Fully Automatic Head and Neck Cancer Prognosis Prediction in PET/CT

Pierre Fontaine[1,2,*], Vincent Andrearczyk[1], Valentin Oreiller[1,3], Joël Castelli[2], Mario Jreige[3], John O. Prior[3], and Adrien Depeursinge[1,3]

[1] University of Applied Sciences and Arts Western Switzerland (HES-SO), Sierre, Switzerland
[2] Univ. Rennes, CLCC Eugene Marquis, INSERM, LTSI - UMR 1099, F-35000 Rennes, France
[3] Centre Hospitalier Universitaire Vaudois (CHUV), Lausanne, Switzerland
[*] pierre.fontaine@hevs.ch

Abstract. Several recent PET/CT radiomics studies have shown promising results for the prediction of patient outcomes in Head and Neck (H&N) cancer. These studies, however, are most often conducted on relatively small cohorts (up to 300 patients) and using manually delineated tumors. Recently, deep learning reached high performance in the automatic segmentation of H&N primary tumors in PET/CT. The automatic segmentation could be used to validate these studies on larger-scale cohorts while obviating the burden of manual delineation. We propose a complete PET/CT processing pipeline gathering the automatic segmentation of primary tumors and prognosis prediction of patients with H&N cancer treated with radiotherapy and chemotherapy. Automatic contours of the primary Gross Tumor Volume (GTVt) are obtained from a 3D UNet. A radiomics pipeline that automatically predicts the patient outcome (Disease Free Survival, DFS) is compared when using either the automatically or the manually annotated contours. In addition, we extract deep features from the bottleneck layers of the 3D UNet to compare them with standard radiomics features (first- and second-order as well as shape features) and to test the performance gain when added to them. The models are evaluated on the HECKTOR 2020 dataset consisting of 239 H&N patients with PET, CT, GTVt contours and DFS data available (five centers). Regarding the results, using Hand-Crafted (HC) radiomics features extracted from manual GTVt achieved the best performance and is associated with an average Concordance (C) index of 0.672. The fully automatic pipeline (including deep and HC features from automatic GTVt) achieved an average C index of 0.626, which is lower but relatively close to using manual GTVt (p-value = 0.20). This suggests that large-scale studies could be conducted using a fully automatic pipeline to further validate the current state of the art H&N radiomics. The code will be shared publicly for reproducibility.

Keywords: Head and Neck Cancer · radiomics · deep learning

1 Introduction

Radiomics allows quantitative analyses from radiological images with high throughput extraction to obtain prognostic patient information [1]. Radiomics features, including intensity-, texture-, and shape-based features, are generally extracted from Volumes Of Interests (VOI) that delineate the tumor volume [2]. These VOIs are often obtained from manual delineations made for treatment planning in radiotherapy [3, 4] or from a simple threshold on the Positron Emission Tomography (PET) image [5]. This approach allowed researchers to perform various radiomics studies without the need for re-annotating the images specifically for these tasks. The annotations made for radiotherapy are, however, very large as compared to the true tumoral volumes and frequently include non-tumoral tissues and parts of other organs such as the trachea. On the other hand, simple PET thresholding may discard important hypo-metabolic parts of the tumoral volume (*e.g.* necrosis) containing important information for outcome prediction. For that reason, it seems important to clean delineations before using them in a radiomics study. Unfortunately, this represents a tedious and error-prone task when done manually, and usually entails poor inter-observer agreement [6]. Moreover, it is not scalable to large cohorts including thousands of patients, which is urgently needed to validate the true value of radiomics-based prognostic biomarkers when used in clinical routine.

Since the introduction of deep Convolutional Neural Network (CNN) segmentation models and in particular of the UNet architecture [7], the automatic segmentation of the tumoral volume in radiological and nuclear medicine images has made tremendous progress during the past ten years. In particular for brain tumor segmentation, expert-level performance was achieved in the context of the Brain Tumor Segmentation (BraTS) challenge [8, 9]. The brain tumor segmentation models were further extended or used for prognostic prediction, allowing to obtain fully automatic methods that do not require manual delineation of VOIs for extracting the radiomics features. In particular, Baid *et. al.* [10] used a 3D UNet for the segmentation task and then extracted radiomics features from the automatically segmented tumor components in multi-sequence Magnetic Resonance Imaging (MRI), achieving good performance for the prediction of overall patient survival. More recently, the automatic segmentation of Head and Neck (H&N) tumors from PET/CT imaging was investigated in the context of the HECKTOR challenge [6], also achieving expert-level delineation performance of the primary Gross Tumor Volume (GTVt). This opened avenues for scalable and fully automatic prognosis prediction in Head and Neck (H&N) cancer as well as their validation on large-scale cohorts.

In this paper, we propose a fully automatic pipeline for predicting patient prognosis (*i.e.* Disease-Free Survival, DFS) in oropharyngeal H&N cancer. We use a 3D UNet for segmenting the tumor contours and to extract deep radiomics features from its encoder. We further extract Hand-Crafted (HC) radiomics features from the automatically generated GTVt contours. Both PET and CT modalities are used to characterize the metabolic and morphological tumor tissue properties. Finally, we compare and combine the deep and HC

radiomics features using either Cox Proportional Hazards (CPH) or Random Survival Forest (RSF) models. This fully automatic approach is compared to a classical radiomics approach based on HC features extracted from manually annotated VOIs.

2 Methods

This section first introduces the dataset. Then, the proposed manual and fully automatic radiomics pipelines are detailed. In particular, we compare the performance between prognostic radiomics models using either *manual* or *automatic* annotations of GTVt. We also compare the prognostic performance when using spatially aggregated activation maps from the bottleneck of the 3D UNet (*i.e.* the bottom of the encoder, referred to as *deep features*) and even combine these with HC radiomics features to investigate their complementarity for DFS prediction.

2.1 Datasets

For this study, we use the entire dataset used in the context of the HECKTOR 2020 challenge [6] (training and test sets) including a total of 239 patient with H&N cancer located in the oropharyngeal region (see Table 1). This dataset is an extension of the data used in the study of Vallières *et al.* 2017 [3].

Table 1. Overview of the dataset. Hôpital Général Juif (HGJ), Montréal, CA; Centre Hospitalier Universitaire de Sherbooke (CHUS), Sherbrooke, CA; Hôpital Maisonneuve-Rosemont (HMR), Montréal, CA; Centre Hospitalier de l'Université de Montréal (CHUM), Montréal; Centre Hospitalier Universitaire Vaudois (CHUV), CH.

Center	#patients	Gender	Age (avg.)	Follow-up (avg. days)	#events (DFS)
HGJ	55	Male 43 Female 12	62	1418	11
CHUS	71	Male 50 Female 21	62	1274	13
HMR	18	Male 14 Female 4	69	1265	4
CHUM	55	Male 41 Female 14	64	1120	7
CHUV	40	Male 35 Female 5	63	705	7

FluoroDeoxyGlucose-PET and Computed Tomography (FDG-PET/CT) imaging and manual GTVt contours are available for each patients. The latter (*i.e.* *manual*) were cleaned from initial VOIs acquired in the context of pre-treatment radiotherapy planning, *i.e.* re-annotated to specifically delineate the true tumoral

volume. These manual GTVt are used both for the extraction of features in the *manual* setting and to train a 3D UNet that will generate *automatic* contours. The latter are generated for the entire dataset using a repeated 5-fold Cross-Validation (CV) scheme.

2.2 UNet Architecture and Training

In this section, we describe the deep learning model used to obtain *automatic* GTVt contours and *deep features*. We use a multi-modal 3D UNet developed in [11]. The input images are first cropped to $144 \times 144 \times 144$ voxels using the bounding boxes provided in the HECKTOR challenge [12]. The images are then resampled to 1mm^3 with linear interpolation. The PET images are standardized to zero mean, unit variance. The CT images clipped to [-1024,1024] and linearly mapped to [-1,1]. Data augmentation is applied to the training data including random shifts (maximum 30 voxels), sagittal mirroring and rotations (maximum 5°). The model is trained with an Adam optimizer with cosine decay learning rate (initial learning rate 10^{-3}) to minimize a soft Dice loss. The model is trained for 600 epochs on the entire HECKTOR 2020 data (see Section 2.1). For each of the five folds of the repeated CV, the training set is split into 80% training and 20% validation. The best model according to the validation loss is used for obtaining both *automatic* contours and *deep features* on the test data. Since we do not have enough data to train the model on cases that are not present for the subsequent prognostic prediction study, we use the repeated 5-fold CV to mimic this scenario, ensuring that the *automatic* contours and *deep features* are all obtained on held-out data.

2.3 Radiomics Workflow

In order to evaluate and compare the DFS prognosis performance when extracting HC radiomics features from either the *manual* or *automatic* GTVt, as well as to evaluate the performance of the *deep features*, we develop a standard radiomics workflow, as detailed in Fig. 1.

We first pre-process the images (CT and PET) with an iso-resampling of $2 \times 2 \times 2$ mm using linear interpolation. This step is performed before feature extraction except for the shape features, for which we extract only from the CT with its initial resolution.

Then, we extract features from CT and PET images within either *manual* or *automatic* GTVt. Table 2 details the parameters and the types of HC features used. A total of 130 HC features are extracted per modality with additional 14 shape-based features. For each patient and for each annotation type, we therefore have 274 features as detailed in the following[4]. From those two modalities per patient (CT and PET), we extract features from the first-order family (18 features) and second-order family (56 features). Regarding the second-order family,

[4] We can unconventionally detail the number of HC features as follows: 274 = 2 modalities × (2 bin × 56 2^{nd}order + 18 1^{st}order) + 14 shape (see Table 2).

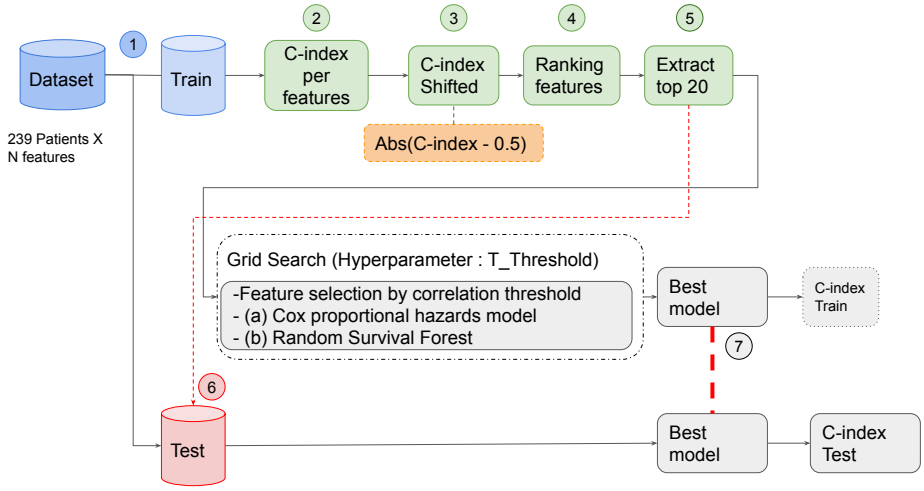

Fig. 1. Radiomics workflow inside one fold of the 5-fold CV and 1 repetition. The univariate and multivariate steps are highlighted in green and grey, respectively.

we extract the 56 features using two different discretization parameters based on Fixed Bin Number (FBN) and Fixed Bin Size (FBS) (see Table 2). Those 56 features are divided into three sub-families: Grey Level Co-occurence Matrix (GLCM), Grey Level Run Length Matrix (GLRLM) and Grey Level Size Zone Matrix (GLSZM). Finally, we extract 14 shape features.

Table 2. List of the different combinations of parameters and types for HC radiomics features. FBN: Fixed Bin Number and FBS: Fixed Bin Size.

Image	Preprocessing	settings	Features
CT	Iso-resampling 2x2x2mm Linear interpolation	FBN = 64 FBS = 50	GLCM (24 features) GLRLM (16 features) GLSZM (16 features)
			First Order (18 features)
			Shape (14 features)
PET	Iso-resampling 2x2x2mm Linear interpolation	FBN = 8 FBS = 1	GLCM (24 features) GLRLM (16 features) GLSZM (16 features)
			First Order (18 features)

Regarding the *deep features*, we extract a total of 256 ReLU-activated feature maps of size $9 \times 9 \times 9$ at the UNet bottleneck, which are spatially averaged to

obtain a collection of 256 scalar-valued features. The latter were directly used in the CPH or RSF model, either alone or concatenated with the HC features.

After extracting features from PET and CT images with either the *manual* or the *automatic* GTVt, we use those features in the pipeline detailed in Fig. 1. We separate the training and the test set using a repeated 5-fold CV method with the same folds as those used for the 3D UNet (see Section 2.2). Based on the training dataset, we compute the univariate Concordance (C) index [13] of each feature. We shift this value ($|C_{\text{index}} - 0.5)|$) to take into account both concordant and anti-concordant features. The resulting top 20 features are kept. We further remove correlated features using a Pearson product-moment correlation threshold $t \in [0.6, 0.65, 0.70, 0.75, 0.80]$, optimized on the validation set using grid-search. This feature selection approach is commonly used in the radiomics community [14, 15] to avoid using highly correlated features as an input to either the CPH model [16] or the RSF [17] model. The CPH and RSF are trained to predict the hazard score and we further compute the average C-index to estimate the performance of DFS estimation across the repeated 5-fold CV. We perform those experiments using Python 3.9 the following libraries: SimpleITK [18] (V2.0.2), PyRadiomics [19] (V3.0.1), scikit-learn [20] (V0.22.2) and scikit-survival [21] (V0.12.0).

3 Results

3.1 GTVt Segmentation

The estimated average Dice Score Coefficient (DSC) on the test sets of the repeated CV is 0.737±0.009. It is worth noting that the model yielded a tumor contour for all patients in the dataset, *i.e.* at least one automatic GTV segmentation for each input. Fig. 2 shows a qualitative result of the 3D UNet segmentation (*automatic* contour) compared with the *manual* contours.

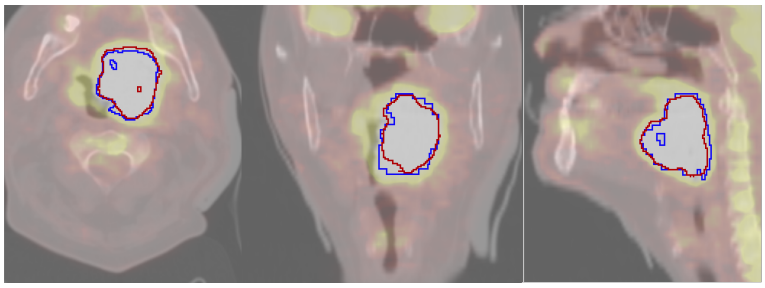

Fig. 2. Example of a *manual* GTVt in blue and the UNet-generated *automatic* GTVt in red on a fused PET/CT image.

3.2 Prognosis Prediction

The DFS prediction performances are reported in Table 3 for the various approaches considered. For a baseline comparison, the performance of simple tumor volume-based DFS prediction is associated with a C-index of 0.6 for both manual and automatic GTVt.

Table 3. Comparison of prognosis performance (DFS) for the various approaches considered using either the CPH or the RSF model. We report the average C-index ± standard-error across 5 folds and 10 repetitions of the CV. The best overall performance is highlighted in bold and the best performance based on automatic segmentation is shown in italic.

	CPH	RSF
HC from *manual* GTVt	**0.672** ± 0.085	0.645 ± 0.096
HC from *Automatic* GTVt	0.623 ± 0.111	0.623 ± 0.095
deep features	0.560 ± 0.114	0.566 ± 0.108
deep ft. + HC from *manual* GTVt	0.648 ± 0.094	0.639 ± 0.096
deep ft. + HC from *Automatic* GTVt	0.606 ± 0.027	***0.626*** ± 0.028

4 Discussion and Conclusions

This work proposes a fully automatic processing pipeline for predicting patient prognosis from PET/CT in the context of H&N cancer. It is based on a 3D UNet providing automatic delineations of GTVt and deep features, as well as additional HC features extracted from the generated GTVt. Deep and HC features are compared and combined when used either by a CPH or RSF model predicting patient hazard for DFS. This fully automatic pipeline is compared to a classical approach based on manually delineated GTVt regions.

While being a secondary objective of our approach, the segmentation performance of the 3D UNet was found to be consistent with the top-performing models of the HECKTOR 2020 challenge [6]. While being close to the state of the art in the field, we will investigate ways to improve the segmentation performance in future work, which is expected to have a important impact on the DFS prediction performance.

The performance of prognostic prediction is reported in Section 3.2. According to Table 3, the best performance for DFS prediction is obtained with HC features from *manual* GTVt contours with a CPH model, which is associated with an average C-index across the repeated 5 folds CV of 0.672. Using the same HC features from *automatic* GTVt contours, results in a lower performance associated with C-indices of 0.623 and 0.623 using CPH or RSF, respectively. When used alone, the *deep features* obtain the worst prognosis performance with a C-index of either 0.560 (CPH) or 0.566 (RSF). Combining the deep features with

HC extracted from *manual* GTVt seems to deteriorate the performance when compared to using HC features only in the *manual* VOIs (C-index of 0.648 versus 0.672 when based on CPH). By comparison, Andrearczyk *et. al* [22] developed a fully end-to-end model (segmentation and prediction) which achieved a good performance of 0.723 (C-index). Nevertheless, we can not directly compare this result with ours due to the difference in the validation splits. However, combining the *deep features* with HC extracted from *automatic* GTVt in a RSF model achieves a promising C-index of 0.626, where the rate of change is only 7.3% lower than the best model based on *manual* contours (CPH and HC features only) while being fully automatic. The statistical significance of this result (HC from *manual* GTVt versus deep ft. + HC from *automatic* GTVt) was assessed using a corrected r-repeated k-fold CV t-test as defined in Bouckaert *et. al* [23] and is associated with a p-value of 0.20. This correction is needed to adapt to underestimated variance related to dependent observations in repeated CVs. It is worth noting that using the standard uncorrected student's t-test yields a p-value of 0.0026. We feel that reporting the latter is important as many studies in the field are not using corrections when breaking the independance assumption of the t-test. Therefore, the proposed fully automatic prognostic model based on the combination of *deep features* with HC extracted from *automatic* GTVt with RSF model seems most relevant to conduct large-scale radiomics studies, which are crucial to further validate the clinical value of such image-based biomarkers. For inferring the prognosis of one patient, the computational time of the fully automatic pipeline (*i.e.* segmentation and *deep feature* extraction, HC features computation and prognosis prediction), is approximately 3 minutes (assuming that the 3D UNet is already trained). Moreover, the proposed approach allows scalable and reproducible contouring, as opposed to manual contouring where inter-observer agreement was reported to be low [24, 6]. We remind the reader that we used a dataset originating from five distinct centers and including various scanner manufacturers and image acquisition protocols.

One limitation of our study is that the *deep features* are pooled by averaging the entire feature maps at the bottleneck of the UNet, which may explain the poor performance given by those set of features either alone or concatenated with the HC features. Future work will investigate feature pooling by averaging inside the VOIs (*e.g. automatic or manual*). Other aggregation methods will be investigated also for the HC features, as it was shown to have a strong influence on the prognosis predictive performance [25]. In addition, we plan to include readily available clinical patient data (*e.g.* age, gender, smoking status, tumor site, human papilloma virus status) [26] to further improve the prognosis performance of the models.

Acknowledgements

This work was partially supported by the Swiss National Science Foundation (SNSF, grant 205320_179069), the Swiss Personalized Health Network (SPHN

via the IMAGINE and QA4IQI projects), and the Hasler Foundation (via the EPICS project, grant 20004).

References

1. Robert J Gillies, Paul E Kinahan, and Hedvig Hricak. Radiomics: images are more than pictures, they are data. *Radiology*, 278(2):563–577, 2016.
2. Alex Zwanenburg, Martin Vallières, Mahmoud A Abdalah, Hugo JWL Aerts, Vincent Andrearczyk, Aditya Apte, Saeed Ashrafinia, Spyridon Bakas, Roelof J Beukinga, Ronald Boellaard, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*, 295(2):328–338, 2020.
3. Martin Vallieres, Emily Kay-Rivest, Léo Jean Perrin, Xavier Liem, Christophe Furstoss, Hugo JWL Aerts, Nader Khaouam, Phuc Felix Nguyen-Tan, Chang-Shu Wang, Khalil Sultanem, et al. Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer. *Scientific reports*, 7(1):1–14, 2017.
4. Vincent Andrearczyk, Valentin Oreiller, Martin Vallières, Joel Castelli, Hesham Elhalawani, Mario Jreige, Sarah Boughdad, John O. Prior, and Adrien Depeursinge. Automatic segmentation of head and neck tumors and nodal metastases in PET-CT scans. In *International Conference on Medical Imaging with Deep Learning (MIDL)*, 2020.
5. Ivayla Apostolova, Ingo G Steffen, Florian Wedel, Alexandr Lougovski, Simone Marnitz, Thorsten Derlin, Holger Amthauer, Ralph Buchert, Frank Hofheinz, and Winfried Brenner. Asphericity of pretherapeutic tumour FDG uptake provides independent prognostic value in head-and-neck cancer. *European radiology*, 24(9):2077–2087, 2014.
6. Vincent Andrearczyk, Valentin Oreiller, Mario Jreige, Martin Vallières, Joel Castelli, Hesham Elhalawani, Sarah Boughdad, John O. Prior, and Adrien Depeursinge. Overview of the HECKTOR challenge at MICCAI 2020: Automatic head and neck tumor segmentation in PET/CT. In *Lecture Notes in Computer Science (LNCS) Challenges*, 2021.
7. Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015.
8. Bjoern H. Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Transactions on Medical Imaging*, 34(10):1993–2024, 2014.
9. Mohammad Havaei, Axel Davy, David Warde-Farley, Antoine Biard, Aaron Courville, Yoshua Bengio, Chris Pal, Pierre Marc Jodoin, and Hugo Larochelle. Brain tumor segmentation with Deep Neural Networks. *Medical Image Analysis*, 35:18–31, 2017.
10. Ujjwal Baid, Sanjay Talbar, Swapnil Rane, Sudeep Gupta, Meenakshi H Thakur, Aliasgar Moiyadi, Siddhesh Thakur, and Abhishek Mahajan. Deep learning radiomics algorithm for gliomas (drag) model: a novel approach using 3d unet based deep convolutional neural network for predicting survival in gliomas. In *International MICCAI Brainlesion Workshop*, pages 369–379. Springer, 2018.
11. Fabian Isensee, Philipp Kickingereder, Wolfgang Wick, Martin Bendszus, and Klaus H Maier-Hein. Brain tumor segmentation and radiomics survival prediction: Contribution to the BRATS 2017 challenge. In *International MICCAI Brainlesion Workshop*, pages 287–297. Springer, 2017.

12. Vincent Andrearczyk, Valentin Oreiller, and Adrien Depeursinge. Oropharynx detection in PET-CT for tumor segmentation. In *Irish Machine Vision and Image Processing*, 2020.
13. Frank E Harrell Jr, Kerry L Lee, and Daniel B Mark. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in medicine*, 15(4):361–387, 1996.
14. Philippe Lambin, Ralph TH Leijenaar, Timo M Deist, Jurgen Peerlings, Evelyn EC De Jong, Janita Van Timmeren, Sebastian Sanduleanu, Ruben THM Larue, Aniek JG Even, Arthur Jochems, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nature reviews Clinical oncology*, 14(12):749–762, 2017.
15. Yannick Suter, Urspeter Knecht, Mariana Alão, Waldo Valenzuela, Ekkehard Hewer, Philippe Schucht, Roland Wiest, and Mauricio Reyes. Radiomics for glioblastoma survival analysis in pre-operative MRI: exploring feature robustness, class boundaries, and machine learning techniques. *Cancer Imaging*, 20(1):1–13, 2020.
16. Cox R David et al. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society*, 34(2):187–220, 1972.
17. Hemant Ishwaran, Udaya B Kogalur, Eugene H Blackstone, Michael S Lauer, et al. Random survival forests. *Annals of Applied Statistics*, 2(3):841–860, 2008.
18. Bradley Christopher Lowekamp, David T Chen, Luis Ibáñez, and Daniel Blezek. The design of simpleitk. *Frontiers in neuroinformatics*, 7:45, 2013.
19. Joost JM Van Griethuysen, Andriy Fedorov, Chintan Parmar, Ahmed Hosny, Nicole Aucoin, Vivek Narayan, Regina GH Beets-Tan, Jean-Christophe Fillion-Robin, Steve Pieper, and Hugo JWL Aerts. Computational radiomics system to decode the radiographic phenotype. *Cancer research*, 77(21):e104–e107, 2017.
20. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
21. Sebastian Pölsterl. scikit-survival: A library for time-to-event analysis built on top of scikit-learn. *Journal of Machine Learning Research*, 21(212):1–6, 2020.
22. Vincent Andrearczyk, Pierre Fontaine, Valentin Oreiller, and Adrien Depeursinge. Multi-task deep segmentation and radiomics for automatic prognosis in head and neck cancer. In *Workshop on PRedictive Intelligence in MEdicine (PRIME) at MICCAI*.
23. Remco R Bouckaert and Eibe Frank. Evaluating the replicability of significance tests for comparing learning algorithms. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 3–12. Springer, 2004.
24. Hilke Vorwerk, Gabriele Beckmann, Michael Bremer, Maria Degen, Barbara Dietl, Rainer Fietkau, Tammo Gsänger, Robert Michael Hermann, Markus Karl Alfred Herrmann, Ulrike Höller, et al. The delineation of target volumes for radiotherapy of lung cancer patients. *Radiotherapy and Oncology*, 91(3):455–460, 2009.
25. Pierre Fontaine, Oscar Acosta, Joël Castelli, Renaud De Crevoisier, Henning Müller, and Adrien Depeursinge. The importance of feature aggregation in radiomics: a head and neck cancer study. *Scientific Reports*, 10(1):1–11, 2020.
26. Tian-Tian Zhai, Lisanne V van Dijk, Bao-Tian Huang, Zhi-Xiong Lin, Cássia O Ribeiro, Charlotte L Brouwer, Sjoukje F Oosting, Gyorgy B Halmos, Max JH Witjes, Johannes A Langendijk, et al. Improving the prediction of overall survival for head and neck cancer patients using image biomarkers in combination with clinical parameters. *Radiotherapy and Oncology*, 124(2):256–262, 2017.