Development and prospective validation of a spatial dose pattern based model predicting acute pulmonary toxicity in patients treated with volumetric arc-therapy for locally advanced lung cancer

Vincent Bourbonne, François Lucia, Vincent Jaouen, Julien Bert, Martin Rehn, Olivier Pradier, Dimitris Visvikis, Ulrike Schick

Development and prospective validation of a spatial dose pattern based model predicting acute pulmonary toxicity in patients treated with volumetric arc-therapy for locally advanced lung cancer.

Bourbonne V.[1,2], Lucia F.[1,2], Jaouen V.[2,3], Bert J.[2], Rehn M.[1], Pradier O.[1,2], Visvikis D.[2], Schick U.[1,2]

1. Department of Radiation Oncology, University Hospital, Brest, France
2. LaTIM UMR 1101 INSERM, University Brest, Brest, France
3. Institut Mines-Télécom Atlantique, Brest, France

Corresponding author:

Dr Vincent Bourbonne, MD, MSc

Radiation Oncology Department

CHRU Brest

2 avenue Foch, 29200 Brest, France

Mail : vincent.bourbonne@chu-brest.fr

Tel: +33298223398

Fax: +33298223087

Funding: none

Conflict of interest: none

Short title: Validation of a spatial dose pattern based model for radiation pneumonitis prediction.

Keywords: radiation pneumonitis, lung cancer, prediction, cluster of voxel, adaptive dosimetry

Statistical analysis: Dr Vincent Bourbonne, MD, MSc, Radiation Oncology Department, University Hospital Brest, 2 avenue Foch, 29200 Brest, France. Mail : vincent.bourbonne@chu-brest.fr

Wordcount of the text: ~ 370 words
Wordcount of the abstract: ~ 3270 words

Highlights:

- Despite strict application of dose constraints, acute toxicities such as acute pulmonary toxicity (APT) remain frequent, and may impact treatment's compliance and patients' quality of life.
- Using a voxel-based approach, a cluster of voxels (Pmap-region) predictive of APT was identified in the posterior right lung.
- Mean dose to the Pmap-region contributes to a significant improvement in acute pulmonary toxicities prediction, opening the possibility of easily implementable adaptive dosimetry planning.

Introduction: (Chemo)-radiotherapy is the standard treatment for patients with locally advanced lung cancer (LALC) not accessible to surgery. Despite strict application of dose constraints, acute toxicities such as acute pulmonary toxicity (APT) remain frequent, and may impact treatment's compliance and patients' quality of life. Previously, on a population treated with intensity-modulated photon therapy or passive scattering proton therapy, spatial dose patterns associated with APT were identified in the lower lungs, especially in the posterior right lung. In the present study, we aim to define these spatial dose patterns on a retrospective cohort treated by volumetric-arctherapy (VMAT) and to validate our findings prospectively.

Methods: For the training cohort, we retrospectively included all patients treated in our institution by VMAT for a LALC between 2015 and 2018. APT was scored according to the CTCAE v4.0 scale. All dose maps were registered to a thorax phantom using a segmentation-based elastic registration. Voxel-based analysis of local dose differences was performed with a non-parametric permutation test accounting for $n = 10.000$ permutations, producing a 3-dimensional significance maps on which clusters of voxels that exhibited significant dose differences ($p < 0.05$) between the two toxicity groups (APT $\geq$ grade 2 vs APT $<$ grade 2) were identified. A prediction model (Pmap-Model) was then built using a neural network approach and then applied to an observational prospective cohort for validation. The model was evaluated using the Area under the curve (AUC) and the balanced accuracy (Bacc: mean of the sensitivity and specificity).

Results: 165 and 42 patients were included in the training and validation cohorts, with respective APT rates of 22.4% and 19.1%. In the training cohort, a cluster of voxels (Pmap-region) was identified in the posterior right lung. In the training cohort, the Pmap-Model combining 11 features among which the mean dose to the Pmap-region resulted in an AUC of 0.99 and a Bacc of 99.2 using an 8% probability threshold. Using the same voxel cluster on the validation cohort, the Pmap-model resulted in an AUC of 0.81 and a Bacc of 82.0.

Conclusion: Our APT-prediction model was successfully validated in a prospective cohort treated by VMAT. Regional radiosensitivity should be considered in usual lung dose constraints, opening the possibility of easily implementable adaptive dosimetry planning.

Despite a decrease in its main risk factors, lung cancer remains the leading cause of death by cancer in western countries such as the United States[1] and Europe[2] with approximately 130.000 and 260.000 new deaths/year in 2020, respectively. For locally advanced lung cancers (LALC) not accessible to surgery, (chemo)-radiotherapy (Ch-RT) is the treatment of reference[3].

While adjuvant immunotherapy (durvalumab) leads to a significant benefit in progression free survival[4] (PFS) and overall survival[5] (OS), toxicities of such combination of treatments are frequent. Substantial advances have been made in radiation therapy in order to further respect dose constraints based on dose-volume histograms (DVHs). Among these techniques, image-guided radiotherapy (IGRT), intensity-modulated radiation therapy (IMRT) and volumetric arc-therapy (VMAT) allow to better spare organs at risk (OAR) such as the lungs without compromising local control[6,7]. Despite these improvements, acute pulmonary toxicity (APT) remains frequent with an approximate rate of 5-25% for grade ≥ 2 APT[8-10] in patients treated by Ch-RT. In patients treated by the association of Ch-RT and adjuvant immunotherapy (durvalumab)[4], the APT rate of any grade increases to 33.9% compared to 24.8% for the Ch-RT group only, with a moderate increase in grade ≥3 APT (3.4% vs 2.6%). Such toxicities can impair the completion of radiotherapy (RT) or adjuvant immunotherapy (IT). For example, the discontinuation rate due to adverse events in the initial PACIFIC trial was higher for the adjuvant IT group (15.4%) compared to the Ch-RT alone group (9.8%).

Previous efforts have been made to better stratify patients based on their risk of APT. Among common approaches, non-tumor complication probability is the most commonly used and is based on dose-volume histograms (DVHs). But, despite the respect of dose constraints, APT remains frequent. Previously, we successfully increased this DVH-based prediction of APT risk using an innovative radiomics-based model to analyse dose heterogeneity to OARs in a retrospective cohort[11]. However, our previous model did not account for topographical radiation sensitivity disparities insides OAR. Indeed, sub-regions sensitive to radiation in the heart and the lungs and especially the lower right lung have already been successfully identified through an analysis based on cluster of voxels[12]. However, this method was never externally nor prospectively validated. As such, no prediction model incorporating such sub-regions was developed.

In this study, we aim to identify anatomical sub-regions in the lungs and heart correlated with the risk of APT in a population treated by VMAT for a LALC, and to further build and prospectively validate a predictive model of APT.

Material and Methods

Retrospective cohort: all patients treated by VMAT with a curative intent for a histologically-proven locally advanced lung cancer (non-small cell or small-cell lung cancer) between 2015 and 2018 were retrospectively considered. Patients with an age > 18 and a minimum follow-up of 1 year after RT completion were included. Patients with an history of thoracic radiation therapy, pneumonectomy or incomplete RT were excluded. When performed, chemotherapy could be delivered as a sequential or a concomitant treatment. The study was approved by the hospital ethical committee (NCT04545658).

Prospective cohort: the prospective cohort consisted of the patients included in a prospective observational trial approved by the hospital ethical committee (the TEFRAC trial - NCT03931356). This trial assesses the consequences of low doses of radiation delivered by VMAT on the respiratory capacity of patients treated for bronchopulmonary carcinoma, as assessed by functional respiratory exploration during follow-up. Patients from the TEFRARC

trial who met the previously detailed inclusion and exclusion criteria were included in the prospective cohort.

RT was delivered with a prescription dose of 60 to 66 Gy to the planning target volume (PTV), with 95% of the dose covering 95% of the prescription volume. At physician's discretion, lower RT doses (minimum of 54 Gy) were allowed. Dosimetric planning was performed using Pinnacle v16.2 (Phillips, Eindhoven, Netherlands) with the objective of 95% of the prescribed dose (60-66Gy) covering 95% of the target volumes, with a 2Gy by fraction regimen. VMAT plans were designed with two half arcs for peripheral target volumes and two full arcs for central target volumes. An energy of 6 Megavolts was used. The maximum dose point was limited to 110% of the prescribed dose.
For the analysis, doses were converted to biologically equivalent doses (BED) using the dose map conversion tool in MIM Maestro (MIM v7.0.0, Cleveland, OH, USA). Alpha/beta ratios of 3 for non-tumour volumes and of 10 for tumour volumes were used.

Acute pulmonary toxicity was scored using the CTCAE v4.0[13]: APT was defined as a binary outcome corresponding to a grade ≥ 2 pulmonary toxicity event occurring during the first 6 months following start of RT. APT was evaluated each week during radiation therapy, at the end of radiotherapy and then one month, 3 months and 6 months after the completion of radiation therapy. Evaluation of toxicities was similarly performed by the doctor in both the training and validation cohorts and for all patients. For the retrospective cohort, APT was retrospectively collected for all patients. Regarding the prospective cohort, APT was prospectively evaluated and collected as a secondary outcome. Actuarial incidences were used.

All dose maps were registered to a thoracic phantom using a segmentation-based elastic registration via MIM Maestro (MIM v7.0.0, Cleveland, OH, USA). The segmentation used for registration was a hybrid volume of interest (VOI) combining the lungs and the heart. In order to evaluate the quality of the registration, the DICE coefficient between the phantom's segmentation and each patient's segmentation was calculated. Correlation between each DICE score and the model's results (correct vs incorrect classification) was analyzed using the Spearman's correlation coefficient in order to assess the possible impact of the registration on the model's performance. For the definition on the radiation-sensitive volume, we applied the previously used methodology developed by Palma *et al*[12,14]. Voxel-based analysis of local dose differences was performed with a non-parametric permutation test accounting for $n = 10.000$ permutations, producing a 3-dimensional significance maps (Pmap-region) on which clusters of voxels that exhibited significant dose differences ($p < 0.05$) between the two toxicity groups (APT $\geq$ grade 2 vs APT $<$ grade 2) were identified. Softwares used for the voxel-based analysis were FSL (FSL v6.0.4) for the general linear matrix creation, SPM12 and Palm Matlab tools (MATLAB_R2014b). Correction for multiple testing error was insured thanks to the FWER (Family Wise Error Rate) function embedded in the Palm tool.
Definition of the significant cluster of voxels was performed on the retrospective cohort. It was then applied to each patient from the prospective cohort.

Usual clinical features such as age, gender, performance status (PS), administration of chemotherapy (Ch), tumour stage according to the American Joint Committee on Cancer (AJCC), mean expiratory volume/second (MEVS), history of smoking, chronic obstructive pulmonary disease (COPD) were considered. Regarding RT, total delivered dose, gross (Vol_{GTV}), clinical (Vol_{CTV}) and planning (Vol_{PTV}) tumour volumes and dosimetric factors extracted from DVHs[15,16] were included, according to current recommendations[17,18]. DVHs

were calculated from the delivered RTPlan using the Pinnacle treatment planning system. Vx (Gy) will further be defined as the percentage of the VOI receiving at least x dose (Gy). DMean and Dmax correspond to the mean and maximal dose received by the VOI, respectively. On the homolateral (LungH) and controlateral (LungC) lungs, V5, V10, V13, V20, and V30, DMax and DMean were collected. Regarding both lungs (2Lungs) VOI, V13, V20, V30 and DMean were considered. Finally, regarding the heart, V30, V40 and DMean were extracted. DMean from the identified radiation sensitive sub-region ($DMean_{Pmap}$) was also analyzed. As a result, a full set of 37 features was available for each patient.
Univariate analysis for each feature was performed using the Receiver Operative Characteristics (ROC) and the Area Under the Curve (AUC). The optimal cut-off was defined as the one maximizing the Youden Index (YI: sensitivity (Se) + specificity (Sp) – 1).
For descriptive analysis, continuous variables were reported with mean, standard deviation, median, confidence intervals (CI) and quartiles. Categorical variables were instead reported with frequencies and percentages. The significance tests used to estimate the inference between the two groups were the T-test for independent variables and the non-parametric Mann-Whitney U test otherwise.

Correlation between the model's results and the DICE coefficient between the phantom's segmentation and each patient's segmentation was analyzed using the Spearman's correlation coefficient.

All features were processed for classification and toxicity modeling via a supervised neural-network (NN) approach relying on the NN library in SPSS Modeler V18.0® (IBM, NY, USA). This tool is based on a multilayer perceptron architecture, with a feedforward approach. Optimization stopped when a minimum precision of 90.0% was reached, the other parameters being set as : an initial lambda of 5.10-7, an initial sigma of 5.10-5, an interval center of 0 with an interval offset of +/- 0.5.
Due to the relatively small cohort, we replaced the classical training/testing/validation split with a training/validation split and cross-validation being realized on the training set. In order to prevent or reduce the risk of overfitting, we used a technique mainly based on bootstrapping with n = 1000 replications. Firstly, the overall population was thus split into a training (retrospective set) and a validation (prospective set) cohort, the validation cohort being used for the evaluation of the final model's performance. Once split, the model combined the provided features, with as a result a risk probability for each patient of the training cohort and a ranking by importance for each provided features, this ranking possibly being different from the correlation (AUC).

Regarding the number of features, 37 features were available before optimization and features' selection. Acknowledging the high risk of overfitting, we limited the number of retained features by iteratively developing several models based on each features' rank. 37 models were thus built; with at each step the least important feature being put apart and the remaining features being provided for the development of the next model. The final model was chosen among these 37 models, as the one with the highest mean accuracy based on the 1000 replications of the boostraping and then evaluated on the validation cohort for the prediction of grade ≥ 2 APT. For completeness, the pre-developed model was also evaluated for the prediction of grade ≥ 3 APT. A sub-set of patients at very low risk of APT was also identified (threshold probability $\leq 2\%$).
Finally, in order to evaluate the prognostic value of the laterality, separate models were developed without providing the $DMean_{Pmap}$ but instead a variable corresponding to the lateralization of the treatment.

Performances of the model was then analyzed using Se, Sp, balanced accuracy (Bacc: mean of the Se and Sp). Decision curves for the training and validation sets were drawn plotting the net benefit with its corresponding threshold. Decision curve analysis calculates a clinical net benefit for the prediction model in comparison to the default strategies of considering all patients at high (or low) risks of APT. A net benefit is calculated across the full range of threshold probabilities and gives the reader a view of the impact of the model and the selected threshold on the clinical benefit[19]. The reference corresponds to the net clinical benefit with all patients considered at high risk of APT.

Results

In the training cohort, 167 patients were considered as eligible with 2 excluded because of an history of pneumonectomy, resulting in 165 included patients with a mean age of 65.0. Overall, 22.4% of patients experienced APT ≥ grade 2 and 81.8% were treated with chemotherapy: 29.1% with neoadjuvant chemotherapy, 32.1% with concomitant chemotherapy and 20.6% with neoadjuvant followed by concomitant chemotherapy.
Forty-two patients were included in the validation cohort, with a mean age of 65.7 and only 21.4% of patients did not receive any chemotherapy. The rate of APT was 19.1% in the validation cohort.
Patients' characteristics according to the occurrence of APT in each cohort are summarized in Table 1.

The mean DICE coefficient between the phantom's segmentation and each patient's segmentation was 0.98 (range 0.98-0.99) and 0.98 (range 0.96-0.99) for the training and validation sets, respectively.
With n = 10.000 permutations, voxels harboring a significance difference between the patients with APT and those without APT were identified in the the posterior lower and upper right lung defining the Pmap-region. A 3D-example of this mask on the phantom is available as Figure 1. The mean dose received by this cluster ($DMean_{Pmap}$) was significantly different between the non-APT and the APT group: 15.1Gy versus 32.1 Gy ($p = 0.0003$).

Out of the 37 evaluated features, only 7 features were significantly correlated with the risk of APT based on the univariate analysis: the $DMean_{Pmap}$, the $DMean_{2Lungs}$, the $DMean_{LungH}$, the Vol_{GTV}, Vol_{CTV} and Vol_{PTV}. After the $DMean_{Pmap}$, the second highest AUC was reached for the Vol_{PTV} (0.62). For the $DMean_{Pmap}$, the Youden Index was maximized with a cut-off of 30.3 Gy (Se 54.1% and Sp 88.3%) with consistent AUCs and Baccs in the training (0.69/71.2%) and validation (0.77/72.5%) cohorts. Full set of individual features and detailed results (AUC) for each feature are available as Supplementary Table 1.
In the training cohort, mean $DMean_{Pmap}$ was significantly different between the two APT groups: 17.7Gy in the non-APT group and 28.3Gy in the APT-group ($p=0.0003$). Similar results were found in the validation cohort with a mean $DMean_{Pmap}$ of 19.4Gy for the non-APT group and 31.9Gy for the APT-group ($p=0.02$). In the validation cohort, patients with a $DMean_{Pmap}$ superior to 30.3Gy were 4.7 times more likely to present with an APT ≥ grade 2.

Thirty-seven models were iteratively developed on the training cohort. The model's mean accuracy ranged from 0.72 to 0.88, the highest accuracy being reached for the model N°26 combining 11 features, described in Table 2. Accounting for more than 50% of the model's predictions the $DMean_{Pmap}$, the $DMean_{2Lungs}$ and the $V30_{2Lungs}$ were the 3 most important features. From now on, the model N°26 will be named the Pmap-Model. Representation of each model's performance is available as Figure 2a (AUC) and Figure 2b (C-Index) for the training and validation cohorts. Full composition of the Pmap-Model can be found as Table 2.

The Pmap-model resulted in an AUC of 0.99 and a BAcc of 99.2% when applying an 8% threshold. From a clinical perspective, 94.9% and 100% of the patients were correctly classified at high and low risk of APT meaning that only 2 patients were misclassified at high-risk and none at low-risk.
On the validation cohort, the Pmap-model resulted in an AUC of 0.81 and a BAcc of 82.0%, with 46.7% correctly classified at high risk while 8 patients being misclassified at high risk of APT despite no APT occurred. Only one patient (3.7%) was misclassified at low-risk. Overall, 98.8% and 78.6% of patients were correctly classified in the training and validation sets. Crossed tables in each cohort are presented as Table 3. ROC curves for the training and validation cohorts are available as Supplementary Figures 1a and 1b, respectively.
Correct classification was independent from the DICE coefficient (Spearman's coefficient of ρ= 0.06, p = 0.41 and ρ=-0.02, p = 0.88 classified in the training and validation sets, respectively).
Decision curve analysis shows a clinically interesting profile for the Pmap-model in both the training and validation sets but only for patients with a risk < 15% (Figure 3). In the validation cohort, 10/42 patients had a risk superior to 15% according to the Pmap-Model, among them 4 experienced an APT ≥ grade 2. Among the 5 patients with a risk between 8 and 15%, 60% (n=3) developed APT ≥ grade 2 while in the 27 remaining patients with a risk < 8%, APT ≥ grade 2 occurred in 1 patient only (3.7%).
In the training cohort, the Pmap-model was significantly associated with the risk of grade ≥ 3 APT : AUC 0.92. Applying the 8.0% probability threshold, the model resulted in a Bacc of 0.91 (Se 100.0%, Sp 80.8%). On the validation cohort, the model achieved a Bacc of 83.8% (Se 100.0%, Sp 67.5%). In the sub-set of patients with a predicted risk ≤2% (n=62 patients), 16 patients (25.8%) presented with an in-field local relapse. A higher radiation dose could be delivered in this selected population and possibly increase the chance of local control.

Regarding the value of laterality and based on the previously presented neural network approach, 37 models were thus newly developed without using the $DMean_{Pmap}$. The best model combined 7 features ($V40_{Heart}$, $V10_{LungH}$, $DMean_{2Lungs}$, Stage, COPD, MEVS and smoking status) and resulted in a mean accuracy of 0.86 in the training cohort. With its corresponding threshold (10.0%), the model achieved an AUC of 0.67, a Se of 100.0% and a Sp of 11.8% (Bacc 55.9%) in the validation cohort. Laterality of the target volumes was considered as the least important feature overall and was not retained in the best model.

Discussion

In this study, we successfully identified a cluster of voxels in the posterior right lung which is significantly correlated with a higher risk of APT. Indeed, in the validation cohort, mean dose to this Pmap-region was significantly higher in patients experiencing grade ≥2 APT (19.4Gy vs 31.9Gy, *p=0.02*).
We applied the voxel-based analysis approach described by Palma *et al*[12]. In that study, based on a population of 178 patients treated with either intensity modulated radiation therapy or passive scattering proton therapy, sub-regions sensitive to RT for the prediction of grade ≥ 2 radiation pneumonitis were identified in the lower lungs (mainly the right lung) and the right side of the heart. While confirming the existence of such a radiation sensitive sub-region, our results differed in two main aspects: Firstly, the Pmap-region was localized in the right lung only with no intersection with the heart or the left lung. The role of the heart in the pathogenesis of APT is debatable with conflicting results to this date[20]. Secondly, Palma *et al* defined the significant region on a single cohort. While delimitating the Pmap-region on a retrospective cohort, we validated our results on an external prospective cohort of 42 patients.
The physiopathology underlying this functional heterogeneity remains unclear. A possibility

could be the vascular structure of the lung (perfusion) but more data are needed to better understand why the posterior right lung appears as particularly sensitive to radiation[21].

Given the relatively high frequency of APT among patients treated with RT and its possible negative consequence on RT completion and adjuvant IT tolerance [4,5], prediction models are necessary.
We could not find any clinical variable significantly predictive of APT, except the GTV, CTV and PTV's volumes[22,23]. On the opposite, after identifying the Pmap-region and proving the significant correlation between the $DMean_{Pmap}$ and the APT risk, we developed a predictive model combining 11 features among which the $DMean_{Pmap}$ accounted for more than a third (36%) of the Pmap-model. Our model successfully classified 98.8% and 78.6% patients of the training and validation cohorts, respectfully. . This method based on the use of the Pmap-region is easily implantable and fastly processed[11]. The Pmap-Region can be implemented as an avoidance region during the radiation planning, and has has also the advantage of relying on tangible and clinically meningful parameters.
The other approach that we previously reported focuses on dose distribution heterogeneity to the OARs and especially the lungs[11]. This model combined 5 radiomic features among which 4 were extracted from the homolateral lung, and achieved high results, with a Bacc of 92.0% in the testing set for APT. It may produce better results, however, it cannot be taken into account before dosimetric planning, and rather appears as a quality analysis tool of the dosimetric planning.

A sub-region was identified in the posterior right lung in the training cohort and further validated on a prospective cohort. Once the radiosensitive region is identified, it becomes possible to adapt the RT planning in order to avoid these regions. Two main approaches have been previously conducted with the aim of a functional dosimetric planning. The first one uses nuclear imaging such as pulmonary ventilation and perfusion plannar scintigraphy or PET/CT[24-26]. IMRT plans adapted to perfused but not ventilated lungs allowed for reduced dose to functional lung without compromising the plan quality. The second one uses four-dimensional computed tomography[27,28]. Despite dose reduction in functional regions, observation of a reduction in the APT risk remains limited because of small cohorts in each study. Both individual functional lung regions as well as our identified Pmap-region could be considered during planning, opening the possibility of easily implementable adaptive dosimetry planning.

Few limitations of our study have to be acknowledged. We partly overcame the retrospective setting by evaluating the model on a prospective cohort. Nevertheless, uncertainties regarding retrospective APT classification persist. All patients having been treated by VMAT-RT, generalizability to other techniques (especially 3D-conformal RT) has yet to be studied. Nevertheless, given the superposability of the Pmap-region with dose patterns identified by Palma et al, one can suppose RT delivery technique has little impact on the Pmap-volume. Additionnaly, asubstantial loss in classification capacity was seen between the training and validation cohorts, suggesting a possible overfitting in the training set despite the boostrap aggregation. Neural networks are often criticized as being "black boxes" [29]. Our approach offers classification by importance of the features, thus partly addressing this issue and providing models with some explainability for the users. Finally, impact of respiratory motion on dosimetric planning on a static CT remains poorly known, whereas motions are supposed to increase in the lower lungs[30].

The proposed approach successfully identified an anatomical sub-region sensitive to radiation for the prediction of APT, this region being localized in the posterior right lung. The Pmap region as well as the Pmap prediction model performed robustly on an external prospective cohort. This anatomical sub-region as well as lung functional sub-volumes should be taken into account for an optimized RT planning. Prospective validation is currently under investigation in a multi-institutional setting.

TABLES

Table 1: Main patients' characteristics in the training and validation sets

Table 2: Classification of each feature included in the Pmap-model by importance and normalized importance

Table 3: Analysis of the Pmap-model's discrimination between patients with or without APT ≥ grade 2 in the training and validation sets

FIGURES

Figure 1: Dose differences between the APT and no APT groups

Figure 2: Results of each decrementally built model in the training set

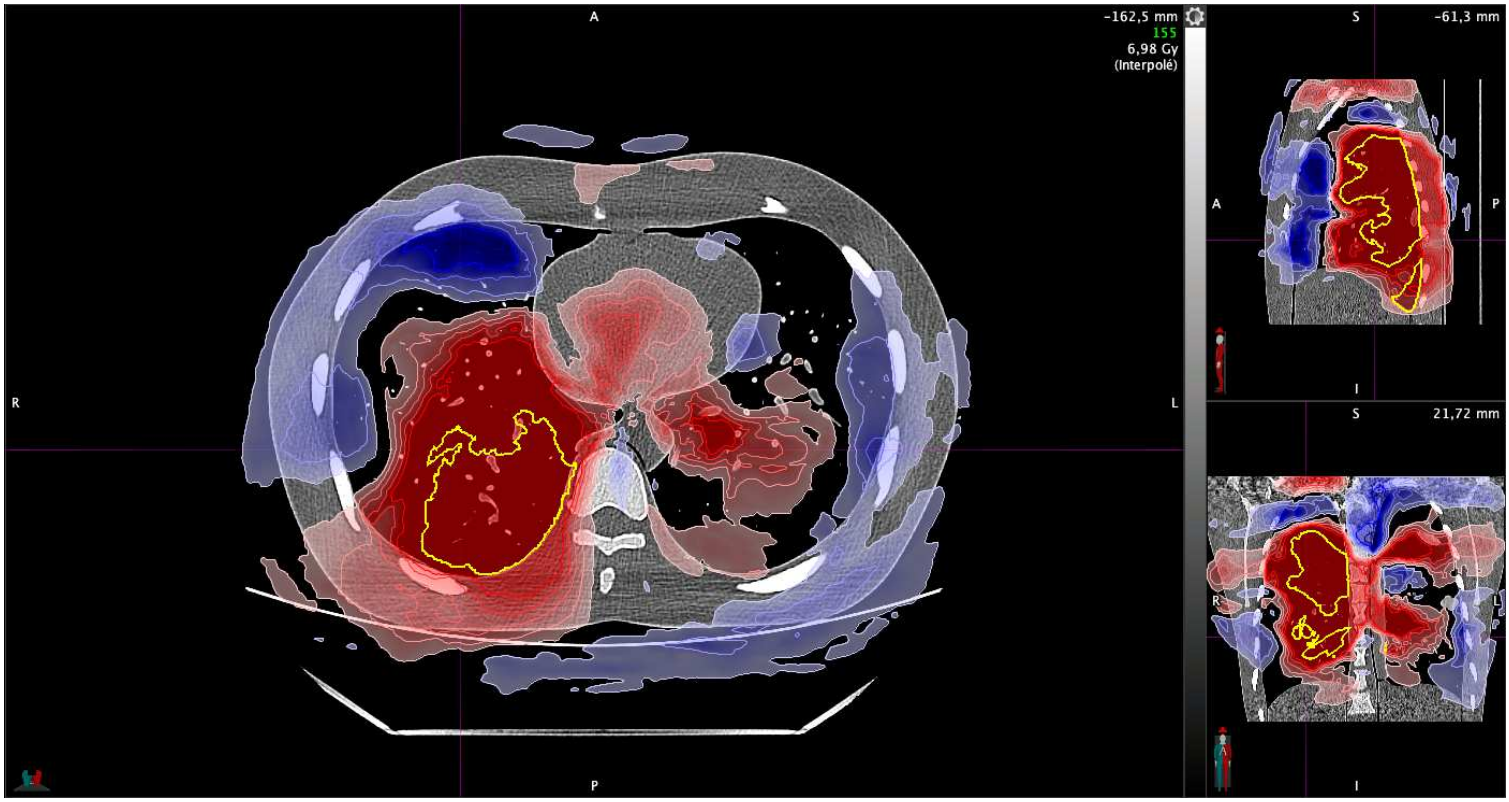
Figure 3: Decision curve analysis in the training (a) and validation (b) sets

REFERENCES

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. CA: a cancer journal for clinicians 2021;71:7-33.
2. ECIS : European Cancer Information System 2020.
3. Ettinger DS, Wood DE, Akerley W, et al. NCCN Guidelines Insights: Non-Small Cell Lung Cancer, Version 4.2016. Journal of the National Comprehensive Cancer Network : JNCCN 2016;14:255-64.
4. Antonia SJ, Villegas A, Daniel D, et al. Durvalumab after Chemoradiotherapy in Stage III Non-Small-Cell Lung Cancer. The New England journal of medicine 2017;377:1919-29.
5. Antonia SJ, Villegas A, Daniel D, et al. Overall Survival with Durvalumab after Chemoradiotherapy in Stage III NSCLC. The New England journal of medicine 2018;379:2342-50.
6. Grills IS, Yan D, Martinez AA, Vicini FA, Wong JW, Kestin LL. Potential for reduced toxicity and dose escalation in the treatment of inoperable non-small-cell lung cancer: a comparison of intensity-modulated radiation therapy (IMRT), 3D conformal radiation, and elective nodal irradiation. International journal of radiation oncology, biology, physics 2003;57:875-90.
7. Christian JA, Bedford JL, Webb S, Brada M. Comparison of inverse-planned three-dimensional conformal radiotherapy and intensity-modulated radiotherapy for non-small-cell lung cancer. International journal of radiation oncology, biology, physics 2007;67:735-41.
8. Khalil AA, Hoffmann L, Moeller DS, Farr KP, Knap MM. New dose constraint reduces radiation-induced fatal pneumonitis in locally advanced non-small cell lung cancer patients treated with intensity-modulated radiotherapy. Acta oncologica 2015;54:1343-9.
9. Wijsman R, Dankers F, Troost EGC, et al. Comparison of toxicity and outcome in advanced stage non-small cell lung cancer patients treated with intensity-modulated (chemo-)radiotherapy using IMRT or VMAT. Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology 2017;122:295-9.
10. Ling DC, Hess CB, Chen AM, Daly ME. Comparison of Toxicity Between Intensity-Modulated Radiotherapy and 3-Dimensional Conformal Radiotherapy for Locally Advanced Non-small-cell Lung Cancer. Clinical lung cancer 2016;17:18-23.
11. Bourbonne V, Da-Ano R, Jaouen V, et al. Radiomics analysis of 3D dose distributions to predict toxicity of radiotherapy for lung cancer. Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology 2021;155:144-50.
12. Palma G, Monti S, Xu T, et al. Spatial Dose Patterns Associated With Radiation Pneumonitis in a Randomized Trial Comparing Intensity-Modulated Photon Therapy With Passive Scattering Proton Therapy for Locally Advanced Non-Small Cell Lung Cancer. International journal of radiation oncology, biology, physics 2019;104:1124-32.
13. National Cancer Institute PROCSG. Common Terminology Criteria for Adverse Events v4.0. Available.
14. Palma G, Monti S, Cella L. Voxel-based analysis in radiation oncology: A methodological cookbook. Physica medica : PM : an international journal devoted to the applications of physics to medicine and biology : official journal of the Italian Association of Biomedical Physics 2020;69:192-204.
15. Marks LB, Bentzen SM, Deasy JO, et al. Radiation dose-volume effects in the lung. International journal of radiation oncology, biology, physics 2010;76:S70-6.

16. Rudra S, Al-Hallaq HA, Feng C, Chmura SJ, Hasan Y. Effect of RTOG breast/chest wall guidelines on dose-volume histogram parameters. Journal of applied clinical medical physics 2014;15:4547.
17. Kim TH, Cho KH, Pyo HR, et al. Dose-volumetric parameters of acute esophageal toxicity in patients with lung cancer treated with three-dimensional conformal radiotherapy. International journal of radiation oncology, biology, physics 2005;62:995-1002.
18. Werner-Wasik M, Yorke E, Deasy J, Nam J, Marks LB. Radiation dose-volume effects in the esophagus. International journal of radiation oncology, biology, physics 2010;76:S86-93.
19. Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. Diagnostic and prognostic research 2019;3:18.
20. Tucker SL, Liao Z, Dinh J, et al. Is there an impact of heart exposure on the incidence of radiation pneumonitis? Analysis of data from a large clinical cohort. Acta oncologica 2014;53:590-6.
21. Johansson A, Vikgren J, Moonen M, Tylen U, Bake B. Regional ventilation and distribution of emphysema - a quantitative comparison. Clinical physiology and functional imaging 2004;24:58-64.
22. Marks LB, Munley MT, Bentel GC, et al. Physical and biological predictors of changes in whole-lung function following thoracic irradiation. International journal of radiation oncology, biology, physics 1997;39:563-70.
23. Yuan ST, Frey KA, Gross MD, et al. Changes in global function and regional ventilation and perfusion on SPECT during the course of radiotherapy in patients with non-small-cell lung cancer. International journal of radiation oncology, biology, physics 2012;82:e631-8.
24. Siva S, Thomas R, Callahan J, et al. High-resolution pulmonary ventilation and perfusion PET/CT allows for functionally adapted intensity modulated radiotherapy in lung cancer. Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology 2015;115:157-62.
25. Bucknell N, Hardcastle N, Jackson P, et al. Single-arm prospective interventional study assessing feasibility of using gallium-68 ventilation and perfusion PET/CT to avoid functional lung in patients with stage III non-small cell lung cancer. BMJ open 2020;10:e042465.
26. Le Roux PY, Hicks RJ, Siva S, Hofman MS. PET/CT Lung Ventilation and Perfusion Scanning using Galligas and Gallium-68-MAA. Seminars in nuclear medicine 2019;49:71-81.
27. Vinogradskiy Y, Schubert L, Diot Q, et al. Regional Lung Function Profiles of Stage I and III Lung Cancer Patients: An Evaluation for Functional Avoidance Radiation Therapy. International journal of radiation oncology, biology, physics 2016;95:1273-80.
28. Yamamoto T, Kabus S, von Berg J, Lorenz C, Keall PJ. Impact of four-dimensional computed tomography pulmonary ventilation imaging-based functional avoidance for lung cancer radiotherapy. International journal of radiation oncology, biology, physics 2011;79:279-88.
29. Clark T, Nyberg E. Creating the Black Box: A Primer on Convolutional Neural Network Use in Image Interpretation. Current problems in diagnostic radiology 2019.
30. Lee S, Stroian G, Kopek N, AlBahhar M, Seuntjens J, El Naqa I. Analytical modelling of regional radiotherapy dose response of lung. Physics in medicine and biology 2012;57:3309-21.

Figure 1: Dose differences between the APT and no APT groups

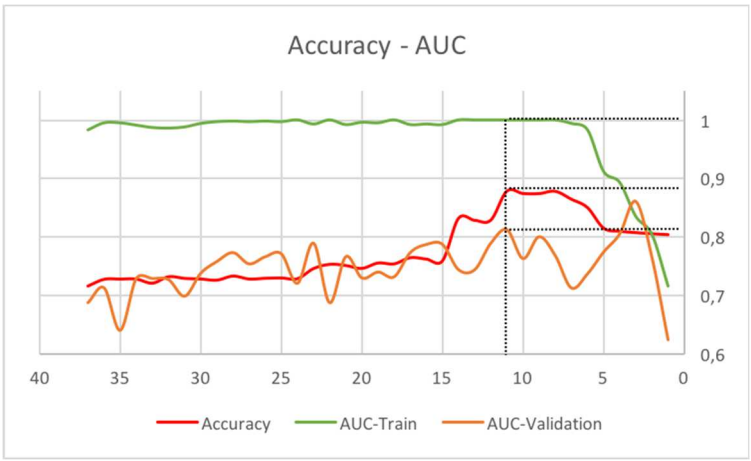


Legend:

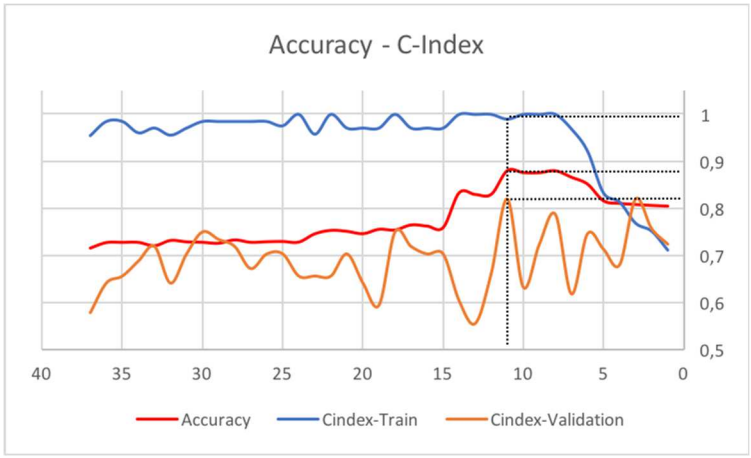
- *Red mask: dose variations between the 2 groups from 1 (light) to 5% (dark)*
- *Blue mask: dose variation between the 2 groups from -1 (light) to -5% (dark)*
- *Yellow mask: Pmap-region*

Figure 2: Results of each decrementally built model in the training set

- 2a: AUC results for each decrementally built model in the training and validation cohorts
- 2b: C-Index results for each corresponding decrementally built model in the training and validation cohorts



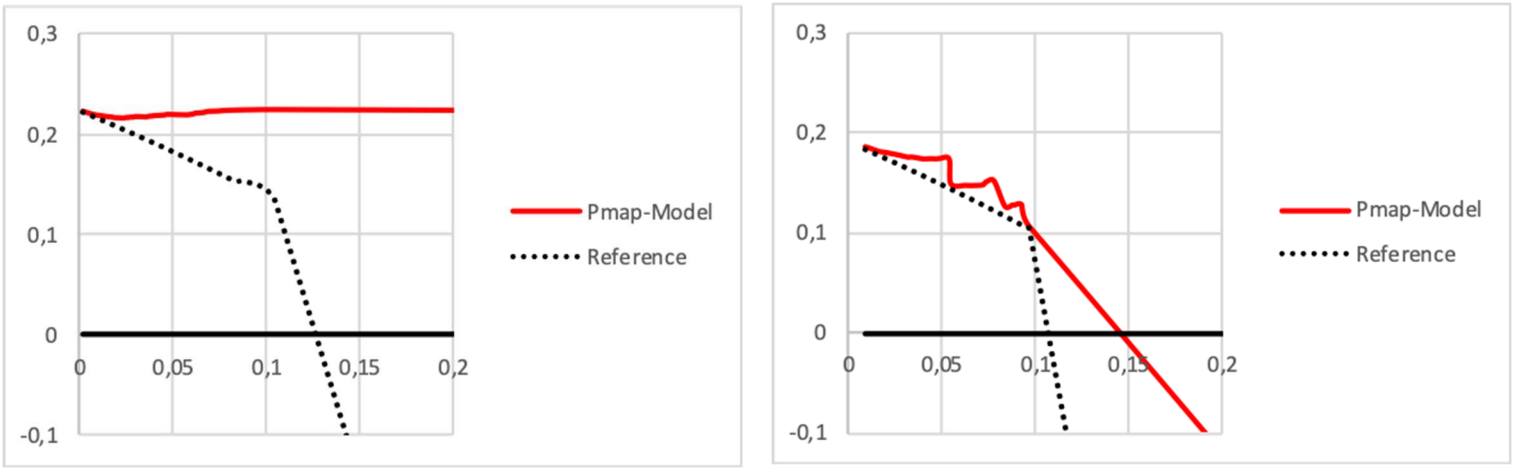
a



b

Legend: x-axis: number of features in the model, y-axis: accuracy of the corresponding model, intersection between the 2 dotted lines = Model N°26 (Pmap-Model). Figure 2a plots the AUC value for each model (y-axis) while the number of included features in the corresponding model is presented on the x-axis. Figure 2b plots the C-Index value for each model (y-axis) while the number of included features in the corresponding model is presented on the x-axis. For any given model, the same features are included in the parts a and b of the figure.

Figure 3: Decision curve analysis for the Pmap-model in the training (a) and validation (b) sets



a b

Legend: x-axis: threshold probability, y-axis: net clinical benefit. The reference corresponds to the net clinical benefit with all patients considered at high risk of APT.

Table 1: Main patients' characteristics in the training and validation sets

	Training cohort n = 165		Validation cohort n = 42		*p*
Age mean (SD)	65.0 (9.38)		65.7 (10.03)		0.67
Gender					
Male (nb, %)	111	67.3	31	73.8	0.42
Female (nb, %)	54	32.7	11	26.2	
Median PS (range)	1	0-2	1	0-2	0.06
Smoking					
Activ (nb, %)	62	37.6	17	40.5	0.52
Former/never (nb, %)	103	62.4	25	59.5	
Known COPD (nb, %)	63 (38.2)		12 (28.6)		0.25
Mean MEVS (%, CI 95%)	74.4 (71.1-77.6)		72.3 (65.7-78.9)		0.58
Histology					
SCC (nb, %)	61	37.0	16	38.1	0.79
ADC (nb, %)	67	40.6	17	40.5	
SCLC (nb, %)	24	14.5	6	14.3	
Others (nb, %)	13	7.9	3	7.1	
AJCC stage (Median)	IIIA		IIIA		0.70
Total RT Dose					
Median (Gy, range)	66.0 (54.0-66.0)		66.0 (54.0-66.0)		0.59
Chemotherapy sequence					
Concomitant (nb, %)	53	32.1	14	33.3	0.99
Induction (nb, %)	48	29.1	12	28.6	
Induction + concomitant (nb, %)	34	20.6	7	16.7	
None (nb, %)	30	18.2	9	21.4	
Adjuvant durvalumab (nb, %)	16	9.7	3	9.5	0.78
APT rate (%)	22.4		19.1		0.64

Abbreviations: SD: Standard Deviation, nb: number, % : percentage, COPD: chronic obstructive pulmonary disease, MEVS: mean expiratory volume/second, SCC: squamous-cell carcinoma, ADC: adenocarcinoma, SCLC: small-cell lung cancer, AJCC: American Joint Commission on Cancer, RT: radiotherapy, APT: Acute pulmonary toxicity ≥ grade 2

Table 2: Classification of each feature by importance and normalized importance – Pmap Model

Feature	Importance	Normalized Importance
$V40_{Heart}$	5.0%	13.9%
$DMean_{LungH}$	5.0%	13.9%
$V5_{LungH}$	5.0%	13.9%
AJCC Stage	5.0%	13.9%
$V10_{LungH}$	6.0%	16.7%
COPD	6.0%	16.7%
MEVS	7.0%	19.4%
Smoking Status	7.0%	19.4%
$V30_{2Lungs}$	7.0%	19.4%
$DMean_{2Lungs}$	11%	30.6%
$DMean_{Pmap}$	36%	100%

Abbreviations: $V40_{Heart}$: Volume of the heart receiving at least 40Gy, $DMean_{LungH}$: Mean dose received by the homolateral lung, $V5_{LungH}$: Volume of the homolateral lung receiving at least 5Gy, AJCC Stage: American Joint Committee on Cancer, $V10_{LungH}$: Volume of the homolateral lung receiving at least 10Gy, COPD: chronic obstructive pulmonary disease, MEVS: mean expiratory volume/second, $V30_{2Lungs}$: Volume of the 2 lungs receiving at least 30Gy, $DMean_{2Lungs}$: Mean dose received by the 2 lungs, $DMean_{Pmap}$: Mean dose received by the Pmap volume

Table 3: Analysis of the Pmap-model's discrimination between patients with or without APT ≥ grade 2 in the training and validation sets

Set	AUC	p	C-index	Se	Sp	BAcc	Number of patients, n (%)					
							Below the cutoff (Low risk of APT ≥ grade 2			Above the cutoff (High risk of APT ≥ grade 2)		
							Total	Without APT	With APT	Total	Without APT	With APT
Overall	0.98	< 0.0001	0.96	97.8	93.8	95.8	153 (73.9)	152 (99.3)	1 (0.7)	54 (26.1)	10 (18.5)	44 (81.5)
Training set	0.99	< 0.0001	0.99	100.0	98.4	99.2	126 (76.4)	126 (100.0)	0 (0.0)	39 (23.6)	2 (5.1)	37 (94.9)
Validation set	0.81	< 0.0001	0.82	87.5	76.5	82.0	27 (64.3)	26 (96.3)	1 (3.7)	15 (35.7)	8 (53.3)	7 (46.7)

Abbreviations: AUC: Area Under the Curve, Se: sensitivity, Sp: specificity, BAcc: Balanced Accuracy, APT: Acute Pulmonary Toxicity