

Radiomics and Machine Learning for Radiotherapy in Head and Neck Cancers

Paul Giraud, Philippe Giraud, Anne Gasnier, Radouane El Ayachy, Sarah Kreps, Jean-Philippe Foy, Catherine Durdux, Florence Huguet, Anita Burgun, Jean-Emmanuel Bibault

▶ **To cite this version:**

Paul Giraud, Philippe Giraud, Anne Gasnier, Radouane El Ayachy, Sarah Kreps, et al.. Radiomics and Machine Learning for Radiotherapy in Head and Neck Cancers. Frontiers in Oncology, 2019, 9, pp.174. 10.3389/fonc.2019.00174 . hal-02094643

HAL Id: hal-02094643

https://hal.sorbonne-universite.fr/hal-02094643

Submitted on 9 Apr 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

frontiers in Oncology

REVIEW
published: 27 March 2019
doi: 10.3389/fonc.2019.00174

Radiomics and Machine Learning for Radiotherapy in Head and Neck Cancers

Paul Giraud[1,2], Philippe Giraud[1,2], Anne Gasnier[1,2], Radouane El Ayachy[1,2], Sarah Kreps[1,2], Jean-Philippe Foy[3,4], Catherine Durdux[1,2], Florence Huguet[5], Anita Burgun[2,6] and Jean-Emmanuel Bibault[1,2,6]*

[1] Radiation Oncology Department, Georges Pompidou European Hospital, Assistance Publique-Hôpitaux de Paris, Paris Descartes University, Paris Sorbonne Cité, Paris, France, [2] Cancer Research and Personalized Medicine-Integrated Cancer Research Center (SIRIC), Georges Pompidou European Hospital, Assistance Publique-Hôitaux de Paris, Paris Descartes University, Paris Sorbonne Cité, Paris, France, [3] Department of Oral and Maxillo-Facial Surgery, Sorbonne University, Pitié-Salpêtriére Hospital, Paris, France, [4] Univ Lyon, Université Claude Bernard Lyon 1, INSERM 1052, CNRS 5286, Centre Léon Bérard, Centre de Recherche en Cancérologie de Lyon, Lyon, France, [5] Department of Radiation Oncology, Tenon University Hospital, Hôpitaux Universitaires Est Parisien, Sorbonne University Medical Faculty, Paris, France, [6] INSERM UMR 1138 Team 22: Information Sciences to support Personalized Medicine, Paris Descartes University, Sorbonne Paris Cité, Paris, France

OPEN ACCESS

Edited by:
Christopher Schultz,
Medical College of Wisconsin,
United States

Reviewed by:
Di Dong,
Institute of Automation (CAS), China
Stephan Bodis,
Kantonsspital Aarau, Switzerland

***Correspondence:**
Paul Giraud
paul.giraud@aphp.fr

Specialty section:
This article was submitted to Radiation Oncology, a section of the journal Frontiers in Oncology

Received: 12 December 2018
Accepted: 28 February 2019
Published: 27 March 2019

Citation:
Giraud P, Giraud P, Gasnier A, El Ayachy R, Kreps S, Foy J-P, Durdux C, Huguet F, Burgun A and Bibault J-E (2019) Radiomics and Machine Learning for Radiotherapy in Head and Neck Cancers. Front. Oncol. 9:174. doi: 10.3389/fonc.2019.00174

Introduction: An increasing number of parameters can be considered when making decisions in oncology. Tumor characteristics can also be extracted from imaging through the use of radiomics and add to this wealth of clinical data. Machine learning can encompass these parameters and thus enhance clinical decision as well as radiotherapy workflow.

Methods: We performed a description of machine learning applications at each step of treatment by radiotherapy in head and neck cancers. We then performed a systematic review on radiomics and machine learning outcome prediction models in head and neck cancers.

Results: Machine Learning has several promising applications in treatment planning with automatic organ at risk delineation improvements and adaptative radiotherapy workflow automation. It may also provide new approaches for Normal Tissue Complication Probability models. Radiomics may provide additional data on tumors for improved machine learning powered predictive models, not only on survival, but also on risk of distant metastasis, in field recurrence, HPV status and extra nodal spread. However, most studies provide preliminary data requiring further validation.

Conclusion: Promising perspectives arise from machine learning applications and radiomics based models, yet further data are necessary for their implementation in daily care.

Keywords: radiomics, machine learning in head and neck cancer, predictive medicine, radiation oncology, treatment planning

INTRODUCTION

Machine Learning (ML) has promising applications in radiation oncology. This technique can process a high number of heterogeneous parameters and its recent development has led to a new paradigm in predictive medicine with hope for improved prognosis classification and toxicity prediction for a better treatment strategy. ML algorithms are all the more powerful that they include data from many different sources that can be clinical, biological, genomic and radiologic. Radiomics are actively being explored as prognostic and predictive tools, correlated with histologic and genomic characteristics in several cancer types (1–3). ML techniques may also enhance the radiation oncology workflow management thanks to automation of treatment planning steps such as automatic delineation or adaptive radiotherapy.

Radiomics

Innovation in medical imaging has led to a dramatic increase of resolution and imaging modalities (MRI, PET TDM, CT), imaging agents, reproducible protocols, and image analysis such as radiomics. Radiomics consists in extracting hundreds of quantitative features by an automated or semi-automated software. It relies on the hypothesis that mineable data can be extracted from medical images and provide additional information on gene protein and tumor phenotype and then used for patient care (4, 5).

Contrary to subjective evaluation of tumor characteristics such as necrosis or heterogeneity, radiomics use data characterization algorithms, automatically extracted from a delineated volume/region of interest (on a CT, PET, or MRI scan), to render a mineable feature space. They can be classified into the following categories (6).

- First order features describing the distribution of voxel intensities (including intensity features from wavelet decompositions of the original image).
- Shape features: related to the shape of the volume.
- Texture features: revealing intra-tumoral heterogeneity differences (including texture features from wavelet decompositions of the original image).

A high number of features is usually extracted and a selection needs to be made, generally by the following steps: each region/volume of interest is contoured by two independent physicians to assess feature stability: only reproducible values independent of delineation uncertainties are kept, using Intra Class Correlation (ICC). Then univariate analysis is performed to select features with a statistically significant correlation with the endpoint. Eventually multivariate analysis and feature ranking are performed. The best performing features are then combined into a predictive model. The performances of the resulting model is then assessed on an external validation dataset (**Figures 1**, **2**). Due to the high number of interdependant features generated, ML algorithms are well-suited to render best performing models. The rise of research in radiomics has relied on the recent accelerated development of ML techniques.

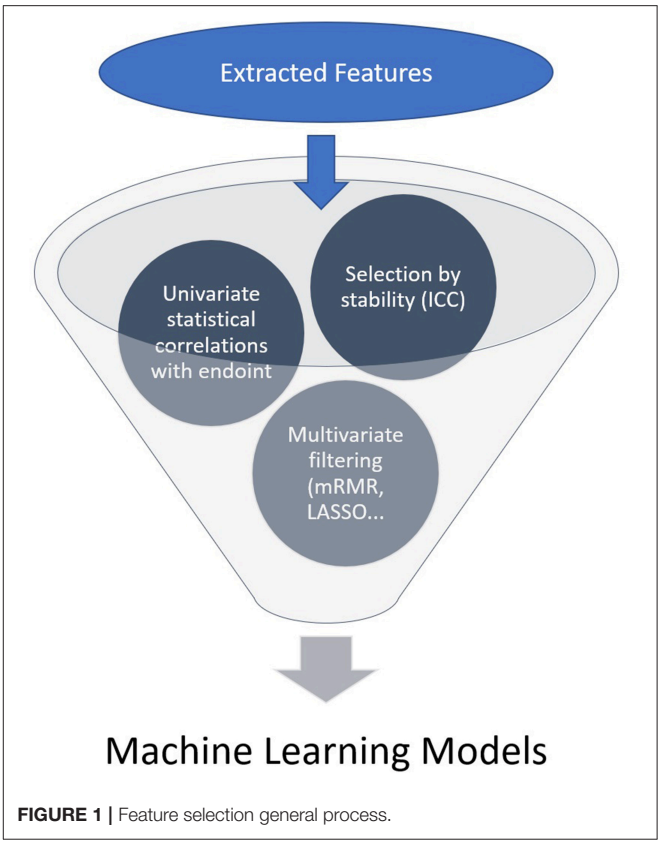


FIGURE 1 | Feature selection general process.

The reproducibility of such a model is also highly dependent of the imaging technique (7, 8). Movement may alter the model performances as well as the time lapse between contrast injection and acquisition, or data processing from acquisition to display (from RAW images to DICOM images), known to vary across manufacturers.

Several institutions have dedicated teams for radiomics research with an open source approach. For instance, the radiomics.io plateform (http://www.radiomics.io/) was developed by Harvard University in order to develop and maintain open source projects, provide go-to resource for radiomic applications and promote radiomics within the science community. It provides links to major radiomics original articles and features extraction tools such as PyRadiomics (9). The CASMI (Chinese Academy of Science, Medical Imaging) laboratory in Beijing has developed a website on which tools such as *3DMed* (feature extraction software) and *Radiomics Software* (automated segmentation and radiomics classifier software) are available for download (http://www.radiomics.

Abbreviations: ML, Machine Learning; ANN, Artificial Neural Network; DT, Decisional Tree; SVM, Support Vector Machine; BN, Bayesian Network; NTCP, Normal Tissue Complication Probability; RT, Radiotherapy; DSC, Dice-Sørensen Coefficient; OAR, Organ at Risk; IMRT, Intensity Modulated RadioTherapy; kV-CBCT, Kilo-Voltage Cone Beam Computed Tomography; MV-CBCT, Mega Voltage Cone Beam Computed Tomography; ART, Adaptive Radiation Therapy; HDIR, hybrid deformable image registration; TPS, Treatment Planning System; ROI, Region Of Interest; DVH, Dose Volume Histogram; RFC, Random Forest Classification; GTV, Gross Tumor Volume; ENE, extra nodal extension; FOV, Field Of View.

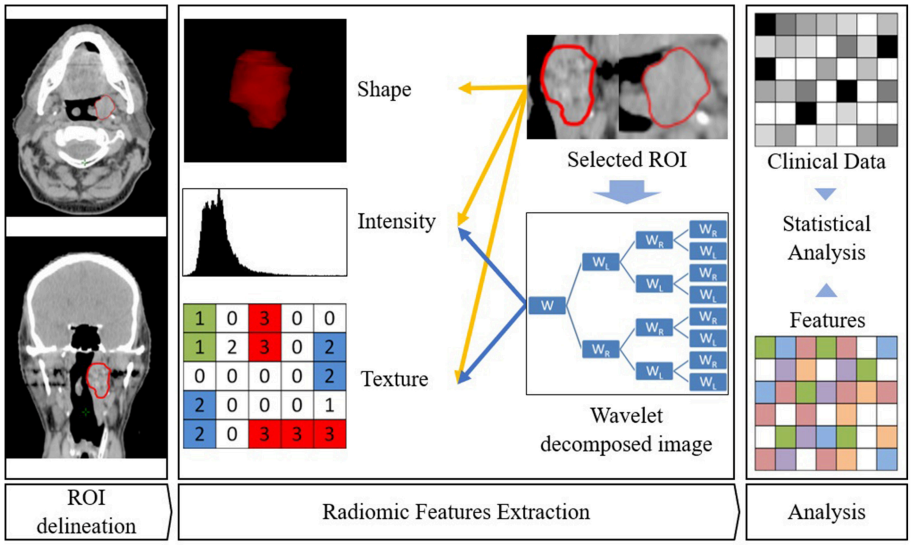


FIGURE 2 | Radiomics workflow. ROI is first delineated. Then features are extracted from this ROI. Eventually, association with clinical parameters or survival are sought.

net.cn). It also launched a cancer data sharing plateform. Other features extraction softwares have also been developed and made available to the scientific community such as IBEX (10), LifeX (11).

Machine Learning

ML technique is a branch of artificial intelligence in which an algorithm learns by inference from a data set (12). Its main objective is to produce a model capable of classification, prediction, and estimation of a situation from selected known data. As a result it may improve our decision-making process as it can encompass a higher number of parameters than humans (13). Parameters coming from clinical observations, biology, genomics and radiomics data may improve the clinician decision-making process (**Figure 3**).

The training phase is the process of finding dependencies in a system for a given dataset. Then, these dependencies are used to predict new outputs. ML methods are categorized into supervised and unsupervised learning. Supervised algorithms use labeled data (to map the data to the desired output classification) while unsupervised learning algorithms do not use labeled examples and provide a classification based on the patterns found in the dataset. Semi supervised learning methods combine supervised and unsupervised learning when dataset combines both labeled and unlabeled data.

The performance of a model is estimated on a validation sample, through sensitivity, specificity, accuracy and area under the curve (AUC) on the testing sample (**Figure 4**). If there is only one dataset, validation methods such as Holdout Method, Random Sampling, Cross-Validation, and Bootstrapping can be used (14).

Machine Learning methods include, but are not limited to:

- ANN (artificial neural network) can be used to classify patients. Its main disadvantage is its high level of abstraction which makes it difficult to determine what features were used by the neural network. This disadvantage is called the "black box" effect.
- DTs (Decisional Trees) renders a decisional tree to classify each input.
- Random forest classifier uses a vote from multiple decisional trees for a better classification compared to a single decisional tree.
- SVM (Support Vector Machine) technique identifies a hyperplane that separates data points in two classes. Its advantages are its generalizability and the possibility to obtain probabilistic outputs.
- BNs (Bayesian Networks) produce probability estimations that are useful for representing dependencies between input features through graphs.

Limitations of machine learning lie in data quality (missing, duplicated, noise, outliers) which is the corner stone of quality algorithms. Besides, a higher number of features (data type) implies a higher number of samples (15). Data sample should be 5–10 times more numerous than the number of features. As a result, data processing is necessary to reduce dimensionality and enhance the model robustness (16). However, for a given question such as prognosis, many signatures with different features are generated resulting in a lack of agreement on the features to take into account (17–20).

The ideal classification model should fit the training set and correctly classify each item in the test dataset. However, if misclassification error rate is low in the training set, and gets higher in the testing set, the model is overfitting and is not generalizable.

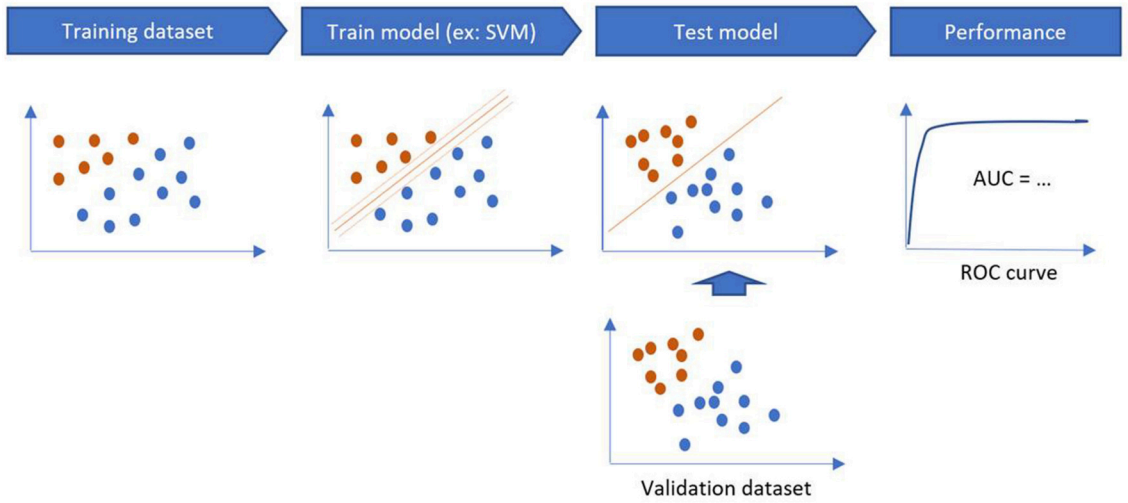


FIGURE 3 | Training and validation steps of a machine learning algorithm.

As a result, validation on external cohorts is critical to prove its generalizability. Many models (not necessarily developed with machine learning techniques) are shared on plateforms such as http://www.predictcancer.org. This open source initiative may ease external validation of models by other teams on other populations and promoted by radiomics websites previously introduced.

MATERIALS AND METHODS

Descriptive Approach of Machine Learning Application

We performed a description of applications of machine learning at each step of treatment of head and neck cancer in radiotherapy.

Review on Radiomics and Machine Learning Survival Models

We then focused on radiomics with a systematic review in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA).

We conducted a comprehensive search of PubMed for relevant peer-reviewed publications from January 2012 to January 2019. A comprehensive list of MeSH terms and keywords was used to query Pubmed ("head and neck cancer," "radiomic," "signature," "survival," "machine learning"). The search strategy also included screening of reference lists of relevant publications ("snowball" search technique). Publications of models based on radiomics or machine learning to predict survival or distant metastasis in head and neck cancer after radiotherapy was performed.

RADIATION ONCOLOGY TREATMENT PLANNING

Radiotherapy treatment planning encompasses time consuming tasks such as delineation, dosimetric planification, and adaptive radiotherapy. Increasing automation of these tasks is a promising prospective. It may shorten and improve reproducibility of contouring that limits the implementation of adaptive radiation therapy and Normal Tissue Complication Probability (NTCP) modeling.

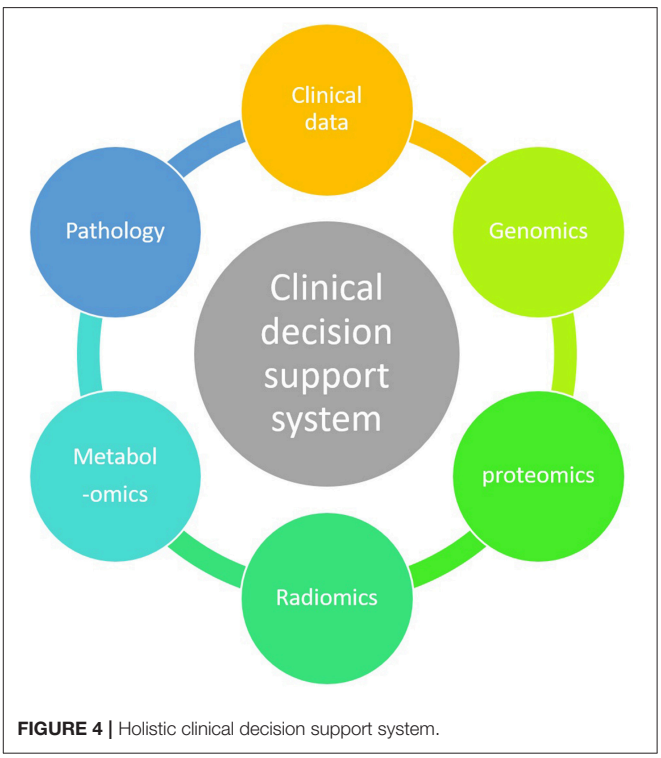


FIGURE 4 | Holistic clinical decision support system.

Automatic Organ at Risk (OAR) Delineation

The delineation process is manual and time-consuming and suffers from inter and intra practitioner variability. Automatic

OAR delineation may help for contouring homogenization and ease the workflow, a key factor for implementing adaptive radiotherapy. The initial approach for automatic delineation was mainly atlas based, using hybrid registration to deform structures from an atlas and map it to the patient anatomy.

Yang et al. (21) created an algorithm for automatic segmentation of the post treatment parotid glands. Fifteen patients were prospectively included. They had head and neck cancer treated by radiation therapy with at least 1 year of follow-up. Parotid glands were manually contoured on a pre-treatment MRI. This volume would then be used as an atlas for each patient. MRIs were performed at 3, 6, and 12 months after radiation therapy. The aim was to have the parotids glands automatically segmented on follow-up MRI to assess their evolution. The first step was an atlas-based registration: first, a hybrid deformable image registration was used to map the pre-RT MRI to the post-RT MRI. The resulting transformation was applied to delineated parotid volume. Multiple features were extracted for input aside with the transformed contour for the support vector machine training in order to detect horizontal, vertical and diagonal edges and strength of edges.

The trained SVM could detect edges on the follow-up MRI to render a parotid contour, then intersected with the transformed parotid volume + 10 mm. Smoothing and 3D morphology operations were then applied to render a volume without disconnection, holes or irregularities.

The ground truth was manually contoured on a follow up MRI. A volumetric Dice-Sørensen Coefficient (DSC) was performed to assess the model performance. There was no significant difference of volume between the automatically and manually contoured parotid glands on follow-up MRI with a volumetric Dice-Sørensen Coefficient of more than 90%.

However, such an atlas based method provided little improvement to the workflow since the result usually needs correction (22–31). Using artificial neural networks is a new approach: Nikolov et al. (32) recently created a convolutional neural network using a 3D U-Net architecture model (33) to automatically contour 21 OAR. The validation process had to account for human variability: each case was segmented by an experienced radiologist arbitrated by a second delineation; their contour were then compared with two further radiographers arbitrated by experienced radiation oncologists as the ground truth. The performance was assessed by a surface DSC as it provides a quantitative agreement between the two surfaces instead of the two volumes.

This assessment method is more clinically relevant since it better represents manual correction time for every misplacement. Delineated OARs were the brainstem, spinal canal, spinal cord, left and right cochleas, lacrimal glands, lenses, lungs, optic nerves, orbits, submandibular, and parotid glands. Each OAR surface DSC was similar to humans in all previously described OAR except for the right lens and the brainstem (deviation of more than 5%). Right lens uncertainties may be correlated with CT quality as borders are easier to see on a higher definition CT. Interobserver variability among experienced oncologists (for instance parotid delineation) was a limitation for learning since the ground truth was not consensual.

Surface DSC is a metric introduced by Nikolov et al. to take into account misplacement even by a small offset that may be time consuming to correct compared to volumetric DSC. This convolutional neural network approach improved the usual performance of automatic delineation model. Clinical acceptability and time saving estimations need to be assessed to value the model's clinical implementation. Integrating other imaging modalities (MRI, PET, CT) co registration to better contour certain OAR has to be assessed.

Adaptive Radiotherapy

Intensity Modulated RadioTherapy (IMRT) is the standard of care for Head and Neck radiotherapy. Daily repositioning by kV-CBCT (Kilo-Voltage Cone Beam Computed Tomography) or MV-CBCT (Mega Voltage Cone Beam Computed Tomography) through rigid registration to detect and correct set up errors (34) can be performed. Adaptive Radiation Therapy (ART) is the adaptation of target volumes and critical OAR volumes to tumor shrinkage and patient emaciation during treatment (35–37). Benefits and feasibility of ART has been assessed by exploratory studies (38–40) ART is not routinely used, as a consequence of limited resources, its time consuming character, and unreliable automatic contouring.

Thus, the main obstacle to test ART on a higher number of patients is the supplementary workload even with hybrid deformable image registration (HDIR) (35). Guidi et al. created (41) and tested (42) a support vector machine tool for adaptive tomotherapy treatment in head and neck cancer to completely automate this process in order to raise an alert when the patient would need a new dosimetric scan to adapt the volumes. Between 2013 and 2014, 40 patients diagnosed with a head and neck cancer treated by radiation therapy with 66 Gy on high-risk volume and 54 Gy on low-risk volume were included. Raystation® (v 4.5.1.14 RaySearch Laboratories AB, Stockholm Sweden) TPS was used for HDIR. First, a rigid registration was performed and ROIs copied on the MVCT. Then, HDIR was performed and the resulting mesh grid—with a voxel to voxel map of the deformed vector field—was eventually applied to new automatically contoured ROIs. Then dose deformation was carried out using the deform dose tool of the TPS.

The algorithm classified data that included ROIs and several dose metrics on a weekly basis.

Eventually, the SVM model rendered a graph and a summary of the output. Unsupervised learning identified four categories:

- Predictive: the data points are within the defined threshold. Re-planning is not needed;
- Adaptive: Points are outside the defined threshold. The patient needs a replanification;
- Errors: points that are outside thresholds because of a software bias;
- Warning: points that are outside thresholds because of emaciation, organ motion, or incorrect set up;

After a learning phase on 27 patients, the algorithm was tested on 13 patients. In the first 3 weeks, no replanification was needed for 84% of the patients. Then a reverse trend became significant at

the beginning of the 4th week with 77% of replanification needed during the last 2 weeks of treatment.

Guidi et al. then performed a clinical validation retrospective cohort of 90 patients (42) from four radiation oncology centers equipped with various accelerators (Tomotherapy or Linac + CBCT) to assess the performance of the previously described algorithm. Two physicians suggested an ideal re-planning time—point if re-planning had to be performed. Decision was based on parotid gland and neck volume shrinkage in comparison with the initial dosimetric CT scan.

Similarly to Yang et al. (21) after 6 weeks of therapy, parotid gland volume decreased by 23.7 ± 8.8%. As shown on their previous work described above (41), there was no re-planning needed during the first 3 weeks (for 86.7% of cases). From the 4th week, 55.2% of patients would need a re-planning. 11.8% were affected by biases and 5.9% would generate a warning.

Giudi et al. provide a machine learning powered, automated tool to discriminate treatments in need of replanification. The clinical validation was focused on the time point of re-planning reproducibility compared with physician decision.

Many adaptive ART approaches are being investigated and questioned (40). Dose recalculation models are based on patient cohorts and are thus limited to data availability. It should be enhanced with the rise of data mining and further validated. The machine learning approach proposed by Guidi et al. opens the prospect of triggering a clinical decision in an automated process, leaving only equivocal cases to the radiation oncologist. However, this would require an automatic workflow handling with a powerful TPS computing power.

NORMAL TISSUE COMPLICATION PROBABILITY ASSESSMENT

Predictive Model Limitations

NTCP models are usually based on dose–volume histograms (DVHs), which are not ideal representations of 3D doses and assume that all the regions of an OAR have an equal functional importance, thus discarding organ specific spatial information. They do not take into account fraction dose variations, nor anatomical variations of the OAR and its dosimetric consequences during treatment (43). Models taking into account more data such as 3D dose distribution in OARs, dependencies between the dose delivered at other OARs may enhance toxicity prediction. Models using machine learning might be well-suited since the bigger amount of interdependent data to take into account.

Xerostomia

Xerostomia is one of the most limiting toxicity during radiation therapy for head and neck cancer due to its consequences on weight loss, tooth decay, speaking difficulties, dysphagia (21, 44). A better prediction of parotid gland function would enhance its protection and the quality of life after RT. Parotid volume diminishes with radiotherapy (21, 41, 45). The volume reduction and the migration toward high dose region can generate a total absorbed dose significantly higher than the planned one. Consequently, a relation between parotid gland volume reduction and xerostomia was found (46). As previously described, Yang et al. (21) generated an atlas based model for parotid delineation. The volume tended to decrease with time after radiotherapy consistently with previous finding (44). New NTCP models predicting xerostomia using volumetric change and other parameters could be tested without the limiting burden of time-consuming contouring. Further longitudinal studies using this algorithm may investigate parotid gland volume reduction as a predictive factor of xerostomia.

Zhang et al. published a model for a plan related clinical complication probability in a multiplan IMRT framework (47). They aimed to predict a given OAR complication as a function of all OAR dose volume constraint settings without explicit additional computation. Using machine learning, such a prediction may be possible. One hundred and twenty-five plans for one head and neck by varying Dose Volume (DV) constraints on right or left parotid, or cord were created

The sample was randomly partitioned between training and testing samples with a 50-time cross validation. Decision trees and support vector machine were explored. SVM yielded superior results in predicting saliva flow rate and decision trees rectal bleeding.

The ground truth on salivary flow estimated by an Equivalent Uniform Dose based model developed by Chao et al. (48) trained on retrospective studies (48–50). Compared to this model, plan related post RT salivary flow rate prediction using SVM had a 0.42% (0.41–0.43) error.

This multiplan approach to display the risk of an OAR adverse event regarding dosimetry constraints on other OAR is global and may help physicians in their prescription. The ground truth relies on relatively weak and somewhat heterogenous data (retrospective cohorts) that may bias the prediction tool, but on the other hand, it easily integrates these data in the clinical decision. This approach would now require a clinical validation.

Mucositis

Mucositis is commonly seen in head and neck radiation therapy treatments, leading to anorexia, pain, dysphagia, weight loss (51), reduced quality of life (52) and eventually missed fractions (53), higher treatment duration compromising tumor control (54). Mucositis limits dose escalation or hyper fractionation designed for better oncology outcomes (55).

Many NTCP models have been developed for clinical decision support, treatment modality selection, and treatment plan optimization (56–59). However, none can predict mucositis severity and guide clinical decision. This may be due to their DVH-based nature, leading to oversimplification of the dose distribution [since it was previously shown that xerostomia was impacted by spatial distribution of the dose (60)]. Oral cavity contains keratinised and non-keratinised mucosa, thus DVH based decision may be biased. Dean et al. conducted a study to generate a severe acute mucositis NTCP and provide with the generated model clinical decision guidance, based on dose constraints (61).

Three hundred and fifty-one patients enrolled in six prospective trials were included. Patients had been treated for

various head and neck cancers and had DICOM RT files available. Mucositis was graded with the common terminology criteria for adverse events CTCAE v2 or 3 which are almost equivalent. Mucositis was dichotomized between severe (grade 3 and more) and non-severe (grade 2 or less). Induction chemotherapy, concurrent chemotherapy regimen, definitive vs. post-operative radiotherapy, primary disease site, age, sex and relative cumulative dose–volume histogram in 20 cGy intervals from 20 to 260 cGy per fraction were co variates of the models. Statistical analysis used for ML methods followed the principles described by Kang et al. (15) for model generation, and on the TRIPOD guidelines (Transparent Reporting of multivariable Prediction model for Individual Prognosis or Diagnosis) (62). Three different types of classifiers were generated: penalized logistic regression (PLR) (63), support vector classification (SVC) (64) and random forest classification (RFC) (65). They had approximately the same discriminative abilities. The addition of the dose distribution description did not improve the discriminative abilities. Therefore, the standard models (without the spatial dosimetric data) were favored. The standard RFC model had better calibration, probability estimates and overall performance than PLR and was the easiest to understand. As a result, the RFC model was preferred for prediction of severe mucositis for individual patient. It is available online at https://github.com/jamiedean/oral-mucositis-model.

The volume which received 220 cGy or more by fraction (V220) was the most important feature in both RFC models. The higher fractional dose volume (v240 and v260) were close to 0% and therefore did not correlate with mucositis. Age was the clinical covariate of highest importance but this may be artifactual due to the large number of possible values compared with the other clinical co variates. Besides, on univariate logistic regression, it was not significantly associated with severe mucositis.

Another model was published (66) with a similar findings on dose response (volume of oral mucosa receiving more than 2 Gy per daily fraction was the most strongly associated feature). Concurrent chemotherapy was also correlated contrary to the Dean et al. model, perhaps since the V220 cGy volume was positively correlated with concurrent chemotherapy. The Dean et al. model was based on delineation standard that does not accurately represent the mucosal surface and may explain the lack of improvement with spatial dosimetric data. Dean et al. recently created and validated a novel automated contouring technique for oral mucosa that would be more reproducible and representative of the mucosal surface (67, 68). Data on tobacco and alcohol consumption were not available for patients in two out of six trials.

TUMOR CONTROL PROBABILITY ASSESSMENT

Improving the tumor control probability assessment before treatment is a promising way to adapt the treatment strategy. Head and Neck tumor treatment strategy mainly depend on the TNM stage. Radiomics are correlated with tumor characteristics, which may provide improved prognosis classification.

Radiomics Revealing Tumor Characteristics

HPV Status

Yu et al. (69) developed a model to assess HPV status with a radiomic signature: they extracted 1,683 features with IBEX (10) from head and neck GTV T or N (or both, in this case the most extreme value was used). Feature selection was carried out in a stepwise manner: (1) inhomogeneous features between training and validation cohorts were discarded; (2) only features which could discriminate HPV + and – were kept; (3) only biserial absolute correlation > 0.3 between each radiomic feature and HPV status were kept (4) only the top 10 features regarding their mean AUC (assessed by various statistical models) were kept (5) final features were selected by forward selection (one by one add to the model until the model AUC ceases to increase).

The General Linear Model (GLM) was used as it provided better AUC compared with various other models (such as random forest, SVM, decision trees, deep learning…). MeanBreadth and Spherical Disproportion were the most important features. MeanBreadth measures width of the ROI, thus it is correlated to tumor size. SphericalDisproportion represents the complication of the surface compared to a sphere. This may imply that HPV + tumors are smaller and less complex compared to HPV—tumors. The model performance was tested on two dataset coming from a public and a private hospital. The model performed an AUC of 0.86667 on public leaderboard, and 0.91549 on the private one which is a good indicator of generalizability. Clinical data such as grade or TNM stage although available, were not used as authors wanted to assess the image discriminatory power, and the discretization induced by these variables could add uncertainty to the final model.

Extra Nodal Extension

Nodal metastasis and tumor extra nodal extension (ENE) are key decision factors for cancer management. However, the performance for ENE detection by imaging measured by AUC range between 0.65 and 0.69 (70). From CT scans of 270 patients from 11 centers, Kann et al. (71) developed a convolutional neural network (CNN) trained on 2875 lymph nodes delineated and labeled with histologic features. After cross validation on 124 samples, the CNN was then tested on 131 samples. To avoid overfitting, data augmentation by random rotations and flips was used. Three different networks were used (Boxnet, SmallNet, DualNet). Dualnet was selected for further testing as it yielded the best AUC. On independent test set, the DualNet neural network demonstrated an AUC of 0.91 (95%CI: 0.85–0.97) with a negative prediction value (NPV) of 0.95 for ENE status, and a AUC of 0.91 with a NPV of 0.82 for nodal metastasis. Using a convolutional neural network directly on images, without prior features extraction, on raw DICOM is a robust approach as it frees from the variability of radiomics features and image pre-processing. Furthermore, they performed a feature extraction with RF analyse with lower

performance compared to convolutional neural network. Only nodes that could definitively be correlated with histologic reports were included, which may lead to a selection bias. CT scan parameters heterogeneity was minimized by normalization and scan resampling.

Machine Learning on Clinical Factors for Prognosis Assessment

Bryce et al. (72) developed an artificial neural network on clinical factors and compared it to a logistic regression approach to create a predictive model of the 2-year recurrence free survival. They analyzed data from a phase III trial (73) that included patients with locally advanced head and neck squamous cell carcinoma (≥T3 or N3 or T2N0 base of tongue, hypopharyngeal, or pyriform sinus carcinomas) undergoing hyperfractionated accelerated radiation therapy with or without concurrent cisplatin and 5FU. Of the 116 patients of the study, 95 were included. Fourteen clinical variable potentially associated with 2 year PFS (Progression Free Survival) were available. Initial univariate analysis followed by stepwise multivariate analysis were first performed to select a subset of promising variables. This subset was pruned through stepwise elimination from an ANN model. The neural network used was a three layers feed forward back propagation neural network (74). For logistic regression models, clinical variables were first selected by stepwise selection and backward elimination with an assessment of the model performance to confirm that the selected variables were the best for 2-year survival prediction. The best logistic regression model used nodal stage, tumor size and race with an AUC of 0.67 ± 0.5. At 70% sensitivity, LR performed a 54% specificity while artificial neural network performed a higher AUC of 0.78 ± 0.05 ($p = 0.07$) with a specificity of 72% for a 70% sensitivity. As a result, ANN performed a better model of the 2-year survival. Although its performance remains low for a clinical use, it was a sensible improvement in prediction compared to TNM alone or logistic regression model within the limits of the available data (≥T3 or N3 or T2N0 base of tongue, hypopharyngeal or pyriform sinus carcinomas). As a result, the 2-year follow-up was too short to reveal the effects of chemotherapy. Due to this non-linear and non-parametric superior ability to model complex patterns, ANN techniques provide a new approach for predictive medicine, but rely on the quality and quantity of data available and its performance is always limited to characteristics of the patients in the training set.

Radiomic Signature Added Value

Radiomic features can predict tumor characteristics linked to survival. But do they provide additional information for a better prognosis classification? Ou et al. (75) developed a radiomic signature to estimate overall survival and assess its incremental value to HPV status. One hundred and twenty patients with stage III—IVb (TNM 2010) were included from 2006 to 2010 in a former monocentric study cohort (76). They received either cetuximab (BRT group) or cisplatin (CRT group) concurrently with a 3 dose levels radiotherapy (70, 60, 50–54 Gy). Oncoradiomics^{TM} was used to extract 544 features, narrowed to 24 statistically significant features by logistic regression, from which a radiomic signature score was generated. On the whole population, the signature score had prediction capacity of 5-year survival with AUC = 0.67 95% CI [0.58–0.76]. The fraction of patient whom HPV status was unknown in BRT or CRT groups was not detailed. P16 status also predicted for OS (Overall Survival). Patients with low score had significantly better OS and PFS. When combining p16 and radiomics signature the prognostic performance improved (AUC = 0.78, $p = 0.01$). This preliminary study showed an incremental added value of a radiomic signature score to HPV status. It would require external validation on a more homogenous population as treatment standards evolved between 2006 and 2012.

Radiomic signature reveals a tumor phenotype that may not be organ specific. For instance, Aerts et al. (6) generated a four feature radiomic signature on a retrospective cohort of 422 patients diagnosed with Non-Small Cell Lung Cancer (NSCLC) treated with curative intent. The signature's validation was performed on a 225 NSCLC patients cohort, two Head and Neck Squamous Cell Carcinoma (HNSCC) patients cohort, and a 89 patients HNSCC cohort with genomic information. The four feature radiomics signature was validated using the concordance index [c-index or CI, a generalization of the ROC AUC (77)]. On Lung2 data, performance was good with a CI of 0.65, $p = 2.91.10^{-9}$ (Wilcoxon test). On H&N1 and H&N2, performance was high with respective CI of 0.69, $P = 7.99 \times 10^{-7}$ (Wilcoxon test), and CI = 0.69, $P = 3.53 \times 10^{-6}$ (Wilcoxon test).

Compared to volume, the radiomic performed significantly better. When combined to TNM classification, the performance was significantly better than TNM alone suggesting complementary information for prognosis. These performances were improved both in the HNSCC and the NSCLC cohorts, in all the treatment groups (radiation or concurrent chemo radiation). This signature was not correlated with HPV status and did preserve its performance in the HPV—group.

Aerts et al. showed the predictive value of features selected for their stability and reproducibility which indicates the power of integrating independent datasets for feature selection and model building. Among those features, keeping only the top four performing features regarding prognosis to perform a single validation on 545 patients allowed a robust statistical analyse and avoided overfitting. The most dominant features of the signature quantified intra-tumor heterogeneity. Intra-tumoral heterogeneity reveals the presence of multiple colonies with different gene expression that cannot be encompassed by biopsies and therefore is an obstacle to personalized medicine and a potential prognosis marker. This results could be supported by associations with gene expression found by Zhu et al. on the TCGA and TCIA cohorts (78). The better prognosis prediction performance in head and neck cancer may be due to the absence of breathing movement but underlines the non-specificity of heterogeneity of this signature in cancer prognosis.

The MD Anderson Cancer Center head and neck quantitative imaging working group (79) has developed a radiomic signature on 465 patients with oropharyngeal cancer to assess the recurrence probability after IMRT. Patients were split into training, tuning and validation sets. For each GTVp, 134 radiomic features were extracted using IBEX and narrowed

to two features with decision tree modeling: Intensity Direct Local Range Max (LRM) and Neighbor Intensity Difference 2.5 Complexity with cut-off values corresponding to 1,616 and 457,808, respectively. Patients with both values were below the aforementioned cut offs, had a more favorable local tumor control with 94% of tumor control rate at 5 years in the training set compared to 62 to 80% depending if one or two values were above cut-offs. Compared to clinical factors such as HPV, smoking status or age, the radiomic signature performed better. Limitations come from CT heterogeneity and some missing clinical data (HPV status, tabaco consumption status).

Nasopharyngeal carcinomas 5-year overall survival is around 50% (80), but drops when in advanced stage. Prediction of outcome may help to make decision regarding treatment. Zhang et al. (81) developed a predictive model for 3-year PFS using MRI-based radiomics in nasopharyngeal carcinomas. One hundred and eighteen patients were included. Nine hundred and seventy features were extracted using Matlab. A radiomics signature from 8 features (5 from T1 weighted and 3 from T2 weighted images) yielded a C index of 0.737 (95% CI: 0.549–0.924). This performance outperformed classic TNM staging system (C-index: 0.514), but both combined yielded a better performance (C-index: 0.761; 95% CI, 0.664–0.858) with a good calibration.

Being able to predict not only the risk of recurrence but also whether it would be locoregional or a distant metastasis (DM) may enhance treatment decisions. As a result, Li et al. (82) explored the predictive power of radiomics for the type of recurrence (in field, out of field, marginal). After selection of the influential subset of radiomic features, Kruskal Wallis test and ROC analysis were employed for each feature to assess its capability on in field recurrence prediction. After further selection, several machine learning models were trained (Artificial Neural Network, K Nearest Neighbor and Support Vector Machine). Three hundred and six patients were included with a median follow up of 26.5 months. Eighteen in field, one marginal and one out of field recurrences occurred. Patients with in field recurrence could be differentiated by 8 features. The best performing model was artificial neural network (AUC: 0.812).

Zhang et al. (83) developed a MRI based model for distant metastasis (DM) prediction in nasopharyngeal carcinoma. DM is a common cause of treatment failure in nasopharyngeal carcinoma. One hundred and seventy-six patients from a retrospective multicentric cohort were included, 123 in the training cohort and 53 in the validation cohort. Pre-treatment MRIs, clinical and biological data were gathered. The primary endpoint was the time to primary MRI to DM or censoring. The follow up ranged from 36 to 60 months after the primary MRI. All patients received concurrent platinum based chemoradiotherapy. MRI were performed with various protocols, thus intensity images were normalized using PyRadiomics plateform. Two senior radiologists segmented images blinded from each other to assess interobserver reproducibility. Two thousand eight hundred and three imaging features were extracted using PyRadiomics platform. Feature selection included many steps: First interclass correlation coefficient (ICC) was calculated to assess the effect of segmentation variability, only features with an ICC above 0.75 were kept. Then, univariate analysis for survival using Chi^2 or Mann–Whitney *U*-tests were performed for each features. Only statistically significant parameters were kept ($p < 0.05$). Then minimum redundancy maximum relevance model was applied to keep the top 10% features. Eventually, the most significant features with highest area under the curve were selected using LASSO (Least Absolute Shrinkage and Selection operator) algorithm. A logistic regression based model to assess the DM risk was then trained. Kaplan Meier survival curves of the two groups were compared using log rank tests. The predictive ability of the model was assessed with AUC. A Distant Metastasis MRI bases Model (DMMM) was created from 7 selected radiomics features with an AUC of 0.827. In the training cohort and 0.792 in the validation cohort. Patients with a high risk of DM had a 5-year overall survival of 12% vs. 26% in the low risk group. The DM related feature with the maximum significance was "CET1-w_wavelet.HLL_ GLCM_Correlation." It reflects the intra tumor heterogeneity from the gray scale extension perspective. The worse performance in the validation cohort suggested overfitting and further training on larger multicentric dataset should be considered. Still, being able to predict the risk of distant metastasis, or in field recurrence are promising prospective for radiomics support for clinical decisions.

Kwan et al. (84) performed a study with a similar goal on HPV related oropharyngeal carcinoma. Three hundred patients with HPV related oropharyngeal carcinoma were included. The 4 features radiomics features form Aerts et al. (6) previously described study were used. The best performing model was a combination of the radiomic signature with clinical data. In this cohort, the signature had better performance for prognostication of DM than of 5 years overall survival while it was initially designed for overall survival. The radiomics signature stratified patients for DM risks especially for cohorts with greater risks such as stage III patients.

DISCUSSION

Radiomics is a new way to mine data from imaging. It may provide specific information on tumoral heterogeneity, HPV status, lymphocytic infiltration and consequently, prognostic signatures.

For instance, as shown with Aerts et al. (6) gray level non-uniformity revealing intra tumoral heterogeneity was found as a bad prognosis factor which is congruent with literature and may reveal multiple clonal populations with a variable response pattern to treatments. Radiomics provide whole tumor information compared to biopsies and thus provide additional information that may become useful for treatment strategy.

As shown by Berenguer et al. (7), many radiomics features may be redundant or not reproducible. Using standardized imaging acquisition with homogenized contrast enhancement delay before acquisition, may enhance the quality of the data. Still, if all CT parameters were fixed except FOV, tube voltage, and milliamperage, then the information provided by radiomics could be summarized in only ten radiomics factors. On the other hand, too restrictive protocols may decrease the applicability

TABLE 1 | Overview of radiomics based outcome prediction models.

Study	Features	Extraction software (ES) and statistical method (SM)	Outcome prediction (OP) and performance (P)	Number of patients	Tumor characteristics
Yu et al. (64)	MeanBreadth Spherical disproportion	ES: IBEX SM: GLM	HPV status AUC: 0.86667 and 0.91549	315 pts	Oropharyngeal cancers
Ou et al. (71)	24 features signature	ES: Oncoradiomics SM: Logistical regression	5 y survival P: AUC = 0.67 CI (0.58–0.76)	120 pts	stage III – IVb Head and Neck cancer from (71)
Aerts et al. (4)	Statistics Energy Shape compactness 2 Grey level non uniformity Run length non-uniformity	ES: IBEX SM: Logistic regression	Overall Survival C-index: 0.69	545 pts	Lung and head and neck cancers
Anderson (79)	Intensity direct local range max Neighbour intensity difference 2.5 complexity	ES: IBEX SM: Decision Tree model	5 y Tumour control classifier (3 groups)	465 pts	Oropharyngeal cancers
Kann et al. (67)	No prior extraction for the selected model	ES: PyRadiomics SM: Random Forest and CNN	Extra nodal extension AUC = 0.91 (95%CI:0.85–0.97). NPV: 0.95	270 pts	Nodal invasion in resected head and neck cancer
Zhang et al. (81)	8 features signature	ES: Matlab: MRI based features. SM: LASSO and Rad-Score	3 y PFS C-index: 0.737 (95% CI: 0.549–0.924)	118 pts	Nasopharyngeal carcinomas
Li et al. (82)	8 features signature	ES: PyRadiomics on SPAIR T2W MRI SM: ANN	In field recurrence Accuracy: 0.812	306 pts	Nasopharyngeal carcinomas
Zhang et al. (83)	7 features signature	ES: PyRadiomics MRI based SM: Logistic regression	Distant Metastatic MRI based Model AUC : 0.827	176 pts	Nasopharyngeal carcinomas

GLM, General Linear model; CNN, Convolutional Neural Network; LASSO, Least Absolute Shrinkage and Selection operator; SPAIR T2W, spectral attenuated inversion-recovery T2-weighted; ANN, Artificial Neural Netwo.

to routine clinical care, all the more that with a given delay to acquisition, the actual enhancement varies depending on subjects. Another approach would be to gather CT scans with varying parameters to rule out noise in the dataset [as done by Aerts et al. (6), presented previously]. The extracted features also depend on the extraction software, patient motion and image treatment. This may be one of the reasons for variability of the generated signatures between each study (**Table 1**). This creates a real challenge for its implementation in daily clinical care, the ultimate goal of radiomics.

Machine learning encompasses many statistical tools suited for complex patterns detection with inter dependent and non-linear parameters. Machine learning models provide a promising enhancement of predictive model performances, suited for the increasing number of parameters available for analysis with the rise of omics. However, one of its main limitations is overfitting, leading to a lack of generalizability to other samples. To avoid overfitting, the number of patients should be five to ten times higher than the number of features (85), which requires a thorough pre-selection of features to keep the number of patients to include reasonable.

Models are only applicable on the range of data used for its generation and dependant on primary data quality. Ground truth definition is fundamental for successful model learning. It is usually defined by experienced practitioner's data labeling. In non-consensual situation, such as delineation, learning may be impaired by inter-practitioners' non-reproducibility. In this case, an approach based on comparing the model's uncertainty with the expert's uncertainty seems valid, as long as a clinically relevant metric is used. Last, each problem has its variables, and no type of machine learning algorithm can always prevail. For instance, logistic regression was better for the study of Ou et al. (75) while ANN was better suited for the clinical prediction of survival by Bryce et al. (72), which underlines the model benchmarking approach, where several models should be tested, and the best performing one kept.

CONCLUSION

Potential applications of machine learning and radiomics methods in the field of head and neck cancer have been explored in several studies. Treatment planning tasks such as OAR delineation or dose recalculation are time consuming and could eventually be automated, opening the possibility for adaptive radiotherapy. Machine learning and radiomics can provide better modeling tools both for adverse events and survival for a step toward personalized and predictive medicine. While these studies provided interesting results, none of them are actually being used in the daily workflow of radiation therapy departments. Before we can reach this goal, they must be thoroughly assessed in prospective, multicentric trials to prove their actual benefit. Collaborating groups will have an important role in the design and conduct of these important studies. The use of interoperable standards and homogeneous treatment planning methods will also be needed before such study can be performed.

AUTHOR CONTRIBUTIONS

PaG wrote the manuscript. J-EB helped for article selection and manuscript redaction. J-EB, AB, and REA helped for technical review of the manuscript for machine learning and radiomics aspects. J-PF performed a technical review of the manuscript on radiomics and clinical aspects. AG performed a technical review of the manuscript on machine learning aspects. SK, FH, CD, and PhG performed a review of the manuscript on the clinical aspects.

REFERENCES

1. Segal E, Sirlin CB, Ooi C, Adler AS, Gollub J, Chen X, et al. Decoding global gene expression programs in liver cancer by non-invasive imaging. *Nat Biotechnol.* (2007) 25:675–80. doi: 10.1038/nbt1306
2. Diehn M, Nardini C, Wang DS, McGovern S, Jayaraman M, Liang Y, et al. Identification of noninvasive imaging surrogates for brain tumor gene-expression modules. *Proc Natl Acad Sci USA.* (2008) 105:5213–8. doi: 10.1073/pnas.0801279105
3. Nair VS, Gevaert O, Davidzon G, Napel S, Graves EE, Hoang CD, et al. Prognostic PET 18F-FDG uptake imaging features are associated with major oncogenomic alterations in patients with resected non-small cell lung cancer. *Cancer Res.* (2012) 72:3725–34. doi: 10.1158/0008-5472.CAN-11-3943
4. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RGPM, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer.* (2012) 48:441–6. doi: 10.1016/j.ejca.2011.11.036
5. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol.* (2017) 14:749–62. doi: 10.1038/nrclinonc.2017.141
6. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun.* (2014) 5:4006. doi: 10.1038/ncomms5006
7. Berenguer R, Pastor-Juan M del R, Canales-Vázquez J, Castro-García M, Villas MV, Mansilla Legorburo F, et al. Radiomics of CT features may be nonreproducible and redundant: influence of ct acquisition parameters. *Radiology.* (2018) 288:407–15. doi: 10.1148/radiol.2018172361
8. Ger RB, Zhou S, Chi P-CM, Lee HJ, Layman RR, Jones AK, et al. Comprehensive Investigation on Controlling for CT imaging variabilities in radiomics studies. *Sci Rep.* (2018) 8:13047. doi: 10.1038/s41598-018-31509-z
9. Griethuysen JJM van, 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. doi: 10.1158/0008-5472.CAN-17-0339
10. Zhang L, Fried DV, Fave XJ, Hunter LA, Yang J, Court LE. IBEX: an open infrastructure software platform to facilitate collaborative work in radiomics. *Med Phys.* (2015) 42:1341–53. doi: 10.1118/1.4908210
11. Nioche C, Orlhac F, Boughdad S, Reuzé S, Goya-Outi J, Robert C, et al. LIFEx: A Freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity. *Cancer Res.* (2018) 78:4786–9. doi: 10.1158/0008-5472.CAN-18-0125
12. Mitchell, T. M. (2006). The Discipline of Machine Learning. Pittsburg, CA: School of Computer Science, Carnegie Mellon University. Available online at: http://www.cs.cmu.edu/~tom/pubs/MachineLearning.pdf
13. Abernethy AP, Etheredge LM, Ganz PA, Wallace P, German RR, Neti C, et al. Rapid-learning system for cancer care. *J Clin Oncol.* (2010) 28:4268–74. doi: 10.1200/JCO.2010.28.5478
14. Kourou K. Machine learning applications in cancer prognosis and prediction. *Comput Struct Biotechnol J.* (2015) 13:8–17. doi: 10.1016/j.csbj.2014.11.005
15. Kang J, Schwartz R, Flickinger J, Beriwal S. Machine learning approaches for predicting radiation therapy outcomes: a clinician's perspective. *Int J Radiat Oncol.* (2015) 93:1127–35. doi: 10.1016/j.ijrobp.2015.07.2286
16. Tan P-N, Steinbach M, Kumar V. Introduction to Data Mining.169.
17. Drier Y, Domany E. Do Two machine-learning based prognostic signatures for breast cancer capture the same biological processes? *PLoS ONE.* (2011) 6:e17795. doi: 10.1371/journal.pone.0017795
18. Dupuy A, Simon RM. Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. *J Natl Cancer Inst.* (2007) 99:147–57. doi: 10.1093/jnci/djk018
19. Ein-Dor L, Kela I, Getz G, Givol D, Domany E. Outcome signature genes in breast cancer: is there a unique set? *Bioinforma Oxf Engl.* (2005) 21:171–8. doi: 10.1093/bioinformatics/bth469
20. Ein-Dor L, Zuk O, Domany E. Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. *Proc Natl Acad Sci USA.* (2006) 103:5923–28. doi: 10.1073/pnas.0601231103
21. Yang X, Wu N, Cheng G, Zhou Z, Yu DS, Beitler JJ, et al. Automated segmentation of the parotid gland based on atlas registration and machine learning: a longitudinal MRI study in head-and-neck radiation therapy. *Int J Radiat Oncol Biol Phys.* (2014) 90:1225–33. doi: 10.1016/j.ijrobp.2014.08.350
22. Veiga C, McClelland J, Moinuddin S, Lourenço A, Ricketts K, Annkah J, et al. Toward adaptive radiotherapy for head and neck patients: Feasibility study on using CT-to-CBCT deformable registration for "dose of the day" calculations. *Med Phys.* (2014) 41:031703. doi: 10.1118/1.4864240
23. Daisne J-F, Blumhofer A. Atlas-based automatic segmentation of head and neck organs at risk and nodal target volumes: a clinical validation. *Radiat Oncol.* (2013) 8:154. doi: 10.1186/1748-717X-8-154
24. Fortunati V, Verhaart RF, Lijn F van der, Niessen WJ, Veenland JF, Paulides MM, et al. Tissue segmentation of head and neck CT images for treatment planning: a multiatlas approach combined with intensity modeling. *Med Phys.* (2013) 40:071905. doi: 10.1118/1.4810971
25. Duc AKH, Eminowicz G, Mendes R, Wong S-L, McClelland J, Modat M, et al. Validation of clinical acceptability of an atlas-based segmentation algorithm for the delineation of organs at risk in head and neck cancer. *Med Phys.* (2015) 42:5027–34. doi: 10.1118/1.4927567
26. Hoogeman MS, Han X, Teguh D, Voet P, Nowak P, Wolf T, et al. Atlas-based Auto-segmentation of CT images in head and neck cancer: what is the best approach? *Int J Radiat Oncol Biol Phys.* (2008) 72:S591. doi: 10.1016/j.ijrobp.2008.06.196
27. Levendag PC, Hoogeman M, Teguh D, Wolf T, Hibbard L, Wijers O, et al. Atlas based auto-segmentation of ct images: clinical evaluation of using auto-contouring in high-dose, high-precision radiotherapy of cancer in the head and neck. *Int J Radiat Oncol Biol Phys.* (2008) 72:S401. doi: 10.1016/j.ijrobp.2008.06.1285
28. Qazi AA, Pekar V, Kim J, Xie J, Breen SL, Jaffray DA. Auto-segmentation of normal and target structures in head and neck CT images: a feature-driven model-based approach. *Med Phys.* (2011) 38:6160–70. doi: 10.1118/1.3654160
29. Sims R, Isambert A, Grégoire V, Bidault F, Fresco L, Sage J, et al. A pre-clinical assessment of an atlas-based automatic segmentation tool for the head and neck. *Radiother Oncol.* (2009) 93:474–8. doi: 10.1016/j.radonc.2009.08.013
30. Thomson D, Boylan C, Liptrot T, Aitkenhead A, Lee L, Yap B, et al. Evaluation of an automatic segmentation algorithm for definition of head and neck organs at risk. *Radiat Oncol.* (2014) 9:173. doi: 10.1186/1748-717X-9-173
31. Walker GV, Awan M, Tao R, Koay EJ, Boehling NS, Grant JD, et al. Prospective randomized double-blind study of atlas-based organ-at-risk autosegmentation-assisted radiation planning in head and neck cancer. *Radiother Oncol.* (2014) 112:321–5. doi: 10.1016/j.radonc.2014.08.028
32. Nikolov S, Blackwell S, Mendes R, De Fauw J, Meyer C, Hughes C, et al. Deep learning to achieve clinically applicable segmentation of head and neck anatomy for radiotherapy. *ArXiv Prepr. ArXiv180904430.* (2018) Available online at: https://arxiv.org/abs/1809.04430
33. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. In: eds. Navab N, Hornegger J, Wells WM, Frangi AF, editors. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015 Lecture Notes in Computer Science.* Springer International Publishing (2015). p. 234–241.

34. Castadot P, Lee JA, Parraga A, Geets X, Macq B, Grégoire V. Comparison of 12 deformable registration strategies in adaptive radiation therapy for the treatment of head and neck tumors. *Radiother Oncol J Eur Soc Ther Radiol Oncol.* (2008) 89:1–12. doi: 10.1016/j.radonc.2008.04.010
35. Olteanu LAM, Madani I, De Neve W, Vercauteren T, De Gersem W. Evaluation of deformable image coregistration in adaptive dose painting by numbers for head-and-neck cancer. *Int J Radiat Oncol.* (2012) 83:696–703. doi: 10.1016/j.ijrobp.2011.07.037
36. Capelle L, Mackenzie M, Field C, Parliament M, Ghosh S, Scrimger R. Adaptive radiotherapy using helical tomotherapy for head and neck cancer in definitive and postoperative settings: initial results. *Clin Oncol.* (2012) 24:208–15. doi: 10.1016/j.clon.2011.11.005
37. Sriram P, Syamkumar SA, Kumar JSD, Prabakar S, Dhanabalan R, Vivekanandan N. Adaptive volumetric modulated arc treatment planning for esophageal cancers using cone beam computed tomography. *Phys Med.* (2012) 28:327–32. doi: 10.1016/j.ejmp.2011.10.006
38. van Kranen S, Mencarelli A, van Beek S, Rasch C, van Herk M, Sonke J-J. Adaptive radiotherapy with an average anatomy model: Evaluation and quantification of residual deformations in head and neck cancer patients. *Radiother Oncol.* (2013) 109:463–8. doi: 10.1016/j.radonc.2013.08.007
39. Schwartz DL, Garden AS, Thomas J, Chen Y, Zhang Y, Lewin J, et al. Adaptive radiotherapy for head-and-neck cancer: initial clinical outcomes from a prospective trial. *Int J Radiat Oncol.* (2012) 83:986–93. doi: 10.1016/j.ijrobp.2011.08.017
40. Brouwer CL, Steenbakkers RJHM, Langendijk JA, Sijtsema NM. Identifying patients who may benefit from adaptive radiotherapy: does the literature on anatomic and dosimetric changes in head and neck organs at risk during radiotherapy provide information to help? *Radiother Oncol.* (2015) 115:285–94. doi: 10.1016/j.radonc.2015.05.018
41. Guidi G, Maffei N, Vecchi C, Ciarmatori A, Mistretta GM, Gottardi G, et al. A support vector machine tool for adaptive tomotherapy treatments: prediction of head and neck patients criticalities. *Phys Med.* (2015) 31:442–51. doi: 10.1016/j.ejmp.2015.04.009
42. Guidi G, Maffei N, Meduri B, D'Angelo E, Mistretta GM, Ceroni P, et al. A machine learning tool for re-planning and adaptive RT: A multicenter cohort investigation. *Phys Medica PM Int J Devoted Appl Phys Med Biol Off J Ital Assoc Biomed Phys AIFB.* (2016) 32:1659–66. doi: 10.1016/j.ejmp.2016.10.005
43. Marks LB, Yorke ED, Jackson A, Ten Haken RK, Constine LS, Eisbruch A, et al. Use of normal tissue complication probability models in the clinic. *Int J Radiat Oncol.* (2010) 76:S10–9. doi: 10.1016/j.ijrobp.2009.07.1754
44. Wu VWC, Ying MTC, Kwong DLW. Evaluation of radiation-induced changes to parotid glands following conventional radiotherapy in patients with nasopharygneal carcinoma. *Br J Radiol.* (2011) 84:843–9. doi: 10.1259/bjr/55873561
45. Scalco E, Fiorino C, Cattaneo GM, Sanguineti G, Rizzo G. Texture analysis for the assessment of structural changes in parotid glands induced by radiotherapy. *Radiother Oncol.* (2013) 109:384–7. doi: 10.1016/j.radonc.2013.09.019
46. Eisbruch A, Kim HM, Terrell JE, Marsh LH, Dawson LA, Ship JA. Xerostomia and its predictors following parotid-sparing irradiation of head-and-neck cancer. *Int J Radiat Oncol.* (2001) 50:695–704. doi: 10.1016/S0360-3016(01)01512-7
47. Zhang HH, D'Souza WD, Shi L, Meyer RR. Modeling plan-related clinical complications using machine learning tools in a multiplan IMRT framework. *Int J Radiat Oncol Biol Phys.* (2009) 74:1617–26. doi: 10.1016/j.ijrobp.2009.02.065
48. Chao KS, Deasy JO, Markman J, Haynie J, Perez CA, Purdy JA, et al. A prospective study of salivary function sparing in patients with head-and-neck cancers receiving intensity-modulated or three-dimensional radiation therapy: initial results. *Int J Radiat Oncol Biol Phys.* (2001) 49:907–16. doi: 10.1016/S0360-3016(00)01441-3
49. Eisbruch A, Ten Haken RK, Kim HM, Marsh LH, Ship JA. Dose, volume, and function relationships in parotid salivary glands following conformal and intensity-modulated irradiation of head and neck cancer. *Int J Radiat Oncol.* (1999) 45:577–87. doi: 10.1016/S0360-3016(99)00247-3
50. Eisbruch A, Ship JA, Kim HM, Ten Haken RK. Partial irradiation of the parotid gland. *Semin Radiat Oncol.* (2001) 11:234–9. doi: 10.1053/srao.2001.23484
51. Sanguineti G, Rao N, Gunn B, Ricchetti F, Fiorino C. Predictors of PEG dependence after IMRT±chemotherapy for oropharyngeal cancer. *Radiother Oncol.* (2013) 107:300–4. doi: 10.1016/j.radonc.2013.05.021
52. Trotti A. Toxicity in head and neck cancer: a review of trends and issues. *Int J Radiat Oncol Biol Phys.* (2000) 47:1–12. doi: 10.1016/S0360-3016(99)00558-1
53. Trotti A, Bellm LA, Epstein JB, Frame D, Fuchs HJ, Gwede CK, et al. Mucositis incidence, severity and associated outcomes in patients with head and neck cancer receiving radiotherapy with or without chemotherapy: a systematic literature review. *Radiother Oncol J Eur Soc Ther Radiol Oncol.* (2003) 66:253–62. doi: 10.1016/S0167-8140(02)00404-8
54. Robertson AG, Robertson C, Perone C, Clarke K, Dewar J, Elia MH, et al. Effect of gap length and position on results of treatment of cancer of the larynx in Scotland by radiotherapy: a linear quadratic analysis. *Radiother Oncol J Eur Soc Ther Radiol Oncol.* (1998) 48:165–73.
55. Bourhis J, Overgaard J, Audry H, Ang KK, Saunders M, Bernier J, et al. Hyperfractionated or accelerated radiotherapy in head and neck cancer: a meta-analysis. *Lancet Lond Engl.* (2006) 368:843–54. doi: 10.1016/S0140-6736(06)69121-6
56. Dehing-Oberije C, De Ruysscher D, Petit S, Van Meerbeeck J, Vandecasteele K, De Neve W, et al. Development, external validation and clinical usefulness of a practical prediction model for radiation-induced dysphagia in lung cancer patients. *Radiother Oncol J Eur Soc Ther Radiol Oncol.* (2010) 97:455–61. doi: 10.1016/j.radonc.2010.09.028
57. Lambin P, Roelofs E, Reymen B, Velazquez ER, Buijsen J, Zegers CML, et al. 'Rapid Learning health care in oncology' – An approach towards decision support systems enabling customised radiotherapy'. *Radiother Oncol.* (2013) 109:159–164. doi: 10.1016/j.radonc.2013.07.007
58. Langendijk JA, Lambin P, De Ruysscher D, Widder J, Bos M, Verheij M. Selection of patients for radiotherapy with protons aiming at reduction of side effects: the model-based approach. *Radiother Oncol J Eur Soc Ther Radiol Oncol.* (2013) 107:267–.73. doi: 10.1016/j.radonc.2013.05.007
59. Kierkels RGJ, Korevaar EW, Steenbakkers RJHM, Janssen T, Veld A van t, Langendijk JA, et al. Direct use of multivariable normal tissue complication probability models in treatment plan optimisation for individualised head and neck cancer radiotherapy produces clinically acceptable treatment plans. *Radiother Oncol.* (2014) 112:430–6. doi: 10.1016/j.radonc.2014.08.020
60. Buettner F, Miah AB, Gulliford SL, Hall E, Harrington KJ, Webb S, et al. Novel approaches to improve the therapeutic index of head and neck radiotherapy: An analysis of data from the PARSPORT randomised phase III trial. *Radiother Oncol.* (2012) 103:82–7. doi: 10.1016/j.radonc.2012.02.006
61. Dean JA, Wong KH, Welsh LC, Jones A-B, Schick U, et al. Normal tissue complication probability (NTCP) modelling using spatial dose metrics and machine learning methods for severe acute oral mucositis resulting from head and neck radiotherapy. *Radiother Oncol.* (2016) 120:21–7. doi: 10.1016/j.radonc.2016.05.015
62. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (tripod): explanation and elaboration. *Ann Intern Med.* (2015) 162:W1. doi: 10.7326/M14-0698
63. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc Ser B Methodol.* (1996) 58:267–88.
64. Cortes C, Vapnik V. Support-vector networks. *Mach Learn.* (1995) 20:273–97. doi: 10.1007/BF00994018
65. Breiman L. Random forests. *Mach Learn.* (2001) 45:5–32. doi: 10.1023/A:1010933404324
66. Sanguineti G, Sormani MP, Marur S, Gunn GB, Rao N, Cianchetti M, et al. Effect of radiotherapy and chemotherapy on the risk of mucositis during intensity-modulated radiation therapy for oropharyngeal cancer. *Int J Radiat Oncol.* (2012) 83:235–242. doi: 10.1016/j.ijrobp.2011.06.2000
67. Dean JA, Welsh LC, Gulliford SL, Harrington KJ, Nutting CM. A novel method for delineation of oral mucosa for radiotherapy dose–response studies. *Radiother Oncol.* (2015) 115:63–66. doi: 10.1016/j.radonc.2015.02.020
68. Dean JA, Welsh LC, McQuaid D, Wong KH, Aleksic A, Dunne E, et al. Assessment of fully-automated atlas-based segmentation of novel

oral mucosal surface organ-at-risk. *Radiother Oncol.* (2016) 119:166–71. doi: 10.1016/j.radonc.2016.02.022
69. Yu K, Zhang Y, Yu Y, Huang C, Liu R, Li T, et al. Radiomic analysis in prediction of human papilloma virus status. *Clin Transl Radiat Oncol.* (2017) 7:49–54. doi: 10.1016/j.ctro.2017.10.001
70. Maxwell JH, Rath TJ, Byrd JK, Albergotti WG, Wang H, Duvvuri Uet al. Accuracy of computed tomography to predict extracapsular spread in p16-positive squamous cell carcinoma. *Laryngoscope.* (2015) 125:1613–8. doi: 10.1002/lary.25140
71. Kann BH, Aneja S, Loganadane GV, Kelly JR, Smith SM, Decker RH, et al. Pretreatment identification of head and neck cancer nodal metastasis and extranodal extension using deep learning neural networks. *Sci Rep.* (2018) 8:14036. doi: 10.1038/s41598-018-32441-y
72. Bryce TJ, Dewhirst MW, Floyd CE, Hars V, Brizel DM. Artificial neural network model of survival in patients treated with irradiation with and without concurrent chemotherapy for advanced carcinoma of the head and neck. *Int J Radiat Oncol.* (1998) 41:339–45. doi: 10.1016/S0360-3016(98)00016-9
73. Brizel DM. Radiotherapy and concurrent chemotherapy for the treatment of locally advanced head and neck squamous cell carcinoma. *Semin Radiat Oncol.* (1998) 8:237–46. doi: 10.1016/S1053-4296(98)80021-0
74. Widrow B, Lehr MA. 30 years of adaptive neural networks: perceptron, Madaline, and backpropagation. *Proc IEEE.* (1990) 78:1415–42. doi: 10.1109/5.58323
75. Ou D, Blanchard P, Rosellini S, Levy A, Nguyen F, Leijenaar RTH, et al. Predictive and prognostic value of CT based radiomics signature in locally advanced head and neck cancers patients treated with concurrent chemoradiotherapy or bioradiotherapy and its added value to Human Papillomavirus status. *Oral Oncol.* (2017) 71:150–5. doi: 10.1016/j.oraloncology.2017.06.015
76. Levy A, Blanchard P, Bellefqih S, Brahimi N, Guigay J, Janot F, et al. Concurrent use of cisplatin or cetuximab with definitive radiotherapy for locally advanced head and neck squamous cell carcinomas. *Strahlenther Onkol.* (2014) 190:823–31. doi: 10.1007/s00066-014-0626-0
77. Harrell F. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York, NY: Springer-Verlag (2001). Available online at: https://www.springer.com/gb/book/9781441929181 (Accessed October 8, 2018)
78. Zhu Y, Mohamed AS, Lai SY, Yang S, Kanwar A, Wei L, et al. Imaging-genomics study of head-neck squamous cell carcinoma: associations between radiomic phenotypes and genomic mechanisms via integration Of TCGA and TCIA. *bioRxiv.* (2017) 214312. doi: 10.1101/214312
79. M. D. Anderson Cancer Center Head and Neck Quantitative Imaging Working Group. Investigation of radiomic signatures for local recurrence using primary tumor texture analysis in oropharyngeal head and neck cancer patients. *Sci Rep.* (2018) 8:1524. doi: 10.1038/s41598-017-14687-0
80. Zhang L, Zhao C, Peng P-J, Lu L-X, Huang P-Y, Han F, et al. Phase III study comparing standard radiotherapy with or without weekly oxaliplatin in treatment of locoregionally advanced nasopharyngeal carcinoma: preliminary results. *J Clin Oncol Off J Am Soc Clin Oncol.* (2005) 23:8461–8. doi: 10.1200/JCO.2004.00.3863
81. Zhang S, Zhang B, Tian J, Dong D, Sheng GD, Hao DY, et al. Radiomics features of Multiparametric MRI as novel prognostic factors in advanced nasopharyngeal carcinoma. *Clin Cancer Res.* (2017) 23:4259–69. doi: 10.1158/1078-0432.CCR-16-2910
82. Li S, Wang K, Hou Z, Yang J, Ren W, Gao S, et al. Use of radiomics combined with machine learning method in the recurrence patterns after intensity-modulated radiotherapy for nasopharyngeal carcinoma: a preliminary study. *Front Oncol.* (2018) 8:648. doi: 10.3389/fonc.2018.00648
83. Zhang L, Dong D, Li H, Tian J, Ouyang F, Mo X, et al. Development and validation of a magnetic resonance imaging-based model for the prediction of distant metastasis before initial treatment of nasopharyngeal carcinoma: A retrospective cohort study. *EBioMedicine.* (2019) (in press) doi: 10.1016/j.ebiom.2019.01.013
84. Kwan JYY, 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. doi: 10.1016/j.ijrobp.2018.01.057
85. Bibault J-E, Giraud P, Burgun A. Big data and machine learning in radiation oncology: state of the art and future prospects. *Cancer Lett.* (2016) 382:110–7. doi: 10.1016/j.canlet.2016.05.033

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Giraud, Giraud, Gasnier, El Ayachy, Kreps, Foy, Durdux, Huguet, Burgun and Bibault. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.