	36 (47)
Sex, n male (%)	111 (63)	54 (70)	42 (55)
Smoking, n (%)	134 (76)	72 (94)	42 (55)
HPV			
Negative, n (%)	77 (44)	77 (100)	-
Positive, n (%)	76 (43)	-	76 (100)
Unknown, n (%)	24 (13)	-	-
T-stage, n (%)			
T1 + T2	94 (53)	25 (33)	53 (70)
T3 + T4	83 (47)	52 (67)	23 (30)
N-stage, n (%)			
N0	36 (20)	18 (23)	8 (11)
N1	26 (15)	11 (15)	12 (16)
N2	110 (62)	47 (61)	52 (68)
N3	5 (3)	1 (1)	4 (5)
Subsite of cancer			
Tonsillar tissue	99 (56)	42 (55)	46 (60)
Soft palate	18 (10)	11 (14)	2 (3)
Base of tongue	56 (32)	20 (26)	28 (37)
Posterior wall	4 (2)	4 (5)	0 (0)
Clinical endpoints			
LRC <2 year, n (%)	144 (81)	55 (71)	66 (87)
Time to LRF in months, median (IQR)	6 (4-17)	6 (4-13)	9 (4-18)
OS after 2 years, n (%)	137 (77)	50 (65)	72 (95)
OS in months for non-survivors, median (IQR)	12 (8-17)	14 (9-18)	12 (10-15)

HPV: Human papillomavirus; LRC: Locoregional control; OS: Overall Survival; LRF: Locoregional failure.

3.1.2. Overall survival

For the prediction of OS, the predictive performance of the clinical model (Test AUC: 0.708, Sens: 0.68, Spec: 0.67, Acc: 0.69) is better than the radiomic model (Test AUC: 0.654, Sens: 0.62, Spec: 0.57, Acc: 0.60). The combined model (Test AUC: 0.744, Sens: 0.71, Spec: 0.78, Acc: 0.71) had the highest performance and outperformed the two other models.

Eight, ten and twenty-two features were prognostic for overall survival, regarding respectively the clinical, radiomic and combined model (supplementary Table 3). In the clinical model, lower T-stage (r: 0.409), younger patients (r: -0.395), HPV positivity (r: 0.348), node-negative disease (r: 0.232), tumors not located in the posterior oropharyngeal

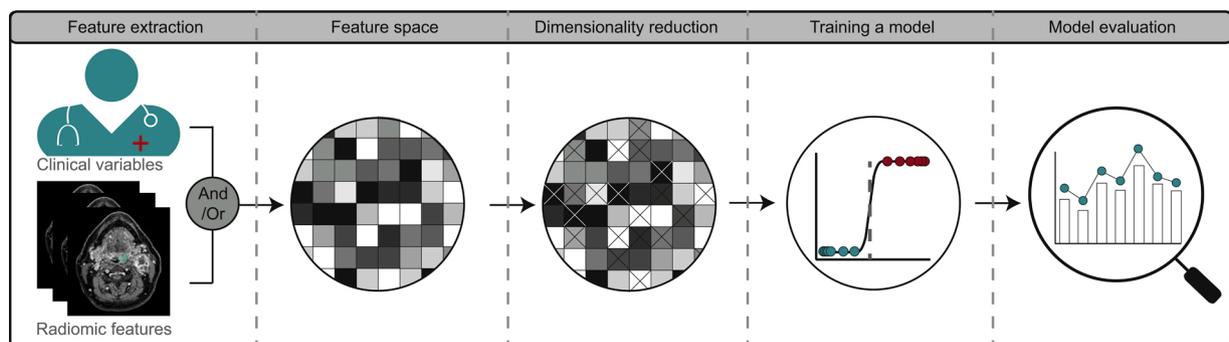


Fig. 2. Flowchart of the radiomics workflow. First clinical variables and/or radiomic features were extracted from the patient and MR image respectively. Feature space including only clinical variables, only radiomic features or the combination were created to build a clinical, radiomic or combined prediction model respectively. After dimensionality reduction, using wrapper feature selection, a logistic regression prediction model is trained and model performance is evaluated.

Table 3

Performance expressed as AUC [95 % CI] for the models predicting LRC and OS within 2 years after chemoradiation for all patients. Confidence intervals were calculated from 500 iterations of bootstrapping.

Model	Training AUC [CV]	Test AUC [CI bootstrap]	Sensitivity [CI bootstrap]	Specificity [CI bootstrap]	Accuracy [CI bootstrap]
LRC					
Clinical	0.637 [0.572–0.702]	0.607 [0.594–0.620]	0.57 [0.56–0.58]	0.60 [0.58–0.62]	0.57 [0.56–0.58]
Radiomic	0.783 [0.690–0.875]	0.740 [0.729–0.750]	0.75 [0.74–0.76]	0.60 [0.58–0.62]	0.71 [0.71–0.72]
Combined	0.747 [0.640–0.855]	0.745 [0.734–0.757]	0.73 [0.72–0.73]	0.71 [0.70–0.73]	0.71 [0.71–0.72]
OS					
Clinical	0.659 [0.558–0.760]	0.708 [0.697–0.719]	0.68 [0.67–0.69]	0.67 [0.65–0.68]	0.69 [0.68–0.69]
Radiomic	0.601 [0.501–0.702]	0.654 [0.646–0.662]	0.62 [0.61–0.62]	0.57 [0.56–0.59]	0.60 [0.59–0.61]
Combined	0.548 [0.519–0.577]	0.744 [0.735–0.753]	0.71 [0.70–0.72]	0.78 [0.76–0.79]	0.71 [0.71–0.72]

CV: Cross validation.

wall (r: -0.147), tumors located at the base of tongue (r: 0.095) and female gender (r: -0.041) were associated with OS. Radiomic features show less complex, coarse and more homogeneous tumors in patients who are more likely to survive (supplementary Table 4).

3.2. Predictive performance of HPV negative tumors

After feature reduction, 123 features remained for the HPV negative subgroup. Table 4 summarizes predictive properties of prediction models in HPV negative tumors.

Performance of all models was generally low for LRC (Test AUCs 0.587 to 0.660) and OS (Test AUCs 0.559 to 0.600). Performance of the clinical model was lower than the model based on radiomic features for both LRC (Test AUC: 0.587 and 0.652 respectively) and OS (Test AUC: 0.559 and 0.593 respectively). Performance of the radiomic model was comparable to the combined model for both LRC (Test AUC: 0.660, Sens: 0.83, Spec: 0.43, Acc: 0.71) and OS (Test AUC: 0.600, Sens: 0.40, Spec: 0.67, Acc: 0.51).

4. Discussion

The main finding of this study was that predictive models based on a combination of clinical variables and MR-based radiomic features have a reasonable ability to predict LRC and OS within 2 years after CRT in OPSCC. Sub analysis of HPV negative patients showed moderate performance in the prediction of LRC and poor performance in the prediction of OS.

Interestingly, predictive performance of models based on only clinical variables was not as good as the combined models. This implies that clinical variables and radiomics features hold independent information for outcome prediction. Radiomic features are likely to add information embedded in tumor structure for the prediction of treatment outcome not captured by clinical variables. Clinical variables may add to the radiomic features in different ways for LRC and OS. For LRC, information is added to the risk of recurrence by clinical factors that influence tumor biology, such as HPV status and age. For OS, non-tumor related information is added to risk of death, like age and comorbidities. These findings, indicates that clinical and imaging features should preferably

Table 4

Performance expressed as AUC [95 % CI] for the models predicting LRC and OS within 2 years after chemoradiation for all patients with HPV negative tumors. Confidence intervals were calculated from 500 iterations of bootstrapping.

Model	Training AUC [CV]	Test AUC [CI bootstrap]	Sensitivity [CI bootstrap]	Specificity [CI bootstrap]	Accuracy [CI bootstrap]
LRC					
Clinical	0.510 [0.442–0.579]	0.587 [0.578–0.595]	0.71 [0.70–0.72]	0.27 [0.26–0.29]	0.57 [0.56–0.58]
Radiomic	0.706 [0.510–0.901]	0.652 [0.642–0.661]	0.83 [0.83–0.84]	0.43 [0.41–0.44]	0.71 [0.71–0.72]
Combined	0.706 [0.510–0.901]	0.660 [0.650–0.670]	0.83 [0.83–0.84]	0.43 [0.41–0.44]	0.71 [0.71–0.72]
OS					
Clinical	0.606 [0.390–0.821]	0.559 [0.543–0.563]	0.47 [0.46–0.48]	0.67 [0.65–0.68]	0.54 [0.54–0.55]
Radiomic	0.501 [0.409–0.593]	0.593 [0.583–0.602]	0.60 [0.59–0.61]	0.32 [0.31–0.34]	0.51 [0.51–0.52]
Combined	0.478 [0.360–0.596]	0.600 [0.591–0.609]	0.40 [0.39–0.41]	0.67 [0.65–0.68]	0.51 [0.51–0.52]

CV: Cross validation.

be combined when constructing models to predict treatment outcome. This is in line with findings of Mes et al. [14] for oral cancer patients and HPV negative OPSCC.

For the prediction of LRC, the combined model consisted of only radiomic features while both clinical variables and radiomic features were included in the construction of the model. The performance of this combined model was slightly better than the radiomic model due to the addition of the radiomics variable skewness. Additional analysis (not shown) revealed that the correlation of clinical variables with selected radiomic features was low. Apparently, the combination of clinical variables and radiomics variables in the model construction sequence makes slight improvements in the eventual combined model compared to the radiomics model, in this case with the addition of skewness. This slight improvement occurred even though clinical variables do not obviously correlate with the radiomics variables. This is an important consideration to take into account in construction of predictive radiomics models.

For prediction of OS, the combined model consisted of a relatively large number of radiomic features and clinical variables. As mentioned previously, risk of death includes a broad range of factors that are not directly tumor related. The large number of clinical and radiomic features with generally low regression coefficients in the combined model for prediction of OS reflects this.

The radiomic features revealed that rounder and more homogenous tumors are associated with a more favorable outcome. This relationship is probably a reflection of genetic tumor diversity/dedifferentiation resulting in more heterogeneous and irregular tumors with worse treatment response and higher rate of locoregional failure. These findings are in line with another MRI-based radiomics study of head and neck cancer showing higher homogeneity and rounder shapes for overall survival [7].

HPV is an important determinant of the biology and behavior of OPSCC, and is known to be a strong predictor of treatment outcome in OPSCC, prompting us to create separate models for HPV positive and negative tumors. As expected, most patients with HPV positive tumors had a favorable outcome, which did not permit us to create a meaningful model for this tumor type. Distribution of outcome variables for HPV negative tumors permitted the construction of a predictive model, but

did not reach consistent meaningful predictive properties. This was probably due to the low number of patients ($n = 77$) in this subgroup. The role of clinical and radiomics models in outcome prediction for HPV positive and, particularly, negative tumors therefore remains unclear.

This study has a relatively large sample size ($n = 177$) compared to other published studies on MR radiomics in head and neck cancer (maximum 118 patients) [10]. However, these results are not generalizable to other hospitals with different scanners and scanner protocols. The next step is to replicate these findings in an external validation cohort from multiple centers [15–17].

MRI based radiomics is still difficult to implement in the clinical workup due to a lack of standardization in acquisition parameters and harmonization between MRI machines, as was shown in previous studies [14,16,17]. Until standardization of acquisition is available, standardization between centers can be reached by harmonizing pre-processing steps and correlation analysis to obtain stable features between centers. Even though stable feature reduces bias introduced by human interaction by manual delineation, some human influence cannot be ruled out completely. In the future, automated delineation techniques may be able to eliminate this unwanted bias

This study extracted radiomic features from primary tumors based on postcontrast T1w MRI. Extracting features from other MR sequences might give a better representation of tumor biology, and may harbor information relevant to treatment outcome. For instance, the dynamic contrast-enhanced MRI parameters have shown its prognostic ability to predict OS and progression-free survival [18]. Sample size considerations and preliminary results prompted us to use only T1w 3D sequences to ensure meaningful results. Evidently this needs to be considered in future studies.

This study shows that predictive models that include radiomic tumor features derived from MR images of OPSCC better predict LRC after chemoradiation than models based on only clinical variables. Predictive models that include clinical variables perform better than models based on only radiomic features for the prediction of OS.

Future studies on MRI based radiomics should confirm these findings in a larger patient cohort and elucidate the potential role of radiomics in outcome prediction in HPV positive and, especially, HPV negative tumors.

CRedit authorship contribution statement

Paula Bos: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Michiel W.M. van den Brekel:** Conceptualization, Methodology, Formal analysis, Investigation, Resources, Writing - review & editing, Supervision, Project administration, Funding acquisition. **Zeno A.R. Gouw:** Resources, Data curation, Writing - review & editing. **Abraham Al-Mamgani:** Resources, Data curation, Writing - review & editing. **Marjaneh Taghavi:** Software, Validation, Formal analysis, Writing - review & editing. **Selam Waktola:** Software, Validation, Formal analysis, Writing - review & editing, Supervision. **Hugo J.W.L. Aerts:** Software, Validation, Writing - review & editing, Supervision. **Jonas A. Castelijns:** Formal analysis, Writing - review & editing, Supervision. **Regina G.H. Beets-Tan:** Conceptualization, Methodology, Formal analysis, Investigation, Resources, Writing - review & editing, Supervision, Project administration, Funding acquisition. **Bas Jasperse:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

H.J.W.L. Aerts: Stockholder Sphera & Genospace, outside submitted

work.

A. Al-Mamgani, R.G.H. Beets-Tan, M.W.M. van den Brekel, P. Bos, J. Castelijns, B. Jasperse, Z.A.R. Gouw, Chintan Parmer, Marjaneh Taghavi, and S. Waktola: None.

Acknowledgements

We would like to acknowledge financial support from the Verwelius Foundation.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.ejrad.2021.109701>.

References

- [1] H.A. Van Monsjou, M. Schaapveld, O. Hamming-Vrieze, J.P. De Boer, M.W.M. Van Den Brekel, A.J.M. Balm, Cause-specific excess mortality in patients treated for cancer of the oral cavity and oropharynx: a population-based study, *Oral Oncol.* 52 (2016) 37–44, <https://doi.org/10.1016/j.oraloncology.2015.10.013>.
- [2] J.A. Bonner, P.M. Harari, J. Giralt, et al., Radiotherapy plus cetuximab for squamous-cell carcinoma of the head and neck, *N. Engl. J. Med.* 354 (2006) 567–578.
- [3] Y.M. Park, H.R. Kim, B.C. Cho, K.C. Keum, N.H. Cho, S.-H. Kim, Transoral robotic surgery-based therapy in patients with stage III-IV oropharyngeal squamous cell carcinoma, *Oral Oncol.* 75 (2017) 16–21, <https://doi.org/10.1016/j.oraloncology.2017.10.014>.
- [4] C. Fakhry, Q. Zhang, P.F. Nguyen-Tan, et al., Human papillomavirus and overall survival after progression of oropharyngeal squamous cell carcinoma, *J. Clin. Oncol.* 32 (2014) 3365–3373, <https://doi.org/10.1200/JCO.2014.55.1937>.
- [5] H.J.W.L. Aerts, E.R. Velazquez, R.T.H. Leijenaar, et al., Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach, *Nat. Commun.* 5 (2014) 4006, <https://doi.org/10.1038/ncomms5006>.
- [6] Z. Liu, S. Wang, D. Dong, et al., The applications of radiomics in precision diagnosis and treatment of oncology: opportunities and challenges, *Theranostics* 9 (2019) 1303–1322, <https://doi.org/10.7150/thno.30309>.
- [7] T.T. Zhai, L.V. van Dijk, B.T. Huang, et al., Improving the prediction of overall survival for head and neck cancer patients using image biomarkers in combination with clinical parameters, *Radiother. Oncol.* 124 (2017) 256–262, <https://doi.org/10.1016/j.radonc.2017.07.013>.
- [8] M. Qiang, C. Li, Y. Sun, et al., A prognostic predictive system based on deep learning for locoregionally advanced nasopharyngeal carcinoma, *JNCI J. Natl. Cancer Inst.* (2020) djaa149, <https://doi.org/10.1093/jnci/djaa149>.
- [9] H. Farhidzadeh, J.Y. Kim, J.G. Scott, D.B. Goldgof, L.O. Hall, L.B. Harrison, Classification of progression free survival with nasopharyngeal carcinoma tumors, *SPIE Med. Imaging* 9785 (2016) 1I, <https://doi.org/10.1117/12.2216976>.
- [10] A. Jethanandani, T.A. Lin, S. Volpe, et al., Exploring applications of radiomics in magnetic resonance imaging of head and neck cancer: a systematic review, *Front. Oncol.* 8 (2018) 131, <https://doi.org/10.3389/fonc.2018.00131>.
- [11] J.J.M. van Griethuysen, A. Fedorov, C. Parmar, et al., Computational radiomics system to decode the radiographic phenotype, *Cancer Res.* 77 (2017) e104–e107, <https://doi.org/10.1158/0008-5472.CAN-17-0339>.
- [12] R. Kohavi, G.H. John, Wrappers for feature subset selection, *Artif. Intell.* 97 (1997) 273–324.
- [13] J. Bergstra, D. Yamins, D. Cox, Hyperopt: a python library for optimizing the hyperparameters of machine learning algorithms, *Proc. SciPy* (2013) 13–20, <https://doi.org/10.1088/1749-4699/8/1/014008>.
- [14] S.W. Mes, F.H.P. van Velden, B. Peltenburg, et al., Outcome prediction of head and neck squamous cell carcinoma by MRI radiomic signatures, *Eur. Radiol.* (2020) 6311–6321, <https://doi.org/10.1007/s00330-020-06962-y>.
- [15] J.E. Park, S.Y. Park, H.J. Kim, H.S. Kim, Reproducibility and generalizability in radiomics modeling: possible strategies in radiologic and statistical perspectives, *Korean J. Radiol.* 20 (2019) 1124–1137, <https://doi.org/10.3348/kjr.2018.0070>.
- [16] A. Traverso, L. Wee, A. Dekker, R. Gillies, Repeatability and reproducibility of radiomic features: a systematic review, *Int. J. Radiat. Oncol. Biol. Phys.* 102 (2018) 1143–1158, <https://doi.org/10.1016/j.ijrobp.2018.05.053>.
- [17] A. Carré, G. Klausner, M. Edjlali, et al., Standardization of brain MR images across machines and protocols: bridging the gap for MRI-based radiomics, *Sci. Rep.* 10 (2020), 12340, <https://doi.org/10.1038/s41598-020-69298-z>.
- [18] S.H. Ng, C.T. Liao, C.Y. Lin, Dynamic contrast-enhanced MRI, diffusion-weighted MRI and ¹⁸F-FDG PET/CT for the prediction of survival in oropharyngeal or hypopharyngeal squamous cell carcinoma treated with chemoradiation, *Eur. Radiol.* 26 (2016) 4162–4172, <https://doi.org/10.1007/s00330-016-4276-8>.