Framework for Machine Learning of CT and PET Radiomics to Predict Local Failure after Radiotherapy in Locally Advanced Head and Neck Cancers

Devadhas Devakumar[1], Goutham Sunny[2,3], Balu Krishna Sasidharan[2], Stephen R. Bowen[4], Ambily Nadaraj[5], L. Jeyseelan[5], Manu Mathew[2], Aparna Irodi[6], Rajesh Isiah[2], Simon Pavamani[2], Subhashini John[2], Hannah Mary T. Thomas[2]

Departments of [1]Nuclear Medicine, [2]Radiation Oncology, [5]Clinical Epidemiology and [6]Radiology, Christian Medical College, Vellore, Tamil Nadu, [3]Department of Radiation Oncology, Baptist Cancer Centre, Bangalore Baptist Hospital, Bengaluru, Karnataka, India, [4]Department of Radiation Oncology, School of Medicine, University of Washington, Seattle, WA 98195, USA

Abstract

Context: Cancer Radiomics is an emerging field in medical imaging and refers to the process of converting routine radiological images that are typically qualitatively interpreted to quantifiable descriptions of the tumor phenotypes and when combined with statistical analytics can improve the accuracy of clinical outcome prediction models. However, to understand the radiomic features and their correlation to molecular changes in the tumor, first, there is a need for the development of robust image analysis methods, software tools and statistical prediction models which is often limited in low- and middle-income countries (LMIC). **Aims:** The aim is to build a framework for machine learning of radiomic features of planning computed tomography (CT) and positron emission tomography (PET) using open source radiomics and data analytics platforms to make it widely accessible to clinical groups. The framework is tested in a small cohort to predict local disease failure following radiation treatment for head-and-neck cancer (HNC). The predictors were also compared with the existing Aerts HNC radiomics signature. **Settings and Design:** Retrospective analysis of patients with locally advanced HNC between 2017 and 2018 and 31 patients with both pre- and post-radiation CT and evaluation PET were selected. **Subjects and Methods:** Tumor volumes were delineated on baseline PET using the semi-automatic adaptive-threshold algorithm and propagated to CT; PyRadiomics features (total of 110 under shape/intensity/texture classes) were extracted. Two feature-selection methods were tested for model stability. Models were built based on least absolute shrinkage and selection operator-logistic and Ridge regression of the top pretreatment radiomic features and compared to Aerts' HNC-signature. Average model performance across all internal validation test folds was summarized by the area under the receiver operator curve (ROC). **Results:** Both feature selection methods selected CT features MCC (GLCM), SumEntropy (GLCM) and Sphericity (Shape) that could predict the binary failure status in the cross-validated group and achieved an AUC >0.7. However, models using Aerts' signature features (Energy, Compactness, GLRLM-GrayLevelNonUniformity and GrayLevelNonUniformity-HLH wavelet) could not achieve a clear separation between outcomes (AUC = 0.51–0.54). **Conclusions:** Radiomics pipeline included open-source workflows which makes it adoptable in LMIC countries. Additional independent validation of data is crucial for the implementation of radiomic models for clinical risk stratification.

Keywords: Computed tomography, head and neck cancer, local failure, machine learning, positron emission tomography, radiomics

INTRODUCTION

Cancer radiomics is an emerging field in medical imaging and refers to the process of converting routine radiological images that are typically qualitatively interpreted to quantifiable descriptions of the tumor phenotypes such as size (volume), shape (sphericity, compactness), texture (voxel heterogeneity, coarseness, contrast) and tumoral intensity (uniformity, entropy), and when combined with statistical analytics can

Address for correspondence: Dr. Hannah Mary T. Thomas,
Department of Radiation Oncology, Christian Medical College, Vellore,
Tamil Nadu, India.
E-mail: hannah.thomas@cmcvellore.ac.in

Access this article online

Quick Response Code:

Website:
www.jmp.org.in

DOI:
10.4103/jmp.JMP_6_21

How to cite this article: Devakumar D, Sunny G, Sasidharan BK, Bowen SR, Nadaraj A, Jeyseelan L, *et al.* Framework for machine learning of CT and PET radiomics to predict local failure after radiotherapy in locally advanced head and neck cancers. J Med Phys 2021;46:181-8.

improve the accuracy of clinical outcome prediction models. Radiomics operates under the assumption that these image features are linked to the tumor gene-protein signatures or the tumor phenotype.[1,2] Unlike tissue biopsies, radiomics accounts for the whole 3D tumor and its heterogeneity and can noninvasively characterize it. This information from routine clinical images obtained before and during the course of treatment with no additional scans or added dose burden to the patients makes it an economical approach toward personalization of treatment. Furthermore, prediction of outcome using pretreatment imaging can help plan treatment strategies and augment the clinical decision-making process, thus preparing clinicians and patients for the appropriate course of treatment.

Researchers have identified and quantified several radiomic features and showed that these imaging descriptors can monitor treatment and predict outcomes in different cancer types.[3,4] Combining the data from radiomics, genomics and clinical assays improved outcome prediction accuracy over that of individual assays.[5] Given the number of studies showing the efficacy of integrating radiomics features for differentiating tumor types (benign or malignant, responders or nonresponders to treatment regimen) or predicting treatment outcomes (toxicity, survival, recurrence),[3,4,6] incorporating them into decision support systems are on the precipice of clinical translation, which will have a significant impact on personalization of cancer treatment, especially in resource-limited settings like India. However, radiomics is not yet deployed in routine clinical settings in India and is limited to a few research investigations.[7-9] To understand the radiomic features and their correlation to molecular changes in the tumour, first, there is a need for the development of robust image analysis methods, software tools and statistical prediction models. These lacunae were the motivation for this study, and an attempt is made to develop a pipeline that can speed the translation of the radiomics research in low and middle income countries (LMIC).

In this study, we focussed on head-and-neck cancers (HNC) since it contributes about 25%–30% of all cancers in India as opposed to 3%–4% in the Western world[10-12] and with >20,000 new cases being reported yearly since 2018.[13] These cancers are different with respect to the patient characteristics, disease presentation, and etiological differences compared to the West. Some of the differences include that Indian patients often present themselves at a very advanced stage of disease, lower incidence of HPV related cancers affecting the oropharynx, unusual presentations of HNC due to oral/chewable tobacco consumption with increased risk due to poor oral hygiene among certain sections (individuals with low educational status, farmers or manual labourers) and higher incidence in younger patients (<40 years) compared to the West (55–60 years).[11,14] These tumours are often challenging to treat due to many factors but not limited to complex regional anatomy, high intra-tumoural heterogeneity, close proximity of critical structures and major changes to the anatomy related to tumor

response. Risk stratification is important for an oncologist to predict early in treatment whether the tumor is likely to respond to a particular therapy. The important outcome parameters that are used for clinical decision-making in HNC are loco-regional failure at 1 year; 2) distant failure and 3) 5-year overall survival. Studies have shown that 1-year disease-free survival strongly correlates to overall survival.[15] With the local recurrence rate in advanced HNC ranging from 15% to 50%[16] more accurate predictors of early disease failure are imperative as treatment could be modified to intensify therapy individualized to the high-risk patients.

The aim of this study is to report a framework for machine learning-based radiomic intensity, shape, and texture features extracted from computed tomography (CT) and positron emission tomography (PET) imaging from head-and-neck patients as a foundational step towards future independent testing and validation.

SUBJECTS AND METHODS

Patient selection
Study participants were retrospectively recruited from patients treated between February 2017 and March 2018, if they had baseline CT and PET scans and evaluation scans within 10 weeks posttreatment. All patients were diagnosed with histologically confirmed locally advanced HNC of the oropharynx, hypopharynx, laryngopharynx or oral cavity. Patients with recurrent HNC or with metastases at presentation, received any prior radiation, and patients on palliative treatment were excluded from the study. The median follow-up period of the cohort was 110 months (range: 70–166). Imaging was performed as per standard clinical routine. The study was conducted in full accordance with ethical principles, including the World Medical Association Declaration of Helsinki (version 2002). The use of the anonymized retrospective data without additional need for informed consent was approved by the Institutional Review Board of the study institution.

Treatment details
Patients included in this study were treated with concurrent chemo-radiation or definitive radiotherapy alone. The radiotherapy regimen was planned with volumetric modulated arc radiotherapy. In most cases (18/31), the concurrent chemotherapy was cisplatin 40 mg/m^2 IV every week; others were given Nimotuzumab (5/31) 200 mg IV once a week. The radiomics worflow is depicted in Figure 1.

Image acquisition
The imaging data consisted of baseline PET/CT that was used for treatment planning. The 3-month posttreatment PET/CT were included to extract delta features to assess the stability of the radiomic features and was excluded for all further analyses. All imaging data were acquired on Siemens Biograph 6 PET-CT scanner (Siemens Medical Solutions, USA) with immobilization in radiation treatment position. The

CT images were first acquired at 130 kVp, and 500–700 mm field-of-view (FOV). The CT images were reconstructed with a transaxial plane resolution of 0.97–1.36 mm and slice thickness of 3–5 mm per slice; the image size was 512 × 512 pixels. The PET images were acquired with 500–700-mm FOV and reconstructed using attenuation weighted ordered subset expectation maximization reconstruction (3 iterations, 21 subsets, and 4-mm FWHM Gaussian filter). During the reconstruction, corrections for point spread function (TrueX), attenuation, and scatter were performed. The sinogram data were decay, dead time, and random corrected. The CT images were used for the attenuation correction. The resultant PET image size was 168 × 168 pixels and the in-plane resolution were 4.07 mm × 4.07 mm with slice thickness ranging between 3 and 5 mm. Pre-and post-treatment PET/CT scans were registered using rigid registration on to a common grid.

Tumor delineation

To reduce the inter-observer variability in the tumor definition, the gross/metabolic tumor volume (MTV) was delineated using an adaptive threshold segmentation followed by a gradient segmentation. As the first step, the MTV defining the extent of the metabolic uptake was delineated on the pretreatment PET. Approximate location of the primary tumor was identified and a fiducial was placed. A sphere ROI was drawn around the fiducial and the maximum standard uptake value (SUVmax) within the ROI was found. The color scale was changed to cold-to-hot and the range of the color scale was adjusted from 0 to SUVmax which helped identify the tumour and its margins clearly. Additional three fiducials were placed around the tumor to represent the background region that lie close to the tumor. The SUV values from each of the background fiducial were averaged to obtain the local background SUVmean. The initial threshold value for segmentation was calculated from the SUVmax and the mean background value. An initial tumour delineation was performed using the initial threshold value. If the segmented region extended beyond the tumor into any adjacent regions, the adjoined regions were manually erased to disconnect them. The tumor island was retained while discarding the adjacent segmented regions. The initial volume of the tumor was calculated from the initial threshold segmentation region. The final threshold value was calculated using the Growcut algorithm as described in the article by Thomas *et al.*[17] The PET contour was propagated to the planning CT. Minor edits were made if required to include any of the regions not included in the PET to generate the gross tumor volume (GTV), which is the extent of the tumor as seen on the CT. The display protocol was a fixed window/level setting for the CT scan (brain window). All the above-mentioned steps for tumor delineation has been customized as python-script based software and operated as a module within 3D Slicer, an imaging platform that is free, open-source and easily downloadable for imaging-related research.[18] The MTV and GTV contours were verified by a radiation oncologist with 12 years of experience in treating HNC. Example delineations are shown in Figure

2. The pretreatment MTV and GTV were propagated to the corresponding posttreatment evaluation scans using rigid registration to allow us to extract features from the same ROIs.

Labeling

When the lesion volume decrease between the pretreatment CT and PET and evaluation scan 3 months posttreatment were <25% and presence of new disease or recurrence confirmed following imaging and clinical examination was found, the patient was classified as having a loco-regional failure. This failure status of the patients at their first follow-up visit following radiotherapy were the labels that was used for the model fitting.

Feature extraction

Before feature extraction, the pre/post-treatment CT and PET scans were resampled to an isotropic 1 mm and 4 mm resolution, respectively using cubic B-spline interpolation. From the MTV and GTV lesion ROIs, 110 features were extracted using Pyradiomics toolbox (v 2.1) made available in 3D Slicer with fixed bin count gray level discretization (CT: 128, PET: 64).[18,19] The feature space included 19 intensity and first-order features, 16 shape and size-based features and 75 textural features representing the spatial distribution of voxel intensities derived from GLCM, GLDM, GLRLM, GLSZM and NGTDM. Compactness 2 and Gray Level NonUniformity-HLH were independently extracted on the CT to compare with the Radiomics features by Vallières *et al.*[20] Those two features were not part of the feature selection and model building approach. The entire list and mathematical definition of these features are made available by Van Griethuysen *et al.*[19] Most features are in compliance with the Imaging Biomarker Standardization Initiative and are reported by Zwanenburg *et al.*[21]

Dimensionality reduction

Although 110 features were extracted for each patient per image set, all features did not contribute equally to discriminating loco-regional failure. The features with low discriminative power or those that are highly correlated with each other tend to overfit the classifiers leading to a poor outcome. To avoid this, rigorous feature selection with dimensionality reduction was performed to find a set of candidate features with excellent discrimination capabilities and significant differences before the outcome prediction. As a first step, all features were scaled *via* Z-score normalization to improve stability and speed of optimization algorithms. Two methods were implemented for dimensionality reduction and model sensitivity perturbation: Method 1 (M1) included (i) inter-patient variance inflation, (ii) pre/post-treatment variance inflation, (iii) co-linearity reduction, (iv) bootstrap iterations of least absolute shrinkage and selection operator (LASSO). These filtering steps were to exclude features that did not vary across patients or did not vary between pre-and post-treatment imaging time points, and those that were strongly correlated to other features, thus ensuring minimal information loss. In M1, the variance of pretreatment features between patients was

computed and the features with inter-patient variance (range/median) >10% and <100% were retained. Following this step 95 features remained. The delta features were calculated from the ratio of pre-and post-CT and PET and features that varied >10% (range/median) were retained. After this step 85 features remained. In M1(iii) the highly correlated feature pairs (Spearman correlation coefficient ρ >0.95), the feature with the least variance from M1 (i) and M1 (ii) were dropped. Following M1 (iii) 69 features remained. Finally, M1 (iv) the LASSO with bootstrapping was applied for 60 iterations and the nonzero weighted features were tallied. A frequency distribution of the nonzero weighted features was constructed and the top-ranked 5 features were retained for the prediction model building. Method 2 (M2) included only M1 (iv), i.e. the LASSO based feature selection over 60 bootstrapped iterations. All 110 features were included, and a frequency distribution table was constructed for the nonzero weighted features at the end of each iteration and tallied. The top 5 features were retained for the prediction model building.

Model performance evaluation

Following dimensionality reduction, multivariate logistic regression models with LASSO and Ridge regularization were constructed to predict the incidence of 1-year loco-regional failure in HNC patients. The true error estimate bias was minimized by applying the stratified 5-fold cross-validation, where "stratified" refers to the different class levels in each fold being represented in identical proportion as that in the full dataset. The benefit of using stratified cross-validation has been highlighted in Witten *et al*.[22] The evaluation metric for the models was the area under the receiver operating characteristic curve (AUC) of the binary outcomes, i.e. "loco-regional failure" "no loco-regional failure." The averages AUCs of the testing runs were reported. The best model for predicting 1 year loco-regional disease failure was determined by maximizing the 5-fold cross-validated AUCs. The DeLong's test[23] was used for estimating the confidence interval of the difference of AUC values between the models generated using features selected on our data and using Aerts' signature features.

For statistical analysis, visualization, model building and testing Orange, an open-source Python-based data mining software[24] was used.

The radiomics analysis was performed on a MacBook Pro 2.6Ghz Quad-Core Intel Core i7 Processor with 16 GB memory, and NVIDIA GeForce GT 650 M I TB graphics card.

RESULTS

Patient demographics

The clinical information of the patients and tumor characteristics in this study are summarized in Table 1.

The training cohort included 31 patients (28 men and 3 women). All patients included were of Indian origin. The median follow-up period was 92 days (range 70–187 days); the median age of this population was 55 years, and the majority of

Table 1: Head-and-neck cancer patients' information and tumor characteristics in the study

Characteristic	Type		Number of patients, *n* (%)
Gender	Male		28 (90)
	Female		3 (10)
Age (years)	Range		30-80
	Mean±SD		57±10
Tumor type	Oropharynx		9 (29)
	HPV positive	4	
	HPV negative	2	
	Not available	3	
	Hypopharynx		9 (29)
	Larynx		8 (26)
	Oral cavity		4 (13)
	Unknown		1 (3)
T-stage	T0		1 (3)
	T1		1 (3)
	T2		10 (32)
	T3		12 (39)
	T4		7 (23)
N-stage	N0		14 (45)
	N1		6 (19)
	N2		11 (36)
TNM stage	Stage-II		5 (16)
	Stage-III		14 (45)
	Stage-IV		12 (39)
Treatment	Radiation only		8 (26)
	Chemo-radiation		23 (74)
Outcome	Loco-regional recurrence		6 (19)

SD: Standard deviation, TNM: Tumor-node-metastasis, HPV: Human papillomavirus

the patients (90%) were men. Only one patient was classified as M1, and the others were classified as M0 in the TNM staging. Six of the 31 patients had loco-regional failure at their first follow-up postradiation treatment.

Feature selection and model building

Following robust and rigorous feature selection both methods (M1 and M2) selected three features from 75% to 84% of bootstrapped LASSO iterations for predicting local disease failure in (6/31) patients. The features were namely MCC (GLCM) SumEntropy (GLCM) and Sphericity (Shape). However, none of the PET features was selected by the LASSO algorithm as they did not appear consistently in the 60 iterative runs. MCC had the highest odds ratio per standard deviation increase in L1 and L2 models followed by sum entropy and sphericity. The L1 and L2 models achieved a 5-fold cross-validation classification performance of AUC = 0.73 and 0.79 respectively [Table 2].

The selected radiomic features (MCC, SumEntropy and Sphericity) were compared to the 4 features (Energy, Compactness, GrayLevelNonUniformity and Wavelet of GrayLevelNonUniformity) of the radiomics signature developed by Aerts *et al*.[25] The odds ratios of these features are

tabulated in Table 3. The shape feature (Sphericity) from our model was the only similar feature to the Aerts' model where Compactness is defined as the cube of sphericity. The Aerts' features based model could not achieve a clear separation between patients with and without loco-regional failure in both L1 (AUC = 0.51) and L2–AUC = 0.53) regression models.

The receiver-operating curves for the different models, both trained using our data and generated using the Aerts' features are shown in Figure 3. The difference between the AUCs for the ROCs for these paired models as per Delong's test was not found to be significant (z scores − 0.18 to − 1.53, $P > 0.05$).

DISCUSSION

Risk stratification is becoming an important strategy to guide treatment decisions in oncology. Molecular studies can assess certain tumor mutations that contribute to treatment failure. The early results for the International Cancer Genome Consortium study on oral squamous cell carcinoma in Indian populations revealed molecular subtypes with distinctive mutational profiles such as patients predominantly harboring mutations in CASP8 with or without mutations in FAT1 with the mean duration of disease-free survival being significantly elevated in some molecular subgroups.[26] However, the molecular tests are not a viable option for all patients in India and other LMIC

Table 2: Odds ratios (per standard deviation increase) and area under the receiver operating characteristic curves with least absolute shrinkage and selection operator (L1) and ridge (L2) logistic regression models

Radiomic features	LASSO (L1)	Ridge (L2)
MCC (GLCM)	3.33	3.06
SumEntropy (GLCM)	2.22	2.26
Sphericity (shape)	1.20	1.40
AUC	0.73	0.79
Classification accuracy	0.81	0.81
Precision	0.76	0.65

LASSO: Least absolute shrinkage and selection operator, AUC: Area under the receiver operating characteristic curve, MCC: Maximal correlation coefficient, GLCM: Gray level cooccurrence matrix

Table 3: Odds ratios (per standard deviation increase) and area under the receiver operating characteristic curve with least absolute shrinkage and selection operator (L1) and ridge (L2) logistic regression models for features from the radiomics signature by Aerts *et al*.

Radiomic features	LASSO (L1)	Ridge (L2)
Energy (first order statistics)	1.43	1.63
Compactness (shape)	1.28	1.45
GrayLevelNonUniformity (GLRLM)	1.76	1.99
GrayLevelNonUniformity (wavelet-HLH)	1.06	1.22
AUC	0.51	0.54

AUC: Area under the receiver operating characteristic curve, LASSO: Least absolute shrinkage and selection operator, GLRLM: Gray-level run-length matrix

owing to their high cost. On the other hand, all HNC patients would routinely undergo radiological imaging such as CT, magnetic resonance and PET[27] as part of their treatment which includes combinations of surgery, radiation and chemotherapy. Radiomics-based prediction models also have a potential for universal implementation as a cost-effective surrogate to genomics-based biomarkers. Risk stratification that rely solely on clinical models may improve with quantitative imaging-based biomarkers (radiomic signatures) in predicting outcome.[4]

However, to understand the radiomic features and their correlation to molecular changes in the tumor, there is a need for the development of robust image analysis methods, software tools and statistical prediction models. It will translate radiomic signatures into routine cancer care to support clinical decision-making. Most of these radiomics analysis pipelines have been in practice in developed countries and their implementation in LMIC is almost absent. Results of a PubMed search of publications with keyword "Radiomics" in the last 5 years is shown in Figure 4. The figure clearly shows that dissemination of radiomics research to LMIC is currently limited. Radiomics research publications from the African continent, Mexico and the rest of Europe and Asia are <1% in the last 5 years and is shown as a pictorial representation in Figure 4.

The novelty of this work is that we have put together a radiomics framework that can be easily adopted to many radiation oncology centers. In this study, radiomics pipeline is presented that uses open-source platforms end-to-end from tumour segmentation and feature extraction (3D Slicer, Pyradiomics) to radiomics based model building and validation (Orange Data mining). The bespoke hybrid tumor delineation software has been developed in Python and it can be easily operated as a module within 3D Slicer. Pyradiomics follow feature definitions prescribed by the Imaging Biomarker Standardization that allows comparison and sharing of reported models.[19] The Orange data mining software[24] is simple with an easy to use interface which makes it be easily adoptable for both single institutional research and multi-institutional collaborative work.[28] This study has evaluated the radiomic features in HNC in the Indian population for the first time in its history and compare it the radiomics signature that has been widely validated in other cohorts.[19,29,30]

In this study, a radiomics framework was tested in building outcome prediction models using machine learning of radiomic features that were derived from baseline planning CT and PET imaging data. The study included 31 HNC patients from single institution treated between 2017 and 2018 and who had baseline and postradiation follow up CT and PET scans. The tumor volumes were defined on the CT and PET images using the adaptive threshold segmentation method and analyzed 110 radiomic features in each dataset, quantifying the differences in the tumor phenotypes based on shape, image intensity and texture. Only 3 CT radiomic features were selected following
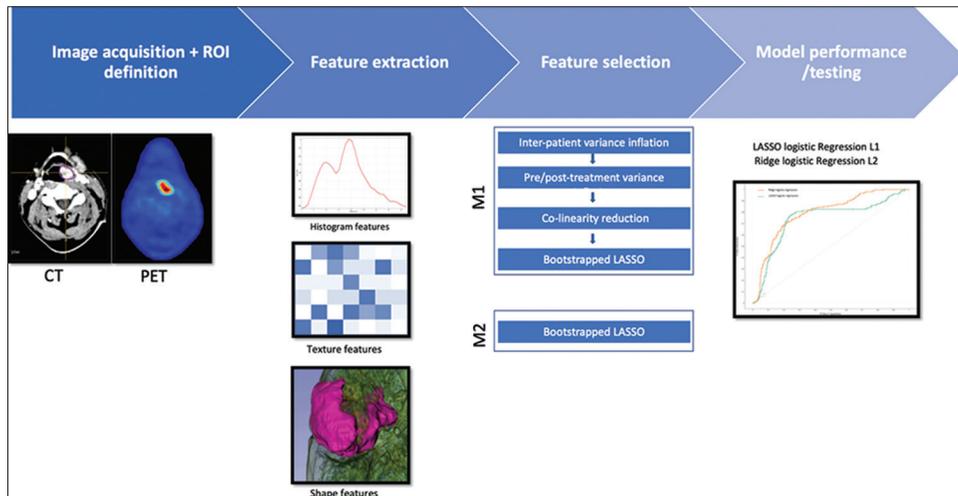
Figure 1: Radiomics and machine learning pipeline for outcome (1-year loco-regional failure) prediction in head-and-neck cancer
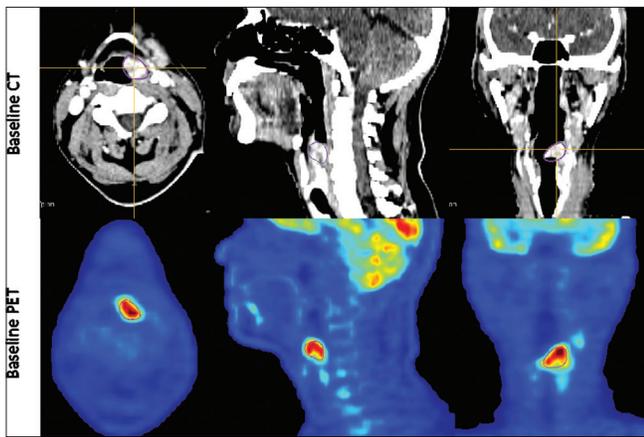
Figure 2: Representative patient images for baseline CT (top row) and PET (bottom row) with tumor defined using automated segmentation algorithm

two LASSO based feature selection methods. None of the PET features was predictive of the loco-regional failure in this cohort unlike the study reported by Bogowicz and Lambin.[31]

The clinical characteristics of the HNC1 and HNC2 cohorts from MAASTRO and VUMC included by van Griethuysen *et al*.[19] were compared to the characteristics of patients in this study while validating the model. Similar to their study all patients included in this study had squamous cell carcinoma. The tumor types included by HNC1 were cancers of the oropharynx and larynx while HNC2 included only oropharyngeal tumors. In contrast, our study included more heterogeneity in the types of tumors. This is also representative of HNC seen in the Indian subcontinent. Cancers of the oral cavity and hypopharynx is notably more in our population compared to the West and this is captured in our study. The incidence is high due to the use of chewing tobacco.[11,13,32] Most of the patients (71%) were from T2–T3, unlike the two cohorts where there was almost equal distribution of patients from stages T1 to T4. No patient was included in stage I, 16% stage II TNM grouping were included in this study in

contrast to 19% and 8% Stage 1 and 8% and 19% Stage II in HNC1 in HNC2, respectively. This is indicative that patients present themselves at a much later stage in India compared to the West. Male preponderance is seen in our cohort with 90% being men which is truly representative of the trends seen in the country compared to 81% and 65% in HNC1 and HNC2, respectively. Only 66% of the patients had p16 testing which is a surrogate of HPV status unlike 100% in the other two cohorts. While the West has seen an increase in the prevalence of HPV related cancers this changing epidemiology has not been extensively studied in our population. Hence, this test is not routinely performed because the data is sparse and the clinical significance is currently unknown.[33,34] The treatment regime was the opposite of the HNC cohort where 74% of patients received radiation alone which is currently the standard practice and 26% had chemo-radiation. In India, >65% follow the practice of the weekly Cisplatin regime.[33]

When the univariate analysis was performed on the clinical parameters none of the characteristics shown in Table 1 showed any significant difference. Tumor volume alone showed significant difference between patients with and without loco-regional recurrence ($P = 0.003$, Mann–Whitney). However, when a clinical model was built with volume as a single feature both the regression models L1 and L2 showed poor performance for this study cohort (AUC = 0. 3). There was a weak correlation between selected radiomic features and tumor volume ($r < 0.5$, Spearman rank). When tumor volume was combined with the radiomics model and predictive performance dropped (AUC = 0.73 L1 and AUC = 0.75 L2) indicating that this feature was just adding more noise to the model.

In our study, we have used a semi-automatic tumor volume delineation method using a framework developed in 3D Slicer and tumor volumes were verified by an expert clinician. Although it is not ideal, it is better than using manual contours that are not reproducible, which in turn affects the radiomic features. The feature extraction was performed
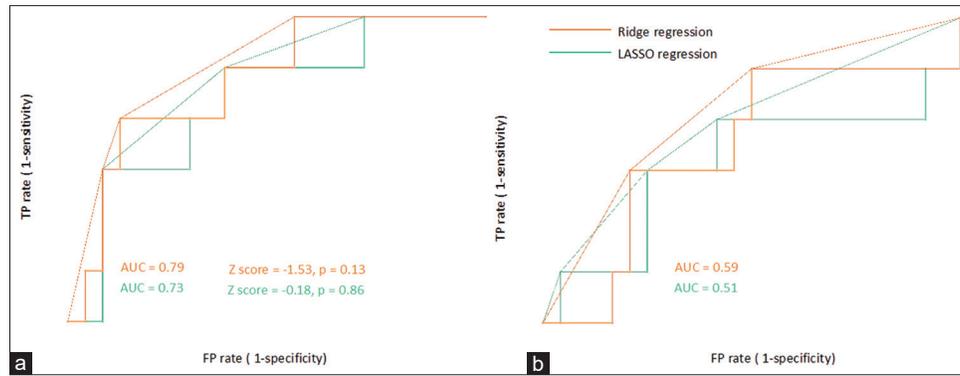
Figure 3: The receiver operating curves (ROC) generated (a) for models trained using retrospective data (b) models tested using features reported in the Aerts' signature for HNC
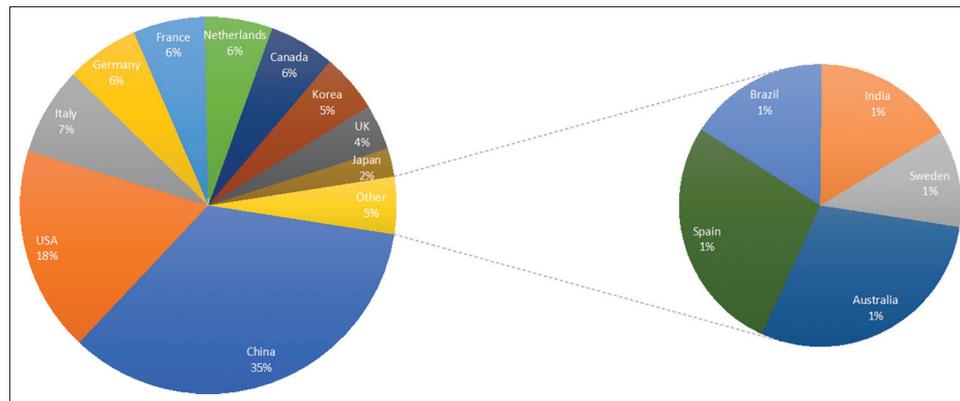
Figure 4: Distribution of publications on radiomics in the last 5 years

using Pyradiomics which makes it possible for comparison with existing radiomics signatures. For this study, the fixed bin width discretization of 64 and 128 for the CT and PET, respectively were used unlike the fixed bin size of 25 employed similar to the study by Aerts *et al.*[20,25] Appropriate discretization of images avoids over or under binning of the image and help with noise reduction while doing texture analysis. There was no volume dependence of the selected radiomics features and when used in a larger validation cohort these features might have the potential to be independent predictive imaging biomarkers.

One of the several limitations of this study included large heterogeneity in image acquisition parameters primarily in the PET images, part of which was harmonized by resampling the data to a common voxel grid. However, none of the PET radiomic features was included in the final model as they were not highly ranked by both M1 and M2 feature selection steps. Our sample size is small, because of our single-institution retrospective study design. However, our single-institution focus enabled us to acquire detailed information on the characteristics of the patients, namely, tumor stage, therapy details and patterns of failure. Although 1 year follow-up may not be ideal, based on the study by Kissun *et al.*[35] the reported median time of recurrence in HNC was 8 months after

the initial surgery and in majority of the patients (90%) was within 2 years. Since this is only the discovery phase of the radiomic models we decided to include patients for whom outcome data were available from the first follow-up visit. The results from these models will be tested in a larger cohort with much longer follow-up. The retrospective nature of the work and the lack of external validation limits the generalizability of our results at this stage. However, this study draws attention to challenges in the application of imaging data science when investigating specific oncologic outcomes, including standardization and harmonization of radiomics pipelines in LMIC. Hence this study only provides preliminary discovery phase models and an external validation is necessary with a larger cohort. Our group has started recruiting patients on a large (>400) prospective observational trial to mitigate such limitations.

CONCLUSIONS

This study reported a framework for radiomics based outcome prediction modeling and was tested as a fundamental step in outcome prediction in locally advanced HNC. In this study, only CT-based radiomic models were able to make a reasonable separation between patients with and without loco-regional failure. Although the feasibility of the radiomics based prediction models has been demonstrated, additional

independent validation data is essential. This study's results will be validated in the future prospective radiomics trial to develop clinical decision support tools.

Conflicts of interest

There are no conflicts of interest.

REFERENCES

1. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images are more than pictures, they are data. Radiology 2016;278:563-77.
2. Lambin P, Leijenaar RT, Deist TM, Peerlings J, de Jong EE, van Timmeren J, *et al.* Radiomics: The bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 2017;14:749-62.
3. Sanduleanu S, Woodruff HC, de Jong EE, van Timmeren JE, Jochems A, Dubois L, *et al.* Tracking tumor biology with radiomics: A systematic review utilizing a radiomics quality score. Radiother Oncol 2018;127:349-60.
4. Parekh V, Jacobs MA. Radiomics: A new application from established techniques. Expert Rev Precis Med Drug Dev 2016;1:207-26.
5. Grossmann P, Stringfield O, El-Hachem N, Bui MM, Rios Velazquez E, Parmar C, *et al.* Defining the biological basis of radiomic phenotypes in lung cancer. Elife 2017;6:e23421.
6. Cook GJ, Azad G, Owczarczyk K, Siddique M, Goh V. Challenges and promises of PET radiomics. Int J Radiat Oncol 2018;102:1083-9.
7. Anbumani S, Jayaraman P, Anchaneyan P, Bilimagga RS, Arunai Nambiraj N. Quantitative radiomic phenotyping of cervix cancer. Int Clin Pathol J. 2018;6:26-8. DOI: 10.15406/icpjl.2018.06.00149.
8. Patil R, Mahadevaiah G, Dekker A. An approach toward automatic classification of tumor histopathology of non–small cell lung cancer based on radiomic features. Tomography 2016;2:374-7.
9. Sapate SG, Mahajan A, Talbar SN, Sable N, Desai S, Thakur M. Radiomics based detection and characterization of suspicious lesions on full field digital mammograms. Comput Methods Programs Biomed 2018;163:1-20.
10. Gatta G, Botta L, Sánchez MJ, Anderson LA, Pierannunzio D, Licitra L, *et al.* Prognoses and improvement for head and neck cancers diagnosed in Europe in early 2000s: The EUROCARE-5 population-based study. Eur J Cancer 2015;51:2130-43.
11. Dandekar M, Tuljapurkar V, Dhar H, Panwar A, DCruz AK. Head and neck cancers in India. J Surg Oncol 2017;115:555-63.
12. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. CA Cancer J Clin 2016;66:7-30.
13. Poddar A, Aranha R, Royam MM, Gothandam KM, Nachimuthu R, Jayaraj R. Incidence, prevalence, and mortality associated with head and neck cancer in India: Protocol for a systematic review. Indian J Cancer 2019;56:101.
14. Sharma JD, Baishya N, Kataki AC, Kalita CR, Das AK, Rahman T. Head and neck squamous cell carcinoma in young adults: A hospital-based study. Indian J Med Paediatr Oncol 2019;40:18.
15. Bourhis J, Le Maître A, Baujat B, Audry H, Pignon JP; Meta-Analysis of Chemotherapy in Head; Neck Cancer Collaborative Group, *et al.* Individual patients' data meta-analyses in head and neck cancer. Curr Opin Oncol 2007;19:188-94.
16. Chang JH, Wu CC, Yuan KS, Wu AT, Wu SY. Locoregionally recurrent

17. Thomas HM, Devakumar D, Sasidharan B, Bowen SR, Heck DK, James Jebaseelan Samuel E. Hybrid positron emission tomography segmentation of heterogeneous lung tumors using 3D Slicer: improved GrowCut algorithm with threshold initialization. J Med Imaging. 2017;4:011009. doi: 10.1117/1.JMI.4.1.011009. Epub 2017 Jan 23. PMID: 28149920; PMCID: PMC5253402.
18. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin J-C, Pujol S, *et al.* 3D slicer as an image computing platform for the quantitative imaging network. Magn Reson Imaging 2012;30:1323-41.
19. van Griethuysen JJ, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, *et al.* Computational radiomics system to decode the radiographic phenotype. Cancer Res 2017;77:e104-7.
20. Vallières M, Kay-Rivest E, Perrin LJ, Liem X, Furstoss C, Aerts HJ, *et al.* Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer. Sci Rep 2017;7:10117.
21. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, *et al.* The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. Radiology. 2020 May;295(2):328-338. doi: 10.1148/radiol.2020191145.
22. Witten IH, Frank E, Hall MA. Data mining: Practical machine learning tools and techniques Morgan Kaufmann 2016.
23. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. Biometrics 1988;44:837-45.
24. Demsˇar J, Curk T, Erjavec A, Demsar J, Curk T, Erjave A, *et al.* Orange: Data mining toolbox in python. J Mach Learn Res 2013;14:2349-53.
25. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, *et al.* Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 2014;5:1-9.
26. India Project Team of the International Cancer Genome Consortium. Mutational landscape of gingivo-buccal oral squamous cell carcinoma reveals new recurrently-mutated genes and molecular subgroups. Nat Commun 2013;4:2873.
27. Wong AJ, Kanwar A, Mohamed AS, Fuller CD. Radiomics in head and neck cancer: From exploration to application. Transl Cancer Res 2016;5:371-82.
28. Koçak B, Durmaz EŞ, Ateş E, Kılıçkesmez Ö. Radiomics with artificial intelligence: A practical guide for beginners. Diagn Interv Radiol 2019;25:485-95.
29. Leijenaar RT, Carvalho S, Hoebers FJ, Aerts HJ, van Elmpt WJ, Huang SH, *et al.* External validation of a prognostic CT-based radiomic signature in oropharyngeal squamous cell carcinoma. Acta Oncol 2015;54:1423-9.
30. Kwan JY, Su J, Huang SH, Ghoraie LS, Xu W, Chan B, *et al.* Radiomic biomarkers to refine risk models for distant metastasis in HPV-related oropharyngeal carcinoma. Int J Radiat Oncol Biol Phys 2018;102:1107-16.
31. Bogowicz M, Leijenaar RT, Tanadini-Lang S, Riesterer O, Pruschy M, Studer G, *et al.* Post-radiochemotherapy PET radiomics in head and neck cancer – The influence of radiomics implementation on the reproducibility of local control tumor models. Radiother Oncol 2017;125:385-91.
32. Francis D. Trends in incidence of head and neck cancers in India. Eur J Cancer 2018;92:S23.
33. Shah SB, Sharma S, D'Cruz AK. Head and neck oncology: The Indian scenario. South Asian J Cancer 2016;5:104-5.
34. Murthy V, Calcuttawala A, Chadha K, d'Cruz A, Krishnamurthy A, Mallick I, *et al.* Human papillomavirus in head and neck cancer in India: Current status and consensus recommendations. South Asian J Cancer 2017;6:93-8.
35. Kissun D, Magennis P, Lowe D, Brown JS, Vaughan ED, Rogers SN. Timing and presentation of recurrent oral and oropharyngeal squamous cell carcinoma and awareness in the outpatient clinic. Br J Oral Maxillofac Surg 2006;44:371-6.