

Chapter 2

Biomarkers for Hypoxia, HPVness, and Proliferation from Imaging Perspective



Sebastian Sanduleanu, Simon Keek, Lars Hoezen, and Philippe Lambin

Introduction

Advances in imaging and treatment technology over the last few decades have brought an improvement in locoregional control among head and neck squamous cell carcinoma (HNSCC). Despite advances in treatment, from robotic surgery to new systemic therapies such as immuno-radiation and programmed cell death protein-1/programmed cell death ligand-1 (PD-1/PD-L1) blockers for metastatic disease, the overall survival rates are still poor with around 50% 5-year survival. This is mainly caused by treatment resistance, recurrence and distant metastasis, which in turn can be caused by hypoxia, resistance due to clonogenic cell populations, and inadequate immune response [1, 2].

Adequate staging and tumor delineation through molecular imaging and imaging biomarkers based on routine clinical images could improve the precision of radiotherapy and surgery, which may lead to a reduction of recurrences.

Radiomics and deep learning are machine learning techniques that have the potential to infer quantitative information from routine medical images in HNSCC [3] (Fig. 2.1). Imaging biomarkers derived from such techniques can be predictive

S. Sanduleanu (✉) · S. Keek · L. Hoezen

The D-Lab and the M-Lab, Department of Precision Medicine, GROW—School for Oncology, Maastricht University, Maastricht, The Netherlands

e-mail: s.sanduleanu@maastrichtuniversity.nl; s.keek@maastrichtuniversity.nl;
l.hoezen@maastrichtuniversity.nl

P. Lambin

The D-Lab and the M-Lab, Department of Precision Medicine, GROW—School for Oncology, Maastricht University, Maastricht, The Netherlands

Department of Radiology and Nuclear Medicine, GROW—School for Oncology and Developmental Biology, Maastricht University Medical Centre+, Maastricht, The Netherlands
e-mail: philippe.lambin@maastrichtuniversity.nl

© The Author(s) 2021

J. B. Vermorcken et al. (eds.), *Critical Issues in Head and Neck Oncology*,
https://doi.org/10.1007/978-3-030-63234-2_2

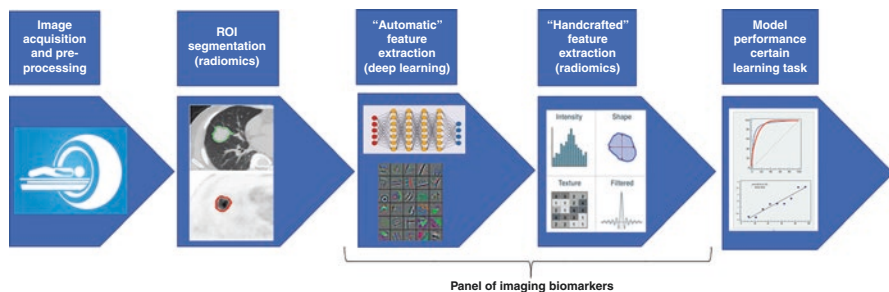


Fig. 2.1 The radiomics and deep learning workflow. Medical images are acquired, pre-processed, and are provided to the deep learning/radiomics workflow. Region of interest (ROI) segmentation is required for radiomics analysis and can be done manually or with automatic segmentation (deep learning). The radiomic features and deep features can be combined using a feature merge layer on which predictions are based. The feature merge layer can comprise a neural network layer but also a machine learning model in which only the most salient features from both pipelines are fed. Eventually the model performance for a specific learning task is assessed

and/or prognostic. A prognostic biomarker provides information about the trajectory/outcome of a patient with cancer, regardless of therapy. Meanwhile, a predictive biomarker is a biomarker that can represent a subgroup of patients who are most likely to respond to the therapy in question. In order to distinguish these two terms, the biomarker-positive and -negative subgroups and experimental and control subgroups are needed. So, when the experimental group shows a difference in survival when tested positive and negative and the survival of the negative response is higher than the control group, this is a prognostic biomarker. If the control group shows no differences in survival when tested positive and negative this is a predictive biomarker. These can also be combined; this means a biomarker can be both prognostic and predictive.

The aim in this chapter is to discuss current trends in head and neck oncology imaging, from imaging biomarkers for HPV-status and hypoxia to recent advances in artificial intelligence (AI) in head and neck oncology.

Imaging Biomarkers for the Assessment of HPV-ness

Human papilloma virus (HPV) positive oropharyngeal squamous cell carcinoma (OSCC) is a rapidly increasing group of patients worldwide (from 16% to 73% in the last 20 years) which responds much better to therapy, whether this is surgery, radiation, or chemotherapy [4, 5]. HPV positive patients have therefore been considered extensively for de-escalation trials [6] in order to decrease toxicity while achieving similar control rates. In 2018, the HPV status of the patient was implemented in the 8th edition of the American Joint Committee on Cancer (AJCC) staging of OSCC [7]. In this staging method, p16 immunohistochemistry (IHC) is used

as a surrogate marker for high-risk HPV [8]. However, p16 IHC is not a perfect surrogate marker for HPV, and consensus on the best way to determine HPV status has currently not been reached [9]. An example of a standard method to determine HPV is polymerase chain reaction (PCR) on paraffin-embedded tissue. However, this method is expensive and time-consuming, and requires the invasive procedure to acquire a biopsy. A study by Molony et al. [10] shows tumor morphology, classified as keratinizing or non-keratinizing, is a significant predictor of HPV status and performs better in determining HPV status in combination with p16 IHC compared to p16 IHC alone. Previous studies have suggested computed tomography (CT) readouts of the tumor showed phenotypical differences between HPV-positive and -negative tumors [11], suggesting an alternative method to determine HPV-status. Indeed, Leijenaar et al. [12] developed a signature based on radiomic features to predict HPV status on routine clinical CT images, showing potential for the determination of HPVness through different methods.

Imaging Biomarkers for Tumor Hypoxia

Tumor hypoxia, also known as the occurrence of oxygen-deficient areas within the tumor, is a known prognostic factor in head and neck cancer. One way to look at both diffusion and perfusion-limited hypoxia is to look at vascular density, vascular permeability, blood volume, and blood flow within the tumor with dynamic contrast enhanced (DCE)-magnetic resonance imaging (MRI)/CT. Although perfusion CT has a dedicated FDA-cleared analysis software and displays greater resolution when compared to DCE-MRI, the required dose of ionizing radiation limits its ability to be used in trials with repeated scanning. In DCE-MRI, moving artifacts from breathing and swallowing and the susceptibility artifacts from interface air-tissue are frequent when scanning the head and neck region with this method, which could substantially affect the tumor-segmentation accuracy and the quantitative imaging biomarker (radiomics) feature extraction. Therefore, at this moment, the data obtained from pre-treatment DCE-MRI seems to be insufficient to allow translation to clinical practice. To our knowledge there is not a single DCE-MRI imaging biomarker study in head and neck looking specifically into association with (histopathologically confirmed) tumor hypoxia, though there are e.g. multiparametric MRI-based prognostic signatures for e.g. advanced nasopharyngeal carcinoma.

Hypoxia imaging PET radiotracers such as ^{18}F -FMISO and ^{18}F -HX4 are promising but not widely available. Hypoxia PET imaging is nevertheless difficult to implement in clinical practice since these PET-agents generally tend to generate smaller signal-to-background ratios compared to e.g. ^{18}F -FDG (and consequently lower target-background image contrast), imaging is labor intensive (instructions of multiple bed positions and acquisitions at multiple time points), costly (chemical process to produce the radioligand is slightly more expensive), and lacking standard

calibration procedures and inconvenient for the patient due to the time-consuming acquisition protocols. Another way would be to infer quantitative imaging biomarkers from routine ^{18}F -FDG PET and (contrast enhanced) CT images using hypoxia PET tracers as gold standard for training these models.

The aim of the study by Crispin-Ortuzar et al. [13] for instance was to design a surrogate biomarker for ^{18}F -FMISO maximum tumor-to-blood uptake ratio (TBR_{max}) based on pre-treatment ^{18}F -FDG PET and contrast-enhanced CT imaging features. The level of hypoxia of a lesion was defined in terms of its TBR_{max} on the last static scan. In particular, in this study a lesion was considered to be hypoxic if $\text{TBR}_{\text{max}} > 1.4$. The further aim was to study its performance in the context of hypoxia-based patient stratification. In her study, 121 lesions from 75 head and neck cancer patients were used in the analysis. Patients received both pre-treatment ^{18}F -FDG and ^{18}F -FMISO PET/CT scans. In total, 79 lesions were used to train a cross-validated least absolute shrinkage and selection operator (LASSO) regression model based on quantitative imaging features, while the remaining 42 were held out for internal testing. The best performance on the unseen test subset in this study was obtained from the combined CT and ^{18}F -FDG PET signature, with an area under the receiver operating characteristic curve (AUC) of 0.833, while the model based on the 90th percentile of ^{18}F -FDG uptake alone had a test AUC of 0.756.

Such imaging biomarkers, when improved to accurately detect hypoxia, could be used to stratify patients for hypoxia-modifying therapy.

Evaluation Treatment Response with RECIST 1.1

Objective assessment of both tumor shrinkage as well as time to development of disease progression after (non-)cytotoxic systemic therapy are important endpoints both in clinical trials as well as on patient-level. The revised response evaluation criteria in solid tumors (RECIST) 1.1 in 2009 [14] sought to improve the accuracy and efficacy of this assessment by (1) reducing the maximum of lesions for longest diameter measurement from 10 to 5 (in maximum two organs) (2) disease progression (PD) not only requires 20% increase in the sum of measurements, but also a 5 mm absolute increase (to guard against over calling PD when the total sum is very small) (3) inclusion of FDG-PET response assessment as an adjunct to determination of progression.

One of the key questions for debate by the RECIST Working group developing RECIST 1.1 was whether it is appropriate to move from anatomic unidimensional assessment of tumor burden to either a volumetric assessment or to a more functional assessment with MRI and/or PET. At that point the Working Group concluded that there is not sufficient standardization or evidence to abandon the current unidimensional anatomical assessment of tumor burden and functional imaging with FDG-PET was only to be used as an adjunct for the determination of progression.

The Potential for Automatic Head and Neck Segmentation and Volumetric RECIST Assessment as Surrogate Imaging Marker for Tumor Proliferation

Treatment planning for high precision radiotherapy of head and neck cancer patients requires accurate delineation of many organs at risk for radiation induced injury as well as gross tumor volumes and (elective) lymph node regions. Manual contouring is a laborious task which suffers from large inter- and intra-rater variability. To reduce manual labor, several fully automated, atlas-based [15] as well as deep learning based [16] methods for head and neck CT image segmentation have been developed. Although these methods save a considerable amount of time as they do not require human input, they are also prone to errors [17].

Significant challenges arise currently using RECIST 1.1 endpoints, which could be mitigated by volumetric methods. First and most foremost, during treatment the longest diameter of the tumors may remain unchanged, while the irregularly shaped and morphologically complex tumors may still shrink in terms of absolute volume (Fig. 2.2).

Secondly, while the RECIST criteria were developed traditionally to assess the efficacy of cytostatic drugs, while other systemic therapies may not shrink tumor size but rather trigger a cytostatic response or alter the physiological properties of a tumor such as metabolism, cell proliferation, and angiogenesis. In the case of immunotherapy, initial tumor enlargement is common, which according to RECIST would be classified as progressive disease.

One of the ways to deal with these challenges is to quantify volumetric measures on CT, MRI, and PET as biomarkers for systemic treatment response is as addressed in the Quantitative Imaging Biomarker Alliance (QIBA) profile initiative in 2007 by the Radiological Society of North America (RSNA). The main purpose of this initiative was to unite researchers, healthcare professionals, and industry to advance quantitative imaging and the use of imaging biomarkers in clinical trials and clinical

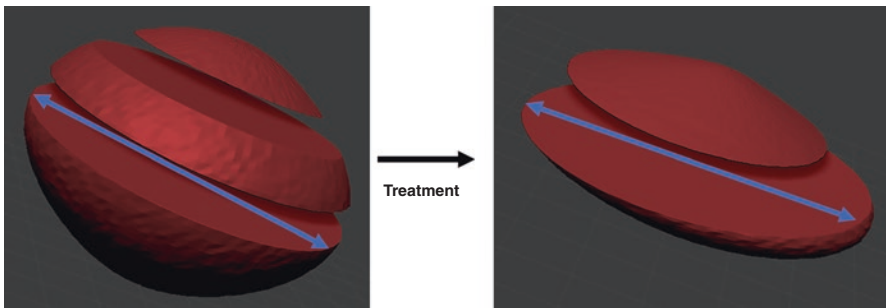


Fig. 2.2 Traditional longest diameter measurement according to RECIST versus volume: longest diameter remains unchanged while overall volume shrinks

practice. One of the QIBA committees is involved in establishing a process map (measurement accuracy, technical feasibility, and comparison with standard RECIST measurements) for qualifying volumetric measures on CT as a biomarker for treatment response as well as determining whether changes in volume are medically meaningful or just add to overall costs and complexity of care [18].

Early treatment response assessment allows the physician to stop an ineffective treatment sooner and enable a transition to a more effective alternative. Generally, volumetric tumor assessment is more costly and time-consuming to perform. Nevertheless, the greater sensitivity [18] associated with volumetric measurement can increase the statistical power per subject, resulting in fewer patient inclusions in clinical trials followed up over shorter periods of time and subsequently decreasing overall time and cost. Additionally, the question is whether volumetric imaging adds value to a clinical trial, in other words if it significantly impacts clinical decision-making. Although this issue still remains to be determined and validated, some preliminary findings find a role for volumetric imaging. In one retrospective study by Hayes et al. [19] on 42 lung cancer patients participating in an open-label phase 2 study, volumetric measurements (semi-automatic segmentation algorithm on CT) on first follow-up (4 weeks after start of treatment) were better able to predict overall survival than RECIST measurements. A second study by Kim et al. [20] found in a cohort of 135 non-small cell lung cancer patients that hyperprogressive disease treated with immune checkpoint inhibitors on the basis of volumetric measurement is more precise than is defining it on the basis of one-dimensional analysis in terms of overall survival. To our knowledge at the moment this chapter has been written there were no such volumetric versus RECIST comparison initiatives in head and neck cancer.

Conclusions and Future Directions

In recent years, explainable AI (XAI), the implementation of transparency and traceability of statistical black-box machine learning methods [21], has been attracting much interest in medicine. The reenactment of the machine decision-making process is necessary not only to comprehend and reproduce the learning and extraction process, but also because for medical decision support it is necessary to understand the causality of learned representations [22–24]. Furthermore, the implementation of explainable AI would help to enhance the trust of medical professionals in future AI-systems. Nevertheless, currently there is still an inherent tension between machine learning performance (predictive accuracy) and explainability, as often the best-performing methods such as deep learning are the least transparent, and the ones providing a clear explanation (e.g. decision trees) are less accurate [25]. This still makes this very much an active area of research. Advanced functional imaging techniques to address the inherent limitations of the current RECIST, such as perfusion CT, dynamic contrast-enhanced MRI, and diffusion-weighted MRI are currently only considered to be experimental endpoints because they have not yet completed the rigorous validation process needed to qualify as

true surrogate endpoints. With the advent and improvement of explainable automatic segmentation algorithms, volumetric endpoints (perhaps with the aid of advanced functional imaging techniques) in the near future will offer increased sensitivity to anatomical measurements and provide the necessary physiological information to interpret response to highly selective, patient tailored therapies, particularly in the cases where RECIST falls short.

References

1. Marcu LG, Reid P, Bezak E. The promise of novel biomarkers for head and neck cancer from an imaging perspective. *Int J Mol Sci.* 2018;19:2511.
2. Alshahafi E, Begg K, Amelio I, Raulf N, Lucarelli P, Sauter T, Tavassoli M. Clinical update on head and neck cancer: molecular biology and ongoing challenges. *Cell Death Dis.* 2019;10:540.
3. 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.
4. Raj Sindwani BJB, Franco RA, Gapany M, Mitchell RB. Year book of otolaryngology – head and neck surgery 2011. 1st ed. London: Elsevier; 2011.
5. De Felice F, Tombolini V, Valentini V, De Vincentiis M, Mezi S, Brugnoletti O, Polimeni A. Advances in the management of hpv-related oropharyngeal cancer. *J Oncol.* 2019; 2019:9173729–9.
6. Masterson L, Moualed D, Liu ZW, Howard JEF, Dwivedi RC, Tysome JR, Benson R, Sterling JC, Sudhoff H, Jani P, Goon PKC. De-escalation treatment protocols for human papillomavirus-associated oropharyngeal squamous cell carcinoma: a systematic review and meta-analysis of current clinical trials. *Eur J Cancer.* 2014;50:2636–48.
7. Amin MB, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, Meyer L, Gress DM, Byrd DR, Winchester DP. The eighth edition AJCC cancer staging manual: continuing to build a bridge from a population-based to a more “personalized” approach to cancer staging. *CA: Cancer J Clin.* 2017;67:93–9.
8. Duncan LD, Winkler M, Carlson ER, Heidel RE, Kang E, Webb D. p16 immunohistochemistry can be used to detect human papillomavirus in oral cavity squamous cell carcinoma. *J Oral Maxillofac Surg.* 2013;71:1367–75.
9. Chaudhary AK, Singh M, Sundaram S, Mehrotra R. Role of human papillomavirus and its detection in potentially malignant and malignant head and neck lesions: updated review. *Head Neck Oncol.* 2009;1:22.
10. Molony P, Werner R, Martin C, Callanan D, Nauta I, Heideman D, Sheahan P, Heffron C, Feeley L. The role of tumour morphology in assigning hpv status in oropharyngeal squamous cell carcinoma. *Oral Oncol.* 2020;105:104670.
11. Chan MW, Yu E, Bartlett E, O’sullivan B, Su J, Waldron J, Ringash J, Bratman SV, Chen YA, Irish J. Morphologic and topographic radiologic features of human papillomavirus-related and-unrelated oropharyngeal carcinoma. *Head & Neck.* 2017;39:1524–34.
12. Leijenaar RT, Bogowicz M, Jochems A, Hoebbers FJ, Wesseling FW, Huang SH, Chan B, Waldron JN, O’sullivan B, Rietveld D, Leemans CR, Brakenhoff RH, Riesterer O, Tanadini-Lang S, Guckenberger M, Ikenberg K, Lambin P. Development and validation of a radiomic signature to predict hpv (p16) status from standard CT imaging: a multicenter study. *Br J Radiol.* 2018;91:20170498.
13. Crispin-Ortuzar M, Apte A, Grkovski M, Oh JH, Lee NY, Schöder H, Humm JL, Deasy JO. Predicting hypoxia status using a combination of contrast-enhanced computed tomography and [(18)f]-fluorodeoxyglucose positron emission tomography radiomics features. *Radiother Oncol.* 2018;127:36–42.

14. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, Dancey J, Arbuck S, Gwyther S, Mooney M, Rubinstein L, Shankar L, Dodd L, Kaplan R, Lacombe D, Verweij J. New response evaluation criteria in solid tumours: revised recist guideline (version 1.1). *Eur J Cancer*. 2009;45:228–47.
15. Han X, Hoogeman M, Levendag P, Hibbard L, Teguh D, Voet P, Cowen A, Wolf T. Atlas-based auto-segmentation of head and neck CT images. *Med Image Comput Comput Assist Interv*. 2008;11:434–41.
16. Van Dijk LV, Van Den Bosch L, Aljabar P, Peressutti D, Both S, Steenbakkens RJHM. Improving automatic delineation for head and neck organs at risk by deep learning contouring. *Radiother Oncol*. 2020;142:115–23.
17. Goldmacher GV, Conklin J. The use of tumour volumetrics to assess response to therapy in anticancer clinical trials. *Br J Clin Pharmacol*. 2012;73:846–54.
18. Mozley PD, Schwartz LH, Bendtsen C, Zhao B, Petrick N, Buckler AJ. Change in lung tumor volume as a biomarker of treatment response: a critical review of the evidence. *Ann Oncol*. 2010;21:1751–5.
19. Hayes SC. Acceptance and commitment therapy, relational frame theory, and the third wave of behavioral and cognitive therapies—republished article. *Behav Ther*. 2016;47:869–85.
20. Kim Y, Kim CH, Lee HY, Lee S-H, Kim HS, Lee S, Cha H, Hong S, Kim K, Seo SW, Sun J-M, Ahn M-J, Ahn JS, Park K. Comprehensive clinical and genetic characterization of hyper-progression based on volumetry in advanced nonsmall cell lung cancer treated with immune checkpoint inhibitor. *J Thorac Oncol*. 2019;14:1608–18.
21. Arietta AB, Diaz-Rodriguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, Garcia S, Gil-Lopez S, Molina D, Benjamins R, Chatila R, Herrera F. Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges towards responsible AI. *Inf Fusion*. 2019;58:82–115.
22. Gershman SJ, Horvitz EJ, Tenenbaum JB. Computational rationality: a converging paradigm for intelligence in brains, minds, and machines. *Science*. 2015;349:273–8.
23. Pearl J. Causal inference in statistics: an overview. *Statist Surv*. 2009;3:96–146.
24. Peters J, Janzing D, Schölkopf B. Elements of causal inference: foundations and learning algorithms. Boston: MIT; 2017.
25. Bologna G, Hayashi Y. Characterization of symbolic rules embedded in deep DIMLP networks: a challenge to transparency of deep learning. *J Artif Intell Soft Comput Res*. 2017;7:265–86.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

