



Prediction of post-radiotherapy locoregional progression in HPV-associated oropharyngeal squamous cell carcinoma using machine-learning analysis of baseline PET/CT radiomics

Stefan P. Haider^{a,b,1}, Kariem Sharaf^{b,1}, Tal Zeevi^c, Philipp Baumeister^b, Christoph Reichel^b, Reza Forghani^d, Benjamin H. Kann^e, Alexandra Petukhova^a, Benjamin L. Judson^f, Manju L. Prasad^g, Chi Liu^h, Barbara Burtneisⁱ, Amit Mahajan^a, Seyedmehdi Payabvash^{a,*}

^a Section of Neuroradiology, Department of Radiology and Biomedical Imaging, Yale School of Medicine, 789 Howard Ave, PO Box 208042, New Haven, CT 06519, United States

^b Department of Otorhinolaryngology, University Hospital of Ludwig Maximilians Universität München, Marchioninistrasse 15, 81377 Munich, Germany

^c Center for Translational Imaging Analysis and Machine Learning, Department of Radiology and Biomedical Imaging, Yale School of Medicine, 333 Cedar Street, New Haven, CT 06510, United States

^d Department of Diagnostic Radiology and Augmented Intelligence & Precision Health Laboratory (AIPHL), McGill University Health Centre & Research Institute, 1650 Cedar Avenue, Montreal, Quebec QC H3G 1A4, Canada

^e Department of Radiation Oncology, Dana-Farber Cancer Institute, Harvard Medical School, 450 Brookline Avenue, Boston, MA 02215, United States

^f Division of Otolaryngology, Department of Surgery, Yale School of Medicine, 330 Cedar Street, New Haven, CT 06520, United States

^g Department of Pathology, Yale School of Medicine, 310 Cedar Street, New Haven, CT 06520, United States

^h Division of Bioimaging Sciences, Department of Radiology and Biomedical Imaging, Yale School of Medicine, 333 Cedar Street, New Haven, CT 06510, United States

ⁱ Section of Medical Oncology, Department of Internal Medicine, Yale School of Medicine and Yale Cancer Center, 25 York Street, New Haven, CT 06520, United States

ARTICLE INFO

Keywords:

Radiomics
Oropharyngeal squamous cell carcinoma
PET/CT
HPV
Imaging biomarker
Risk stratification

ABSTRACT

Locoregional failure remains a therapeutic challenge in oropharyngeal squamous cell carcinoma (OPSCC). We aimed to devise novel objective imaging biomarkers for prediction of locoregional progression in HPV-associated OPSCC. Following manual lesion delineation, 1037 PET and 1037 CT radiomic features were extracted from each primary tumor and metastatic cervical lymph node on baseline PET/CT scans. Applying random forest machine-learning algorithms, we generated radiomic models for censoring-aware locoregional progression prognostication (evaluated by Harrell's C-index) and risk stratification (evaluated in Kaplan-Meier analysis). A total of 190 patients were included; an optimized model yielded a median (interquartile range) C-index of 0.76 (0.66–0.81; $p = 0.01$) in prognostication of locoregional progression, using combined PET/CT radiomic features from primary tumors. Radiomics-based risk stratification reliably identified patients at risk for locoregional progression within 2-, 3-, 4-, and 5-year follow-up intervals, with log-rank p -values of $p = 0.003$, $p = 0.001$, $p = 0.02$, $p = 0.006$ in Kaplan-Meier analysis, respectively. Our results suggest PET/CT radiomic biomarkers can predict post-radiotherapy locoregional progression in HPV-associated OPSCC. Pending validation in large, independent cohorts, such objective biomarkers may improve patient selection for treatment de-intensification trials in this prognostically favorable OPSCC entity, and eventually facilitate personalized therapy.

Introduction

Head and neck squamous cell carcinoma (HNSCC) is among the most morbid cancers [1]. Sustained high-risk human papillomavirus (HPV)-

infection in the oropharynx is the cause of a large and increasing proportion of oropharyngeal squamous cell carcinomas (OPSCC) [2], characterized by distinct demographic, biologic and – most notably – prognostic attributes compared to HPV-negative OPSCC [3]. Consequently, the

Abbreviations: AJCC, American Joint Committee on Cancer; AUC, area under the receiver operating characteristic curve; CT, computed tomography; GTV, gross tumor volume; Gy, Gray; HNSCC, head and neck squamous cell carcinoma; HPV, human papillomavirus; IQR, interquartile range; LRP, locoregional progression; OPSCC, oropharyngeal squamous cell carcinoma; PET, [18F]fluorodeoxyglucose positron emission tomography; RCF, random classification forest; RSF, random survival forest; TCIA, The Cancer Imaging Archive; VOI, volume of interest.

* Corresponding author.

E-mail address: sam.payabvash@yale.edu (S. Payabvash).

¹ Both the authors contributed equally to this work.

<https://doi.org/10.1016/j.tranon.2020.100906>

Received 27 July 2020; Received in revised form 27 September 2020; Accepted 28 September 2020

1936-5233/© 2020 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

American Joint Committee on Cancer (AJCC) adopted separate staging schemes for HPV-associated and HPV-negative OPSCC in the 8th edition staging manual [4,5].

Improved responsiveness to treatment and more indolent natural history of HPV-associated OPSCC may potentially render this prognostically favorable subtype amenable to treatment de-intensification with reduced treatment-related toxicity [6,7]. Nevertheless, treatment failure with locoregional disease progression (LRP) is a negative prognostic factor in HPV-associated OPSCC, often entailing salvage resection or irradiation which are commonly associated with increased morbidity and impaired functionality, and ultimately resulting in reduced overall survival [8–10]. Thus, there is a pressing need for novel biomarkers to identify patients amenable for safe treatment de-escalation and ultimately personalized clinical decision-making.

The notion that quantitative characterization of increasingly larger sets of biomedical data may pave the way for precision diagnosis, prognostication and treatment decision-making has shaped the “-omics” concept – e.g. genomics, metabolomics, proteomics. Radiomics analysis has expanded the scope of “-omics” to quantitative characterization of medical images by extracting high-dimensional sets of “features” from volumes of interest (VOI) such as primary tumor lesions, which capture lesion shape, image intensity and texture patterns. The resulting imaging biomarkers may be correlated with treatment outcome, tumor microenvironment, tissue heterogeneity and pathophysiology; and may enable development of prognostic tools substituting or supplementing traditional outcome predictors such as cancer staging [11–15]. Depending on the imaging modality used, radiomic features can represent a variety of tumor characteristics; [¹⁸F]fluorodeoxyglucose positron emission tomography (PET) radiomics may provide wholistic quantification of tumor metabolic activity and activity distribution; whereas computed tomography (CT) radiomics can describe structural properties and tissue density. In many centers, PET/CT imaging is an integral part of cancer staging and work-up.

Prior studies have demonstrated the predictive value of radiomic biomarkers for LRP in HNSCC, but HPV status was rarely available in all studied OPSCC patients, and subgroup analysis of HPV-associated OPSCC was not reported [16–19]. Radiomics analysis can predict HPV status, and thus the results of prior studies may in part reflect the differences between HPV-associated and HPV-negative subgroups [20,21]. In this study, we aim to apply machine-learning algorithms using combined PET and non-contrast CT radiomic features extracted from baseline clinical scans for prediction and risk stratification of post-radiotherapy LRP in an HPV-associated OPSCC cohort. We acquired a multi-institutional cohort, and devised prognostic biomarkers using radiomic features from the primary tumor as well as metastatic cervical lymph nodes in addition to clinical variables.

Material and methods

Imaging and clinical data

Imaging data and corresponding clinical information were retrospectively acquired from (1) Yale’s Smilow Hospital cancer registry from 2009 to 2019; and (2) public collections in The Cancer Imaging Archive (TCIA) [22]: (2a) the “Head-Neck-PET-CT” collection provides data from four institutions in Canada (“Canadian” cohort) [23]; and (2b) the “Data from Head and Neck Cancer CT Atlas” collection holds an MD Anderson Cancer Center dataset (“MD Anderson” cohort) [24]. Our institutional review board approved this study under IRB protocol #2,000,024,295 and waived informed consent; TCIA provides de-identified data with consents obtained and ethical compliance ensured by source institutions.

We included cases of histopathologically confirmed OPSCC with (1) confirmed HPV-association, (2) pre-treatment PET and non-contrast CT scans of the neck, (3) LRP events or ≥ 18 months of adequate follow-up documentation, and (4) patients who received radiotherapy

as part of definitive or adjuvant treatment after surgery, with or without concurrent platinum-based chemotherapy or targeted therapy with cetuximab.

We excluded (1) HPV-negative subjects, (2) patients receiving palliation only and/or denying treatment, (3) patients with recurrent OPSCC, (4) with M1 disease at initial staging, (5) with $>50\%$ of the primary gross tumor volume affected by artifacts on visual evaluation of CT scans [25], and (6) with <60 Gray (Gy) in the adjuvant, and <66 Gy in the definitive radiotherapy setting delivered to the gross tumor volume [26].

Post-treatment cancer surveillance at our institution included regular physical examinations, endoscopy and imaging, with additional tissue sampling performed at specialists’ discretion. Locoregional disease progression was ascertained by tissue sampling or unequivocal imaging evidence; the latter was confirmed in retrospective data review by documented response to therapy or additional histopathological examination. Study endpoints in TCIA cohorts were based on annotations provided in the datasets.

Lesion segmentation and staging

The segmentation, radiomic feature extraction and disease progression modelling pipeline employed in our study is illustrated in Fig. 1. Separate PET and CT VOI corresponding to the primary tumor lesion and each individual metastatic cervical lymph node were generated as a first step in our radiomics pipeline. Regional metastatic spread was determined based on tissue sampling or unequivocal PET scan findings. We utilized 3D-Slicer version 4.10.1 [27] for image review and segmentation.

The co-registered pre-treatment PET/CT scans were retrieved and reviewed in 3D-Slicer, and the gross tumor volume (GTV) as defined by the “ICRU 83” report [28] was assessed. Using the “Paint” and “Erase” tools in the 3D-Slicer “Segment Editor” module, hypermetabolic areas of the primary tumor and every metastatic lymph node were manually delineated (i.e. slice-by-slice segmentation on axial PET reconstructions). PET segmentations were then copied onto the co-registered CT and manually adjusted to the GTV outline on CT using the “Paint” and “Erase” tools to generate the CT VOI, excluding air, adjacent uninvolved bone, and preserved fat planes. Axial CT slices with streak artifacts involving the lesion upon visual assessment were excluded from analysis; and metastatic lymph nodes with $>50\%$ of the GTV involved were entirely excluded [25].

A trained research associate (SPH) initially segmented all lesions; followed by VOI verification and adaption by a neuroradiologist (SP) with greater than 8 years of experience in head and neck cancer imaging. SP and AM (neuroradiologist with greater than 12 years of experience) performed OPSCC staging in accordance with the AJCC 8th edition staging manual [5].

Pre-processing, feature extraction and stability-based feature pre-selection

PET/CT imaging and image reconstruction were performed at the scan source institutions utilizing standard clinical protocols. PET and CT pre-processing was applied before radiomic feature extraction to homogenize the imaging data: PET grey scale normalization, PET and CT voxel size interpolation to isotropic dimensions, CT re-segmentation, generation of ten image derivatives per original PET or CT enhancing certain image characteristics, and grey scale discretization were consecutively performed – a detailed description of our automated pre-processing pipeline is included in the supplementary methods [20].

A set of 1037 PET and 1037 CT radiomic features was subsequently extracted from each primary tumor and metastatic node VOI; comprised of volumetric shape features ($n=14$ features) extracted from only the original image ($n=1$ image); and first-order ($n=18$ features) and texture-matrix features ($n=75$ features) extracted from

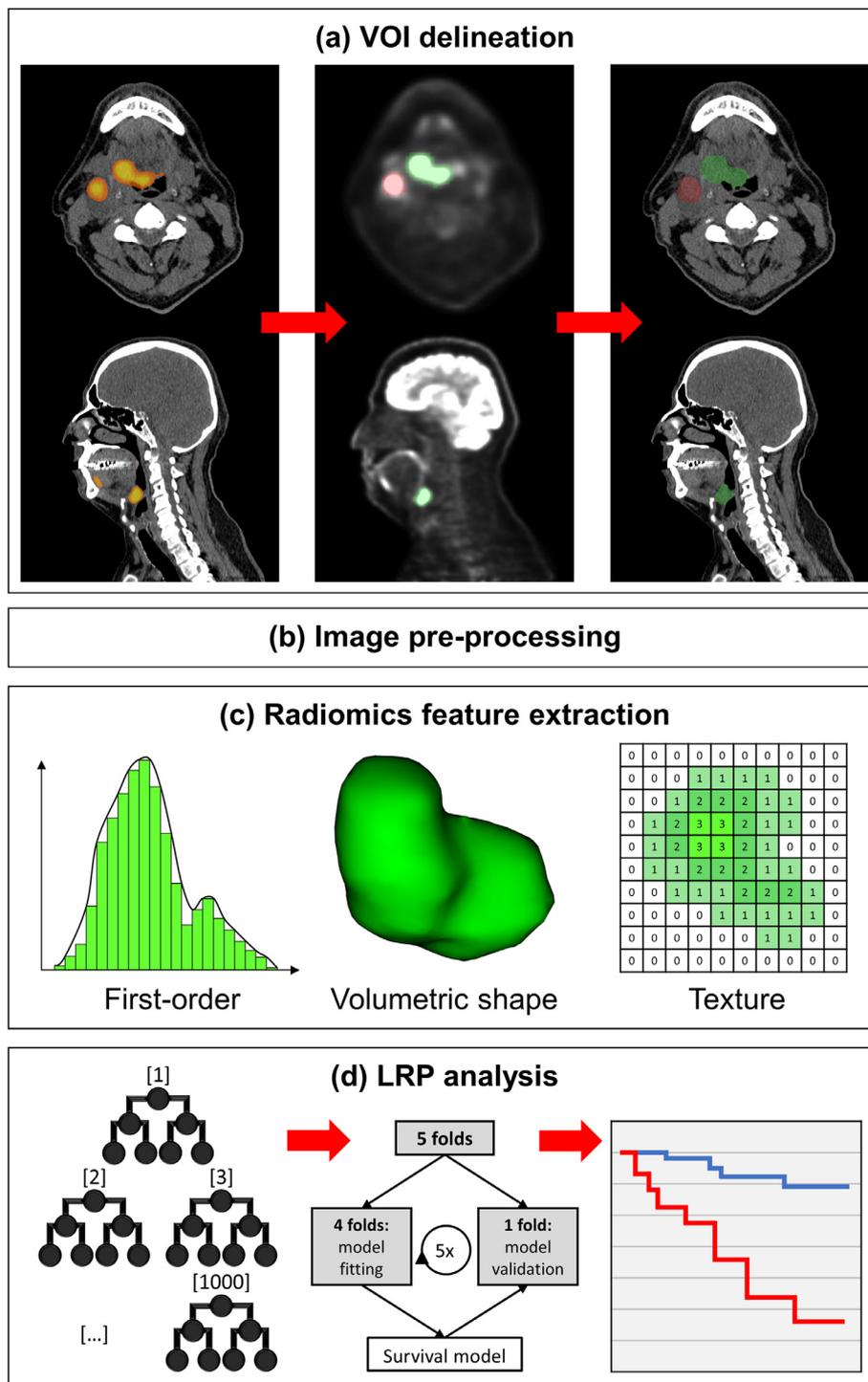


Fig. 1. Radiomics pipeline: (a) VOI delineation – after reviewing the co-registered scans, all lesions were manually delineated on PET axial images, and segmentations were transferred and adapted to the corresponding CT; (b) image pre-processing – details are included in the supplementary methods; (c) radiomics features extraction – 1037 PET and 1037 CT features corresponding to three categories (first-order, volumetric shape, texture) were extracted from each lesion, a comprehensive feature list is included in the supplement; (d) LRP analysis – prognostication and risk stratification was based on random forest machine-learning models with 1000 decision trees internally validated in 20-repeat 5-fold cross-validation, wherein models were iteratively trained on 4 folds, and evaluated in the 5th fold.

both the originals ($n=1$ image) and image derivatives ($n=10$ images) generated in pre-processing. This approach yielded a total of $(14 \times 1 + 18 \times 11 + 75 \times 11 =)$ 1037 PET and 1037 CT features per lesion. A complete list of radiomic features utilized in this study is included in supplementary Table 1. A Pyradiomics version 2.1.2 pipeline was customized and applied for radiomics analysis [20, 29].

We investigated radiomic feature stability in an inter-rater and intra-rater setting to pre-select features prior to disease progression modelling, given the volatile robustness of individual features to delineation variability reported in previous studies [30–32]. Unstable features were excluded; the methodology and results are reported in the supplementary methods and supplementary Table 2, respectively [20].

Disease progression modelling and prognostication

We defined locoregional progression (LRP) as the event of interest, with time-to-LRP defined as the time interval from OPSCC initial diagnosis to progression. Right-censoring was applied for loss to follow-up, death, or diagnosis of distant metastases. Subsequently, patients without an LRP event and <18 months of follow-up from diagnosis to censoring were excluded.

We devised and compared three types of LRP models [15]: (1) “Radiomics” models used radiomic signatures, (2) the “clinical” model incorporated AJCC staging (T-, N- and overall-stage), patient age at initial diagnosis, and the treatment modality, and (3) “combined” mod-

els utilized the combined set of radiomics and above-mentioned clinical predictors for LRP prognostication. AJCC T-, N- and overall-stage were included as ordinal variables with four (T1-T4), four (N0-N3) and three (overall stages I-III) levels, respectively. No overall-stage IV cancers were present, since subjects with distant metastases were excluded. Radiomic features, feature clusters and patient age were numeric variables, and the treatment modality was included as a categorical variable.

An array of different approaches was implemented and compared to generate and select features for prognostically optimized radiomic signatures in combined and radiomic models: (1) Radiomic features from three imaging modalities (PET, CT, PET and CT) were used for LRP modelling; (2) we derived radiomic features from two VOI sources: the primary tumor and a “virtual” VOI combining the primary tumor and all metastatic nodes in a given subject as described by Yu et al. [21]; and (3) the prognostic ability of unreduced feature sets was compared to three dimensionality reduction techniques (abbreviations in Fig. 2, details in the supplementary methods). We applied and compared all methodological approaches (3 imaging modalities x 2 VOI sources x 4 dimensionality reduction techniques) for LRP modelling.

The generic model types (“radiomics”, “clinical”, “combined”) introduced above were implemented applying random survival forest (RSF) [33] machine-learning algorithms for prognostication, which were configured to grow 1000 decision trees using a C-index split rule [34] with the remaining parameters in default. Statistical analysis was performed in R version 3.6.0 [35] using extension packages, R base functions and custom-written code. We used the “ranger” package (version 0.12.1) [33] for RSF modelling.

To limit overfitting and enhance generalizability, all models were internally validated in a framework applying 20 repeats of stratified 5-fold cross-validation (i.e. 100 permutations) using the event/non-event groups and follow-up duration as strata. Consensus VOI generation (if applicable), radiomic feature standardization, dimensionality reduction (if applicable), and RSF fitting were consecutively performed on the training folds, and RSF performance was quantified in the validation fold of each cross-validation iteration. This approach avoids “information leakage” from training to validation data and generates realistic estimates of RSF performance in new datasets.

We quantified models’ prognostic abilities in each validation fold with a right-censoring adjusted concordance index (Harrell’s C-index [34]), and the median score was calculated across 20 cross-validation iterations to represent models’ overall performance.

We further investigated the performance of three select models: the clinical model, and the best – in terms of C-index score – radiomic and combined model, respectively. Models’ validation fold C-index distribution across 20-repeat 5-fold cross validation was compared against random predictions (i.e. C-index calculated with the same model predictions but randomly resampled validation fold LRP outcome) using a corrected paired *t*-test (“corrected repeated k-fold cv test” [36]). *P*-values <0.05 ascertained significance.

We generated time-dependent performance curves to track and compare model performance throughout follow-up by calculating Uno’s estimator of cumulative/dynamic area under the curve (AUC) for right-censored survival data [37] in each validation fold (“survAUC” package [38] for R), and averaging AUC scores across 20 x 5-fold cross validation. The resulting performance curves were plotted for the first five years of follow-up.

Risk stratification and Kaplan-Meier analysis

The potential role of radiomics for LRP risk stratification was investigated by generating radiomics risk groups (high-risk vs. low-risk) in binary classification analysis [15]. We subsequently conducted Kaplan-Meier analysis with radiomics risk groups. For comparison, AJCC-staging (T-, N- overall-stage), patient age (age ≥ cohort median vs. < cohort median), and treatment modality variables served as Kaplan-Meier

risk groups. A log-rank test generated *p*-values with *p*<0.05 considered significant.

To generate radiomics risk groups, our framework applying 20-repeat stratified 5-fold cross-validation was adapted for binary classification, using event/non-event groups as strata, and a random classification forest (RCF) algorithm (“ranger” package version 0.12.1) [33] configured to grow 1000 decision trees for risk score computation (i.e. probability of experiencing an event). RCF case weights in a given outcome class (event or non-event) were specified to be inversely proportional to the class distribution in the training data to account for imbalance, with the remaining RCF parameters in default.

Patients’ RCF risk scores were averaged across validation folds, and a risk cutoff was calculated by maximizing Youden’s statistic in receiver operating characteristic-analysis. Patients with averaged risk scores greater than the cutoff were allocated to the radiomics high-risk group. RCF models were trained on the radiomics-only dataset of the radiomic LRP model selected for further evaluation (previous subsection) without feature selection applied.

Patients were labelled for Kaplan-Meier analysis using 2-, 3-, 4- and 5-year follow-up cutoffs; subjects diagnosed with LRP before a given cutoff were labelled positive, subjects lost to follow-up before a cutoff were excluded, and the remainder was labelled negative and censored at the cutoff. Separate RCF models were generated for each cutoff, and the resulting radiomics risk groups and were investigated in separate Kaplan-Meier plots. Equivalently, Kaplan-Meier analysis with clinical variables was conducted separately for each follow-up cutoff. This strategy avoids censoring before a cutoff (i.e. “dense” survival data) and thus enables RCF performance maximization, while allowing both exploring the majority of the documented follow-up period as well as comparing radiomics risk stratification with clinical variables in an easily interpretable fashion.

Results

Cohort characteristics

A total of 190 patients with HPV-associated OPSCC met inclusion criteria; thereof, 15 (~8%) had LRP events at a median (interquartile range, IQR) of 14.5 (11.0–21.6) months after initial diagnosis. Patients were followed-up for a median (IQR) of 40.7 (30.7–53.5) months after initial diagnosis. Table 1 summarizes demographics, treatment, imaging and staging characteristics of our study cohort.

In addition to 190 OPSCC primary tumors, 266 metastatic lymph nodes were segmented. Thereof, 422 (19.2%) out of 2193 primary tumor lesion axial slices, and 155 (5.6%) out of 2778 lymph node lesion axial slices were affected by streak artifact on CT, and were excluded (details in supplementary Table 3).

Prognostication of locoregional disease progression

The best radiomics LRP model in our study yielded a median (IQR) C-index of 0.76 (0.66–0.81; *p* = 0.01) using the full set of PET/CT primary tumor radiomic features (Fig. 2). The model using clinical variables did not exhibit prognostic value in cross validation, yielding a C-index (IQR) of 0.49 (0.39–0.58; *p* = 0.46), and combined models achieved median scores similar to those of corresponding radiomic models (Fig. 2). Combined PET/CT radiomic models achieved higher prognostic performance than single imaging modality models in the majority of permutations. Models combining radiomic features from primary tumors and metastatic cervical lymph nodes (“virtual” consensus VOI) improved PET-based LRP prognostication, with the best combined model yielding a median C-index (IQR) of 0.65 (0.52–0.76) using random forest-based feature selection (“pRF”); whereas the corresponding PET primary tumor model yielded a median C-index of 0.64 (0.50–0.71).

Select models subjected to further evaluation are highlighted in Fig. 2. Performance curve plotting (Fig. 3) again revealed similar

Table 1
Cohort characteristics.

Number of OPSCC patients – n	190
Included metastatic lymph nodes – n	266
LRP events – n (%)	15 (7.9%)
Follow-up [months] – median (IQR)	40.7 (30.7–53.5)
Time-to-event [months] – median (IQR)	14.5 (11.0–21.6)
Data source – n (%)	
Yale	112 (58.9%)
TCIA	78 (41.1%)
Sex – n (%)	
male	154 (81.1%)
female	36 (18.9%)
Age [years] – mean (SD)	59.83 (8.51)
HPV status – n (%)	
positive	190 (100%)
Smoking – n (%)	
never-smoker	48 (25.3%)
smoker	77 (40.5%)
pack-years – median (IQR)	15 (7.75–30)
pack-years unknown – n	15
unknown	65 (34.2%)
T stage ^a – n (%)	
T1	26 (13.7%)
T2	77 (40.5%)
T3	64 (33.7%)
T4	23 (12.1%)
N stage ^a – n (%)	
N0	35 (18.4%)
N1	108 (56.8%)
N2	43 (22.6%)
N3	4 (2.1%)
Overall stage ^a – n (%)	
I	85 (44.7%)
II	78 (41.1%)
III	27 (14.2%)
Included lymph nodes / patient – range	0 – 6
Primary treatment – n (%)	
CCRT or CBRT	135 (71.1%)
Surgery with adjuvant RT, CCRT or CBRT	34 (17.9%)
RT alone	21 (11.1%)
PET ^b – mean (SD)	
slice thickness [mm]	3.44 (0.40)
in-plane pixel spacing [mm]	4.28 (0.90)
in-plane image matrix [n x n]	148.25 (60.17) x idem
CT ^b – mean (SD)	
slice thickness [mm]	3.06 (0.60)
in-plane pixel spacing [mm]	1.12 (0.18)
in-plane image matrix [n x n]	512 x 512

^a AJCC 8th edition staging manual T/N/overall stage [5].

^b Values from image originals before preprocessing.

CBRT = concurrent bioradiotherapy with cetuximab; CCRT = concurrent platinum-based chemoradiotherapy; RT = radiotherapy alone; SD = standard deviation.

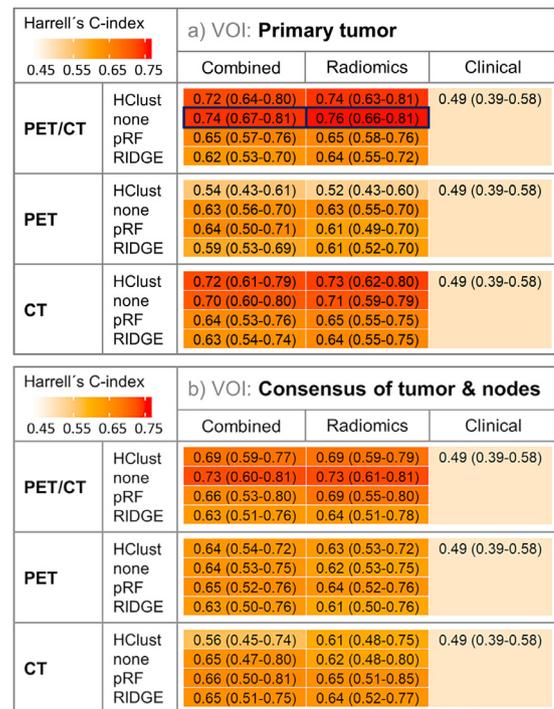


Fig. 2. Heatmap summary of LRP model performance quantified by the median (IQR) validation fold Harrell's C-index across 20-repeat 5-fold cross-validation. The radiomics and combined models selected for further evaluation are highlighted (blue frame). All methodological combinations to generate radiomics signatures for radiomics and combined models were applied (3 imaging modalities x 2 VOI sources x 4 dimensionality reduction techniques (HClust, none, pRF, RIDGE)). Clinical = clinical model; Combined = combined model; HClust = hierarchical clustering; none = no dimensionality reduction applied; pRF = Pearson correlation-based redundancy reduction with random survival forest variable importance; Radiomics = radiomics model; RIDGE = Cox regression with RIDGE regularization adapted for feature selection.

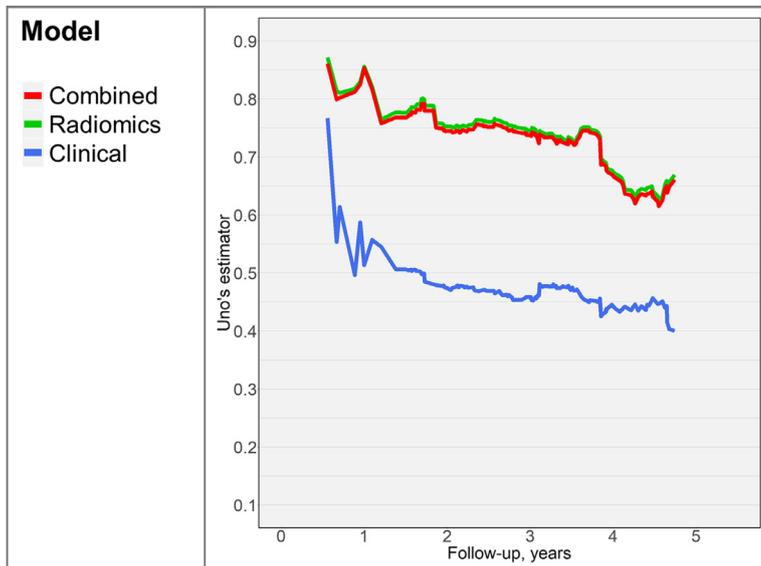


Fig. 3. Time-dependent performance curves depict selected models' (highlighted in Fig. 2) prognostic performance throughout 5-years of follow-up. The corresponding clinical model is presented for comparison.

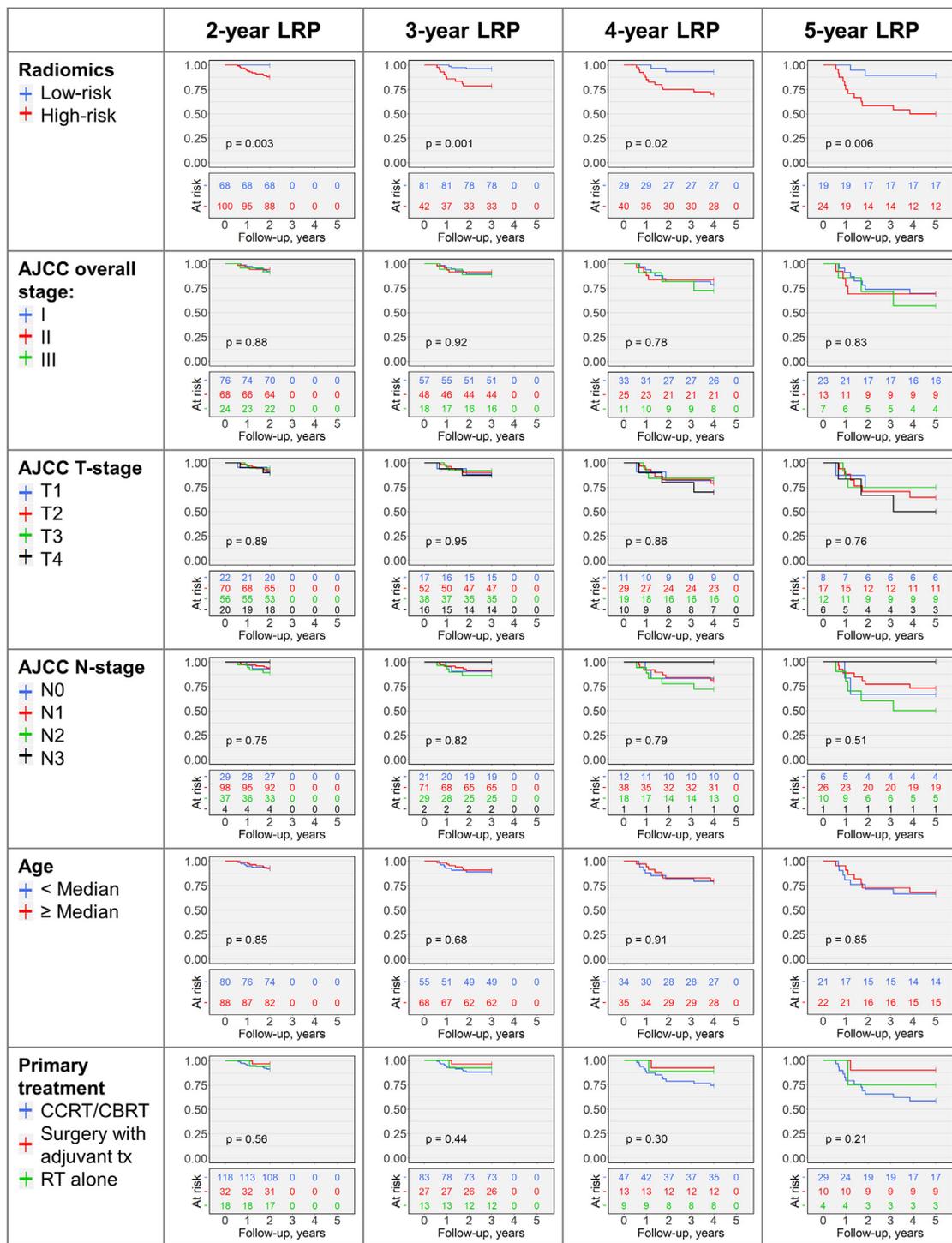


Fig. 4. Kaplan-Meier plots and log-rank test p-values depicting risk stratification based on radiomics analysis and clinical variables.

superiority of the radiomic model and the combined model over the clinical model. Notably, while radiomic modelling exhibited high prognostic performance in follow-up years 1 through 4, model performance was moderate in the fifth year.

Risk stratification of locoregional disease progression

In Kaplan-Meier analysis, radiomics high-risk groups exhibited significantly higher rates of LRP than corresponding low-risk groups in analysis of all follow-up cutoffs, achieving log-rank p-values of $p = 0.003$, $p = 0.001$, $p = 0.02$, $p = 0.006$ for the 2-, 3-, 4-, and 5-year

follow-up intervals, respectively (Fig. 4). Risk groups derived from clinical variables (AJCC staging, age, treatment) did not differ significantly ($p > 0.05$, Fig. 4).

Discussion

Improved responsiveness to treatment and more indolent natural history of HPV-associated OPSCC – as compared to HPV-negative cancers – may render this prognostically favorable subtype amenable to de-intensified therapy with reduced treatment-related toxicity and morbidity [6,7]. Accurate prognostication and risk stratification are, how-

ever, the first steps in personalized treatment decision-making. Using a multi-institutional cohort, we investigated the prognostic value of baseline PET/CT radiomics in prediction of LRP. Applying machine-learning analysis, we devised prognostic models utilizing PET/CT features from primary tumor lesions (with and without metastatic cervical nodes). Pending validation in larger cohorts, these novel objective biomarkers can provide decision assistance tools for precision treatment planning in patients with HPV-associated OPSCC.

LRP represents treatment failure of definite therapy in curative intent, with few remaining satisfactory options: Salvage surgery and irradiation are commonly associated with increased morbidity and impaired functionality; and while new immune checkpoint inhibitors alone or in addition to conventional systemic treatment improved outcome in patients not amenable to localized therapy, long-term control of relapsed HNSCC often remains fairly poor [39–41]. Locoregional relapse is also strongly tied to poor overall survival in HPV-associated OPSCC [8, 10], substantiating the importance of this endpoint for therapeutic decision making in OPSCC.

While pre-treatment PET/CT imaging is a mainstay of disease work-up and cancer staging, human visual interpretation cannot seize the full prognostic utility encoded in metabolic and structural bioimaging patterns [11–14]. By capturing such bioimaging features, radiomic biomarkers may help identify patients who are at increased risk for LRP, and may potentially improve patient selection in future trials of treatment de-intensification for HPV-associated OPSCC, and guide personalized clinical treatment planning.

In this study, we showed the merits of radiomic features quantifying tumor intensity, volumetric shape and texture for LRP prognostication and risk stratification. Employing a pre-processing pipeline designed to mitigate heterogeneity in imaging data, a multiple-delineation feature pre-selection approach retaining only stable radiomic features, and a rigorous cross-validation scheme avoiding data leakage from training to test sets, our results depict the prognostic potentials of machine-learning-generated radiomic biomarkers for LRP in a realistic fashion. A model utilizing the full set of combined PET/CT primary tumor radiomic features was reliably prognostic of LRP in cross validation, yielding a median (IQR) C-index of 0.76 (0.66–0.81; $p = 0.01$); whereas a clinical model combining AJCC overall-, T- and N-stage as well as treatment modality and patient age did not exhibit prognostic abilities. This result may be linked to the low event rate of ~8% and the relatively small cohort size. Models combining clinical variables and radiomic signatures performed similarly to the corresponding radiomic models. Additionally, radiomics-based risk-stratification biomarkers identified patients at increased risk of LRP in different follow-up cutoffs (2-, 3-, 4- and 5-year follow-up with $p < 0.05$); whereas clinical variables could not significantly stratify LRP risk ($p > 0.05$). These findings suggest radiomics analysis may be a more powerful means for LRP risk stratification and prognostication than the tested set of potential clinical predictors.

Notably, models integrating PET and CT radiomics outperformed single modality models in most permutations in both primary tumor and combined tumor/lymph node analysis, suggesting complementary prognostic value of “metabolic” and “structural” features derived from PET and CT imaging, respectively. Additionally, consensus VOI combining PET radiomics information from primary tumors and metastatic cervical lymph nodes yielded performance improvements over most corresponding PET primary tumor models. This finding may suggest added prognostic value from PET lymph node features.

Performance curves (Fig. 3) are a valuable tool to investigate model prognostic accuracy throughout a relevant follow-up period, providing a more granular understanding of model performance than summary measures such as Harrell’s C index. We plotted performance curves for select radiomic and combined models as well as the clinical model, again revealing superiority of radiomics-based prognostication. While radiomic models achieved high prognostic performance in follow-up years 1 through 4, model performance was moderate in the fifth year,

which could be related to data sparsity in model training secondary to right censoring.

Accurate contouring of head-and-neck cancer lesions on CT is challenging – especially on pre-contrast images. In this study, we applied manual PET-guided segmentation, allowing full utilization of both accurate PET-guided lesion contouring and standardized CT tissue densities devoid of contrast-induced variability. Notably, analysis of contrast-enhanced CT scans may be limited due to variabilities in contrast accumulation, affecting radiomic feature extraction and reproducibility [14]. Additionally, combined PET and non-contrast CT radiomics analysis extracts both “metabolic” and “structural” tissue density features, allowing comprehensive assessment of primary tumor and metastatic nodes.

Methodologically, our modelling approach relied on random forest machine-learning algorithms: we applied RSF algorithms designed to handle right-censored survival data as well as “classical” RCF models for binary classification [33]. Machine-learning has proven effective in handling the high variable dimensionality commonly associated with radiomics analysis, with random forest models in particular often outperforming other approaches due to superior robustness [11,14,19].

We acquired a multi-national and multi-institutional cohort incorporating data from our institution and several additional centers in Canada and in the United States to increase the cohort size. Additionally, using multi-center data may help augment model robustness to variations among imaging protocols, scanner hardware and image reconstruction and ultimately lead to more generalizable models and model performance estimates.

LRP in HPV-associated OPSCC treated with radiotherapy is rare – in our cohort of 190 subjects, ~8% experienced events – making allocation of independent validation sets in our study challenging. Thus, we applied a rigorous cross validation framework, with particular attention given to avoiding data leakage from training to validation folds; i.e. consensus VOI generation, feature standardization, dimensionality reduction, and RSF fitting were performed on the training folds, and model performance was quantified in the corresponding validation folds. This approach is expected to yield realistic quantification of model performance in new datasets. Nevertheless, future prospective studies with larger study cohorts and higher absolute event counts are required to confirm the prognostic value of quantitative imaging models for LRP prediction. Additionally, our models require independent validation in external cohorts before translation to clinical application may be considered.

Our cohort of 190 patients with HPV-associated OPSCC was acquired from Yale’s Smilow Hospital (2009 to 2019) and two public collections in The Cancer Imaging Archive. PET/CT acquisition and image reconstruction protocols varied over the years and between different cancer centers. This limitation was addressed by adopting a comprehensive image pre-processing pipeline designed to reduce heterogeneity, denoise our dataset, and homogenize PET/CT scans. Nonetheless, standardization of both PET and CT image acquisition across centers and scanner manufacturers may harbor potential for improved radiomics capabilities in OPSCC outcome prognostication and should be pursued as a long-term goal in the field of quantitative imaging.

Manual lesion segmentation is inherently prone to inter- and intra-rater variability as well as limited reproducibility. Despite our efforts to pre-select a subset of robust radiomics features in multiple delineation analysis, fully or partially automating the lesion delineation process may help reduce the aforementioned limitations and ultimately contribute to improved LRP prognostic performance.

Future studies should also incorporate further established LRP predictors into clinical and combined models – e.g. smoking status was unavailable in a considerable portion of our dataset. Finally, metastatic involvement of cervical lymph nodes was determined by expert radiologist assessment, but without histopathological examination of all nodes.

Conclusion

Radiomics analysis decoding metabolic and structural bioimaging patterns of the primary tumor lesion and metastatic nodes in pre-treatment PET/CT scans can provide novel quantitative imaging biomarkers for risk stratification and prediction of post-radiotherapy LRP in HPV-associated OPSCC. Pending independent validation in large external cohorts, such biomarkers may supplement patient selection for trials of treatment de-intensification for prognostically favorable HPV-associated OPSCC, and ultimately guide personalized treatment decision-making.

Declaration of Competing Interest

SPH: None declared. KS: None declared. TZ: None declared. PB: None declared. CR: None declared. RF has acted as speaker and consultant for GE Healthcare and has a research agreement (beta tester) and support from GE Healthcare. RF is also a founder and stockholder of 4intelligent Inc.; and a clinical research scholar (chercheur-boursier clinician) supported by the Fonds de recherche en santé du Québec (FRQS). BHK: None declared. AP: None declared. BLJ: None declared. MLP: None declared. CL has research agreements with Siemens Medical Solutions and GE Healthcare. BB: None declared. AM: None declared. SP: None declared.

CRedit authorship contribution statement

Stefan P. Haider: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing - original draft, Writing - review & editing. **Kariem Sharaf:** Conceptualization, Data curation, Investigation, Writing - original draft, Writing - review & editing. **Tal Zeevi:** Formal analysis, Investigation, Methodology, Writing - review & editing. **Philipp Baumeister:** Investigation, Writing - review & editing. **Christoph Reichel:** Investigation, Writing - review & editing. **Reza Forghani:** Investigation, Writing - review & editing. **Benjamin H. Kann:** Data curation, Investigation, Writing - review & editing. **Alexandra Petukhova:** Investigation, Writing - review & editing. **Benjamin L. Judson:** Investigation, Writing - review & editing. **Manju L. Prasad:** Data curation, Investigation, Writing - review & editing. **Chi Liu:** Investigation, Writing - review & editing. **Barbara Burtness:** Investigation, Writing - review & editing. **Amit Mahajan:** Data curation, Formal analysis, Investigation, Validation, Writing - review & editing. **Seyedmehdi Payabvash:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Writing - review & editing.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.tranon.2020.100906](https://doi.org/10.1016/j.tranon.2020.100906).

References

- [1] R.L. Siegel, K.D. Miller, A. Jemal, Cancer statistics, 2019, *CA Cancer J. Clin.* 69 (1) (2019) 7–34.
- [2] M.L. Gillison, et al., Epidemiology of Human Papillomavirus-Positive Head and Neck Squamous Cell Carcinoma, *J. Clin. Oncol.* 33 (29) (2015) 3235–3242.
- [3] E. Benson, et al., The clinical impact of HPV tumor status upon head and neck squamous cell carcinomas, *Oral Oncol.* 50 (6) (2014) 565–574.
- [4] W.M. Lydiatt, et al., Head and Neck cancers-major changes in the American Joint Committee on cancer eighth edition cancer staging manual, *CA Cancer J. Clin.* 67 (2) (2017) 122–137.
- [5] AJCC Cancer Staging Manual, 8th edition, Springer International Publishing, 2017.
- [6] D.J. Adelstein, et al., Role of treatment deintensification in the management of p16+ oropharyngeal cancer: ASCO provisional clinical opinion, *J. Clin. Oncol.* 37 (18) (2019) 1578–1589.
- [7] S. Marur, et al., E1308: phase II trial of induction chemotherapy followed by reduced-dose radiation and weekly cetuximab in patients with HPV-associated resectable squamous cell carcinoma of the oropharynx- ECOG-ACRIN cancer research group, *J. Clin. Oncol.* 35 (5) (2017) 490–497.
- [8] C. Gronhoj, et al., Pattern of and survival following loco-regional and distant recurrence in patients with HPV+ and HPV- oropharyngeal squamous cell carcinoma: a population-based study, *Oral Oncol.* 83 (2018) 127–133.
- [9] A.W. Joseph, et al., Disease-free survival after salvage therapy for recurrent oropharyngeal squamous cell carcinoma, *Head Neck* 38 (Suppl 1) (2016) E1501–E1509.
- [10] A. Caula, et al., Correlation between the duration of locoregional control and survival in T1-T2 oropharyngeal cancer patients, *Eur. Arch. Otorhinolaryngol.* 276 (4) (2019) 1161–1166.
- [11] R.J. Gillies, P.E. Kinahan, H. Hricak, Radiomics: images are more than pictures, they are data, *Radiology* 278 (2) (2016) 563–577.
- [12] S.S. Yip, H.J. Aerts, Applications and limitations of radiomics, *Phys. Med. Biol.* 61 (13) (2016) R150–R166.
- [13] S.P. Haider, et al., Applications of radiomics in precision diagnosis, prognostication and treatment planning of head and neck squamous cell carcinomas, *Cancers Head Neck* 5 (1) (2020) 6.
- [14] R. Forghani, et al., Radiomics and artificial intelligence for biomarker and prediction model development in oncology, *Comput. Struct. Biotechnol. J.* 17 (2019) 995–1008.
- [15] S.P. Haider, et al., Potential added value of PET/CT radiomics for survival prognostication beyond AJCC 8th edition staging in oropharyngeal squamous cell carcinoma, *Cancers (Basel)* 12 (7) (2020).
- [16] M.R. Folkert, et al., Predictive modeling of outcomes following definitive chemoradiotherapy for oropharyngeal cancer based on FDG-PET image characteristics, *Phys. Med. Biol.* 62 (13) (2017) 5327–5343.
- [17] M. Bogowicz, et al., Combined CT radiomics of primary tumor and metastatic lymph nodes improves prediction of loco-regional control in head and neck cancer, *Sci. Rep.* 9 (1) (2019) 15198.
- [18] G. Feliciani, et al., Radiomic profiling of head and neck cancer: (18)F-FDG PET texture analysis as predictor of patient survival, *Contrast Media Mol Imaging* 2018 (2018) 3574310.
- [19] S. Lever, et al., A comparative study of machine learning methods for time-to-event survival data for radiomics risk modelling, *Sci. Rep.* 7 (1) (2017) 13206.
- [20] S.P. Haider, et al., PET/CT radiomics signature of human papilloma virus association in oropharyngeal squamous cell carcinoma, *Eur. J. Nucl. Med. Mol. Imaging* (2020).
- [21] K. Yu, et al., Radiomic analysis in prediction of Human Papilloma Virus status, *Clin. Transl. Radiat. Oncol.* 7 (2017) 49–54.
- [22] K. Clark, et al., The cancer imaging archive (TCIA): maintaining and operating a public information repository, *J. Digit. Imaging* 26 (6) (2013) 1045–1057.
- [23] M. Vallières, et al., Data from Head-Neck-PET-CT, The Cancer Imaging Archive, 2017.
- [24] A. Grossberg, et al., Data from Head and Neck Cancer CT Atlas, The Cancer Imaging Archive, 2017.
- [25] R.B. Ger, et al., Practical guidelines for handling head and neck computed tomography artifacts for quantitative image analysis, *Comput. Med. Imaging Graph.* 69 (2018) 134–139.
- [26] D.J. Sher, et al., Radiation therapy for oropharyngeal squamous cell carcinoma: executive summary of an ASTRO evidence-based clinical practice guideline, *Pract Radiat Oncol* 7 (4) (2017) 246–253.
- [27] R. Kikinis, S.D. Pieper, K.G. Vosburgh, 3.D. Slicer, Springer New York, New York, NY, 2014, pp. 277–289. Editor.
- [28] Definition of Volumes, *J. ICRU* 10 (1) (2010) 41–53.
- [29] Pyradiomics-community Pyradiomics Documentation Release 2.1.2, 2018 [cited 2019 December 15th, 2019]; Available from: <https://readthedocs.org/projects/pyradiomics/downloads/pdf/2.1.2/>.
- [30] A. Traverso, et al., Repeatability and reproducibility of radiomic features: a systematic review, *Int. J. Radiat. Oncol. Biol. Phys.* 102 (4) (2018) 1143–1158.
- [31] L. Lu, et al., Robustness of radiomic features in [(11)C]Choline and [(18)F]FDG PET/CT imaging of nasopharyngeal carcinoma: impact of segmentation and discretization, *Mol. Imaging Biol.* 18 (6) (2016) 935–945.
- [32] R.T. Leijenaar, et al., Stability of FDG-PET Radiomics features: an integrated analysis of test-retest and inter-observer variability, *Acta Oncol.* 52 (7) (2013) 1391–1397.
- [33] M.N. Wright, A. Ziegler, Ranger: a fast implementation of random forests for high dimensional data in C plus plus and R, *J. Stat. Softw.* 77 (1) (2017) 1–17.
- [34] M. Schmid, M. Wright, A. Ziegler, On the use of Harrell's C for clinical risk prediction via random survival forests, arXiv e-prints (2015).
- [35] R Development Core Team: A Language and Environment For Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2019.
- [36] R.R. Bouckaert, E. Frank, in: Evaluating the Replicability of Significance Tests For Comparing Learning Algorithms, Springer Berlin Heidelberg, Berlin, Heidelberg, 2004, pp. 3–12.
- [37] H. Uno, et al., Evaluating prediction rules for t-year survivors with censored regression models, *J. Am. Stat. Assoc.* 102 (478) (2007) 527–537.
- [38] Potapov, S., W. Adler, and M. Schmid, survAUC: Estimators of Prediction Accuracy for Time-to-Event Data. 2012.

- [39] P. Pipkorn, et al., Outcomes of relapsed human papillomavirus-related oropharyngeal squamous cell carcinoma treated with curative intent, *Head Neck* 41 (5) (2019) 1312–1319.
- [40] J. Zenga, et al., Salvage of recurrence after surgery and adjuvant therapy: a multi-institutional study, *Otolaryngol. Head Neck Surg.* 161 (1) (2019) 74–81.
- [41] B. Burtneß, et al., Pembrolizumab alone or with chemotherapy versus cetuximab with chemotherapy for recurrent or metastatic squamous cell carcinoma of the head and neck (KEYNOTE-048): a randomised, open-label, phase 3 study, *Lancet* 394 (10212) (2019) 1915–1928.