Neoplasia 39 (2023) 100889

Contents lists available at ScienceDirect

Neoplasia

journal homepage: www.elsevier.com/locate/neo

Neural network based ensemble model to predict radiation induced lymphopenia after concurrent chemo-radiotherapy for non-small cell lung cancer from two institutions

Yejin Kim[a,b], Ibrahim Chamseddine[b], Yeona Cho[c], Jin Sung Kim[d], Radhe Mohan[e], Nadya Shusharina[b], Harald Paganetti[b], Steven Lin[e], Hong In Yoon[d,*], Seungryong Cho[a,**], Clemens Grassberger[b]

[a] Department of Nuclear and Quantum Engineering, Korea Advanced Institute of Science and Technology, Daejeon, Korea
[b] Department of Radiation Oncology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA
[c] Department of Radiation Oncology, Gangnam Severance Hospital, Yonsei University College of Medicine, Yonsei University Health System, Seoul, Korea
[d] Department of Radiation Oncology, Yonsei Cancer Center, Yonsei University Health System, Yonsei University College of Medicine, Seoul, Republic of Korea
[e] Division of Radiation Oncology, MD Anderson Cancer Center, Houston, TX, USA

ARTICLE INFO

Keywords:
Radiation-induced lymphopenia
Prediction model
Chemo-radiotherapy
Immunotherapy

ABSTRACT

The use of adjuvant Immune Checkpoint Inhibitors (ICI) after concurrent chemo-radiation therapy (CCRT) has become the standard of care for locally advanced non-small cell lung cancer (LA-NSCLC). However, prolonged radiotherapy regimens are known to cause radiation-induced lymphopenia (RIL), a long-neglected toxicity that has been shown to correlate with response to ICIs and survival of patients treated with adjuvant ICI after CCRT.

In this study, we aim to develop a novel neural network (NN) approach that integrates patient characteristics, treatment related variables, and differential dose volume histograms (dDVH) of lung and heart to predict the incidence of RIL at the end of treatment. Multi-institutional data of 139 LA-NSCLC patients from two hospitals were collected for training and validation of our suggested model. Ensemble learning was combined with a bootstrap strategy to stabilize the model, which was evaluated internally using repeated cross validation.

The performance of our proposed model was compared to conventional models using the same input features, such as Logistic Regression (LR) and Random Forests (RF), using the Area Under the Curve (AUC) of Receiver Operating Characteristics (ROC) curves. Our suggested model (AUC=0.77) outperformed the comparison models (AUC=0.72, 0.74) in terms of absolute performance, indicating that the convolutional structure of the network successfully abstracts additional information from the differential DVHs, which we studied using Gradient-weighted Class Activation Map.

This study shows that clinical factors combined with dDVHs can be used to predict the risk of RIL for an individual patient and shows a path toward preventing lymphopenia using patient-specific modifications of the radiotherapy plan.

Introduction

Lung cancer is one of the leading causes of cancer-related death worldwide and its underlying biological heterogeneity as well as the large number of confounding factors complicate clinical decision making [1]. Following the completion of RTOG 9410 in the early 2010s [2], concurrent use of Chemo-Radiotherapy (CCRT) has become the definitive treatment for unresectable Locally-Advanced Non-Small Cell Lung Cancer (LA-NSCLC). In the late 2010s, the PACIFIC trial demonstrated superior survival for the addition of Immune Checkpoint Inhibitors (ICIs) after the completion of CCRT, which quickly became the new standard of care [3,4], and similar regimens are studied in a range of other indications [5].

Concurrent chemoradiotherapy is known to cause both acute and late toxicities including esophagitis, pneumonitis, fibrosis, and hematologic toxicities, some of them significantly affecting not only quality of life but also survival [6,7]. Among these toxicities, Radiation-Induced Lymphopenia (RIL) has long been disregarded as a common but non-critical side effect of CCRT [8]. However, the addition of adjuvant ICIs to the treatment regimen for lung cancer [3] and other tumor

* Corresponding author.
** Co-corresponding author.
E-mail addresses: yhi0225@yuhs.ac (H.I. Yoon), scho@kaist.ac.kr (S. Cho).

https://doi.org/10.1016/j.neo.2023.100889
Received 7 September 2022; Received in revised form 10 November 2022; Accepted 13 February 2023

types [9,10] have brought renewed interest to RIL and how to mitigate it, due to the realization that the immuno-suppressive effects of radiation may potentially impact the subsequent response to immunotherapy [11–13].

Emerging clinical data support this view – while it has been known for a while that the incidence of RIL correlates to decreased survival after CCRT alone in lung cancer patients [14,15], recent data suggest that the correlation between RIL and outcome further strengthens in the post-PACIFIC era of adjuvant ICI [16,17]. As RIL has been shown to depend on individual factors and the radiotherapy treatment plan [14], efforts have started to predict RIL using statistical or deep learning models based on patient and treatment characteristics [18,19]. However, RIL depends on dose to multiple organs, particularly the heart and lung in our case [20], that are inherently correlated, and on multiple baseline factors that could interact with these dose metrics.

Therefore, in this study, we aim to develop a novel deep-learning-based model to predict RIL for LA-NSCLC patients undergoing definitive CCRT that efficiently takes into account dose to multiple structures as well as confounding factors. We developed this model based on a pre-registered study analysis plan [21], compared it to conventional approaches (logistic regression and random forests) based on the same predictors, and investigated the network to understand which dose regions activate high-risk predictions.

Method & materials

Patient cohort and characteristics

This retrospective study includes patients with LA-NSCLC who were treated with CCRT from two institutions, and the study was approved by each institution's institutional review board [20,22]. Patients were > 18 years old, had weekly blood tests, and received Intensity Modulated Radiation Therapy (IMRT). The median total radiation dose was 68 Gy (range 60-78 Gy) given in 2.0 Gy per fraction (range 1.8-2.2 Gy). The ALC at baseline, after the first week of CCRT, and at the end of CCRT, organ volumes of the tumor, normal lungs and heart, and mean doses to normal lungs and heart were collected. The difference in baseline patient characteristics and treatment related characteristics between the two cohorts was analyzed using Welch's t test (*R* software package, version 3.5).

Input data and endpoint

Differential Dose Volume Histograms (dDVHs) of normal lungs and heart with 0.5Gy dose bins and clinical variables were used as input data for the prediction model. Examples of two different cases are shown in Fig 1.A. Differential DVHs, which have multiple peaks at dose levels that were received by large areas, were selected instead of cumulative DVHs (cDVH), which are monotonously decreasing. Although both fundamentally convey the same information, the idea was to have a compact network with a limited number of layers. The underlying idea is that dDVHs simplify extraction of useful features via convolutional layers and facilitate model interpretation in a given compact network structure. The very low dose bath below 3Gy in the dDVH has a high peak that is common in every case, dominating the spectrum. Therefore we removed this area from the input. In addition to the dDVHs, we used the following clinical variables as features: baseline Absolute Lymphocyte Count (ALC), ALC after 1st week of RT, organ volumes of normal lungs, heart, and tumor, and mean doses of normal lungs and heart. The endpoint of the prediction model was Grade 4 Radiation-Induced Lymphopenia (G4 RIL) at the end of CCRT, which is defined as ALC $< 200\ /\mu l$.

Network structure and hyperparameters

The model starts with two pathways, one for the dDVHs and the other for the clinical variables. In this hybrid structure, dDVHs of normal lungs and heart are fed into the convolutional layers while clinical variables are fed into the fully-connected layers, extracting features from these two types of features separately. At the end of these two paths, the two extracted feature groups are concatenated using a fully connected layer so that the interplay between dDVHs and clinical variables can be taken into account by the model. We note that a similar structure has been used recently by Cui et al. [23] in integrating multi-omics information with dosimetric data, a schematic structure of our network is shown in Fig 1.C with detailed description of the hyperparameters specified.

Model evaluation and ensemble model

In order to avoid overfitting and to develop a robust and stable model, we applied repeated 5-fold internal cross validation integrated with a bootstrapped ensemble strategy (Fig 1.B). After dividing the dataset into 5 folds in a stratified manner, i.e. with a similar proportion of G4 RIL events in each fold, we used bootstrapping with replacement on the 4 training folds to create 10 bootstrapped subgroups, and trained a separate model on each of them. The final model prediction was computed as the average output of these 10 sub-models. We repeated this process multiple times to evaluate how robust the model is against random data splits – due to the limited size of the cohort and the heterogeneity among institutions we did not form a separate validation cohort in this work. We also report results for the single neural network (sNN) separately from the ensemble neural network (eNN), to demonstrate the difference in performance and robustness gained by using an ensemble approach.

Logistic regression (LR) and random forests (RF) were also implemented and applied for comparison [24,25]. As those models may be easily overfitted given a large number of highly colinear input features from the dosimetric data, we provided these models with the dosimetric input in 5Gy increments to avoid overfitting, similar to published approaches [18]. For the RF model, the number of trees in the forest, the maximum depth of the tree, and the minimum number of samples to split an internal node were 100, 5, and 2 respectively. These parameters provide a good general choice but were, similar to the hyperparameters of the CNN described above, not optimized.

Model comparison – feature importance and examining neural network activation patterns

The comparison models were fit and evaluated in parallel with the neural network using the same cross-validation process and were compared using the Receiver Operating Characteristics (ROC) curve to calculate the Area Under Curve (AUC). We quantified the robustness of the model, i.e. the variability of its performance across data splits, using confidence intervals of the observed AUCs. To avoid overly optimistic reporting of the selected results, our proposed model structure, validation methodology, endpoints, and metrics were pre-specified in a study analysis plan that was published before data collection was initiated [21], deviations from which are discussed below.

To investigate possible differences in performance between the linear, i.e., LR vs more complex models such as RF/NN, we performed three consecutive analyses. First, we analyzed the correlation between clinical and dosimetric variables and G4 RIL using uni- and multi-variable logistic regression. Second, we evaluated the feature importance in the LR using the coefficients and in the RF using the impurity, averaged over the repeated crossvalidation.

Third, for the neural network we investigated which areas of the DVH are most correlated to the risk, i.e. which dose levels activate the network most, using a Gradient-weighted Class Activation Map (Grad-CAM) methodology [26]. This class-discriminative visualization allows us to interrogate the model, indicating the activated region at the end of convolutional layers. We averaged the activation maps for the entire population and compared the observed activation in patients that experienced G4 RIL with those who did not in order to understand which

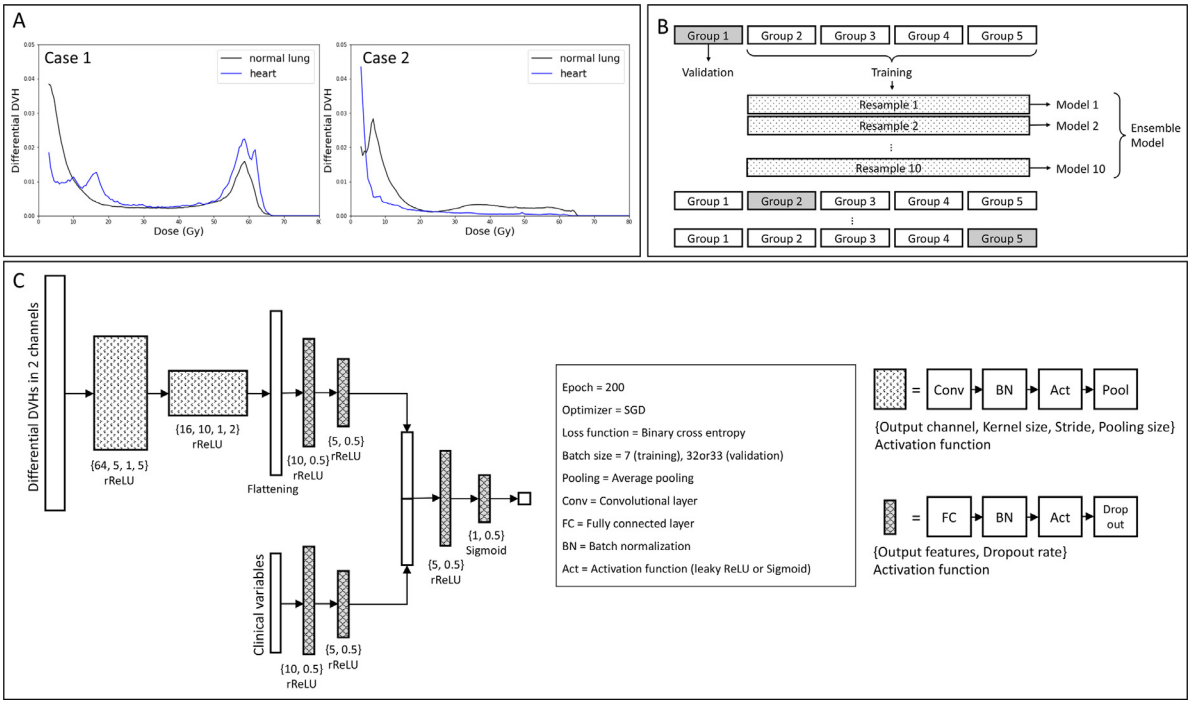

Fig. 1. Schematic diagram of the network. **A.** Two example cases of dDVHs for normal lungs and heart. **B.** Schematic diagram of five-fold cross-validation using ensemble strategy. **C.** Network structure with input channels and hyperparameters.

Table 1
Patient characteristics.

Organ volume (L)	Normal lungs	3.17 (1.47 - 7.38)
	Heart	0.71 (0.18 - 1.47)
	Tumor	0.51 (0.01 - 2.55)
Mean dose (Gy)	Normal lungs	18.0 (2.43 - 49.9)
	Heart	7.49 (0.15 - 27.42)
ALC ($/\mu L$)	Baseline	1.68 (0.3 - 4.27)
	First week of CCRT	1.03 (0.08 - 2.34)
	End of CCRT	0.36 (0.06 - 0.92)

dose levels are important for the development of lymphopenia. Differences in the activation levels between groups were analyzed using the two-sample Kolmogorov-Smirnov test.

Results

In total 139 LA-NSCLC patients treated with CCRT from two hospitals were included in this study. The median and range of the patient characteristics are summarized in Table 1 including organ volumes, mean doses for organs and ALC at baseline, first week and end of CCRT. The endpoint for the prediction model was Grade 4 RIL, which had an incidence of 16% across the cohort – 15% at institution A vs 17% at institution B (no significant difference). The median ALC at the end of CCRT across all patients was 0.38 $/\mu$L, ranging from 0.06-0.92 $/\mu$L. 133 patients developed G2+ RIL (ALC <800), 108 developed G3+ (ALC <500) and 22 developed G4 RIL (ALC<200).

Table 2 summarizes the difference in patients and treatment characteristics between the two institutions and between the groups of patients with G4 RIL at the end of treatment and those without. When comparing the two hospitals, we concluded that it showed clear differences in most of the variables, particularly including organ volumes and ALC before and during the treatment. For example, the lung volumes and ALCs before treatment were 3.5 L and 1.66 $/\mu L$ at hospital A and 3.1 L and 2.03 $/\mu L$ at hospital B. Among treatment-related factors, the total dose and the mean dose to normal lungs were substantially different between hospitals. On the other hand, mean heart dose was not statistically different between the cohorts. These large differences between the institutions led to the decision to perform repeated cross-validation instead of using one cohort as training and the other as a validation set. Similarly, when comparing patients with G4 RIL and no G4 RIL, we observe differences in normal lung volume, baseline and first week ALCs, and mean heart dose. Amongst the variables, the mean heart dose showed the most significant difference with p-value of 0.004.

Fig. 2 shows the resulting performance of the models: the single Neural Network (sNN), logistic regression (LR) and random forest (RF) models showed AUCs of 0.71, 0.73, and 0.73 with a 95% CI of 0.1, 0.04 and 0.06 respectively. The ensemble strategy (eNN) improved the performance of the neural network to 0.77 in AUC while increasing the robustness by lowering the 95% CI from 0.2 to 0.14. We investigated Wilcoxon signed rank sum test to compare the models across all iterations of the repeated cross validation to investigate if there was a statistically significant improvement in AUC values when introducing our proposed ensemble NN over comparison models. A signed rank sum test was selected since each training and validation was performed with the same stratified split of the data set for all four models. The single NN model had a statistically inferior performance, while when comparing RF and LR there was no statistical difference (p-value 0.3). The ensemble NN model on the other hand showed statistically significant improvement of AUC compared to all three models with p-values <0.0001 for all.

To understand the difference in performance between the logistic regression and the two more complex algorithms (RF and NN), we ran uni- and multi-variable logistic regressions of G4RIL and the clinical factors and investigated the possible importance of interactions and collinearity. We also included the dosimetric variables that had the strongest correlation to G4RIL, which were mean heart dose and lung V15, in our analysis (see Table 3). ALCs before and after the first week of CCRT, tumor volume, lung V15 and mean heart dose were correlated to G4RIL. Including these in a multivariable regression (see last column Table 3) leaves only ALC after the first week of CCRT and PTV volume as statistically significant and trending towards significance respectively.

The analysis of feature importance for both logistic regression and random forest models mirror the multivariable regression results,

Table 2
Patient- and treatment-related characteristics for the two cohorts, with p value from Welch's t test.

		Hospital A	Hospital B	p-value	G4 RIL	No G4 RIL	p-value
Organ volume (L)	Normal lungs	3.5	3.1	0.007	3.41	3.06	0.05
	Heart	0.69	0.76	0.09	0.70	0.79	0.13
	Tumor (PTV)	0.71	0.46	<0.001	0.58	0.85	0.06
ALC ($/\mu L$)	Baseline	1.66	2.03	0.003	1.8	1.5	0.03
	First week of CCRT	1.01	1.17	0.041	0.32	0.47	0.03
	End of CCRT	0.38	0.43	0.15			
CCRT regimen	Total dose (Gy)	70.9	61.8	<0.001	68.17	66.25	0.11
	Mean normal lung dose (Gy)	21.2	15.1	<0.001	19.0	20.0	0.46
	Mean heart dose (Gy)	8.9	7.6	0.22	7.8	12.2	0.004

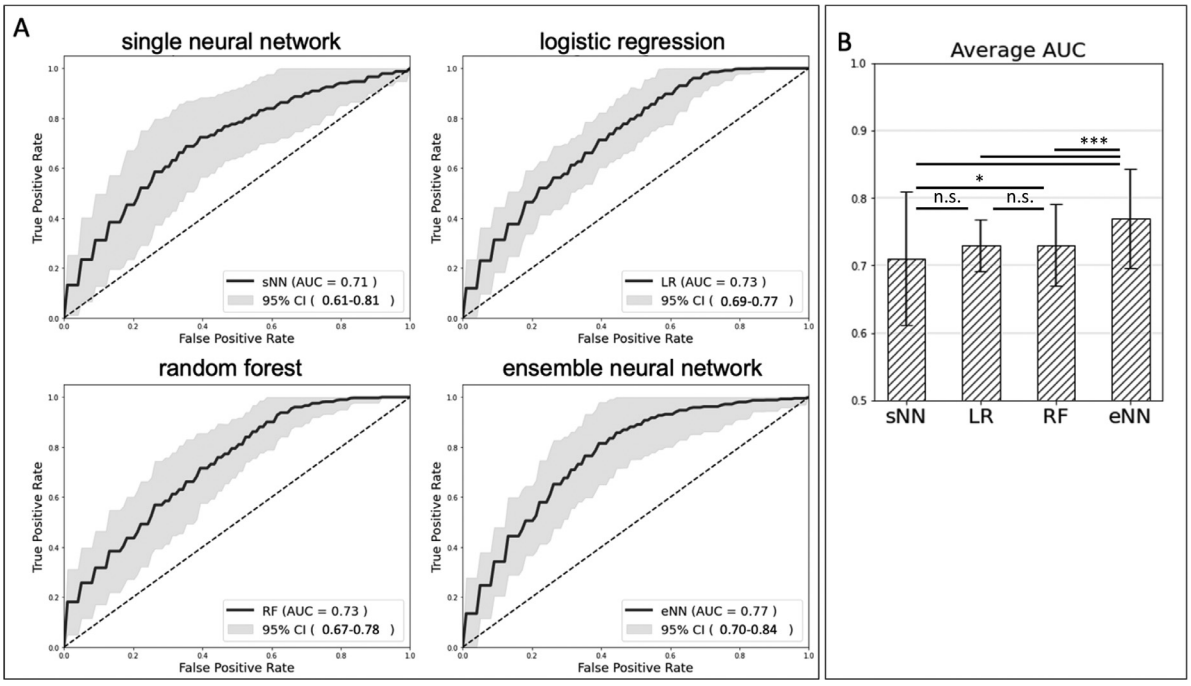

Fig. 2. Receiver Operating Characteristics (ROC) curve and bar chart of AUCs throughout repeated cross-validation. **A.** ROC curves and according AUCs for sNN, LR, FR and eNN. **B.** Bar chart of average AUCs with CI 95% error for all four models. Significance according to Wilcoxon ranked sum test among all data splits and cross validation iterations; n.s. = not significant, $^{*} = p < 0.05$, $^{***} = p < 0.001$.

Table 3
Uni- and multi-variable logistic regression correlating G4 RIL with clinical and important dosimetric features.

	p value of univariable LR	p value of multivariable LR
ALC before CCRT	0.05	0.87
ALC after 1st week	0.001	0.007
PTV	0.01	0.07
Mean heart dose	0.002	0.93
Lung V15	0.05	0.99
Total dose	0.14	

demonstrating that the clinical features are more relevant to these models compared to the dosimetric features. In the feature importance ranking all dose levels seem to contribute similarly when averaged across all data splits, with heart dose contributing more to the prediction than lung dose (see supplementary Fig. 1).

Fig. 3A shows the average Grad-CAM activation maps of all patients (left), those without G4 RIL (middle), and those with G4 RIL (right), together with the corresponding dDVHs of normal lungs and heart. Overall, the network mostly focuses on the very low dose area (<10 Gy). However, in the G4 RIL patients, the model also reacted to middle and high dose areas, more so than for the group without G4 RIL. Fig 3B shows two example patients with their specific Grad-CAM results. For the patient without radiation-induced lymphopenia (left), lung and heart dDVHs have peaks in the low dose areas and the activation of the network happens in the same region. On the other hand, the network focused on broader regions for the G4 RIL case (right), and the heart dDVH shows a broad range of dose distribution from the low to the high dose region. Statistical analyses of the activation maps for the patients with G4 RIL shows that they are significantly different compared to the patients experiencing no lymphopenia ($p < 0.0001$).

Discussion

In this study, we developed an NN-based ensemble model to predict radiation-induced lymphopenia after CCRT for LA-NSCLC patients based on a very heterogeneous cohort from two institutions. Our model showed superior performance compared to logistic regression and random forests when evaluated using repeated cross-validation on the same data splits using the same input data, i.e. same clinical factors and dosimetric data. The robustness of the neural network was greatly improved by introducing an ensemble approach, indicating this strategy could be useful when fitting complex models that include dosimetric data to

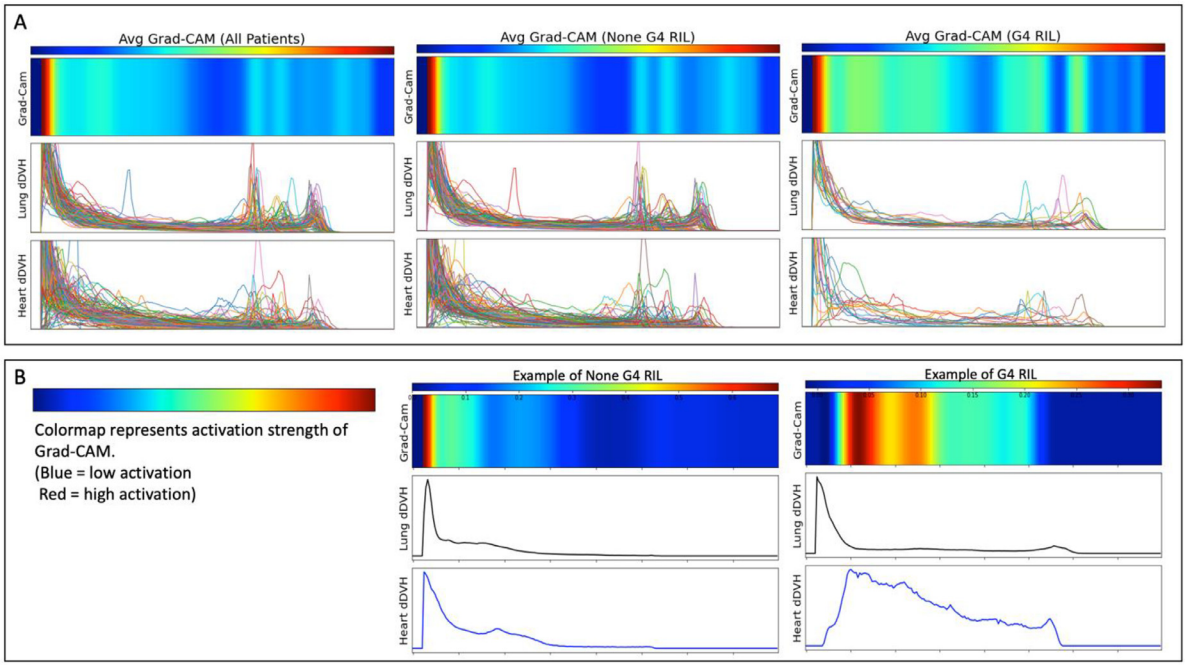

Fig. 3. Grad-CAM **A.** Averaged Grad-CAM over all, those without G4 RIL, and G4 RIL patients and corresponding dDVHs of normal lungs and heart. **B.** Example cases of Grad-CAM for two specific patients, without (left) and with (right) radiation-induced lymphopenia.

the limited datasets common in radiation oncology. Logistic regression showed the most stable predictions, though with the poorest overall performance.

We compared our model to logistic regression and random forests, the former a very standard and robust approach and the latter a different robust model that allows modeling of the feature interactions. To first understand the effects of the input variables on radiation-induced lymphopenia, we performed a uni- and multi-variable logistic regression (Table 3). As expected, the factors most correlated to the ALC at the end of CCRT were ALC before and at the first week of CCRT, tumor volume, and dosimetric parameters for lung and heart, which agrees with the findings in previous studies [14,20]. However, the multi-variable regression showed how, due to collinearity, only two of these factors remained important when combined together (Table 3). The fact that ALC after the first week became the most dominant factor in the multivariable regression analysis does not imply that dose level is not important and that severe RIL cannot be prevented by focused treatment planning strategies – it rather shows that the dose is causing that steep decline of ALC in the first week, which has been recently also shown by Ellsworth et al. [29].

This explains why when analyzing the feature importance in the LR models among all data splits (see Supplementary Fig. 1), the dosimetric features have in general a low feature importance compared to the clinical features and even turn negative sometimes, meaning dose would be negatively correlated to our endpoint. This is common behavior in the case of feature collinearity, when weaker collinear features are driven by the noise among data splits in the dataset. The random forest model is an improvement in that regard, as it is more robust towards collinearity due to the built-in bootstrap aggregating of data as well as predictors among the different decision trees.

To study the feature importance in the neural network, we used the well known Grad-CAM methodology to interpret the network and study which regions of the dDVH led to the strongest activation. In contrast to the LR and RF models, the neural network focused in general on the low dose bath area, which is in line with previous clinical studies investigating lung cancer patients [14,27,28]. As the clinical and dosimetric features available to the models are the same, our results indicate that this type of network structure, consisting of multiple dDVHs feeding as separate channels into a convolutional layer, is an efficient way to abstract DVH information into higher-order features and let them interact with clinical factors in the following layers. Taking the entire dDVHs as inputs has an advantage as it preserves the continuous changes and obvious collinearity of neighboring regions in the histogram, which is efficiently considered by convolutional layers.

Using a similar approach Cui et al. predicted radiation pneumonitis and local control using dDVH and multi-omic features [23]. In that study, the group experiencing pneumonitis did not show clear differences in activation maps, indicating a common dose range of activation. In our study, however, the activated area for the G4 RIL patient group was distinct from the other patients, the network paying more attention to the broader dose range from low to high dose.

Initialization of the neural network parameters and weights can substantially impact the network convergence for training [30], which is why many studies in this field fix the random seed to preserve reproducibility. However, the effect of the random seed becomes dominant for small networks, and the results can be sensitive to initialization. Therefore, we randomized our initialization in every repeated cross-validation, confirming that the outperformance of our model is generalizable, though this might have caused the rather larger confidence intervals in the performance of the NN models.

An important constraint of this study exists in the statistical difference between the cohorts from the two institutions. Except for the endpoint - ALC incidence was similar between cohorts - most baseline factors differed significantly between the two cohorts (Table 2). Considering this difference, we opted to evaluate the model using repeated internal cross-validation and not using one institutional cohort for training and the other one for testing, as this would only yield one AUC for each model, making robust conclusions impossible. Internal cross-validation however can lead to over-optimistic results, particularly for complex algorithms for which a lot of hyper-parameters can be adjusted. We therefore did not use hyperparameter optimization to fine-tune the performance of the random forest or the neural network but stayed within the limits set in our study analysis plan [21], which was submitted during the data collection phase of the project.

At the same time, this large heterogeneity in the population gives us more confidence that the performance estimates of our models are

not overly optimistic. Nevertheless, further prospective validation using an independent dataset is required to strengthen the credibility of the prediction model, and larger cohorts are expected to lead to better overall performance. The code is available to interested researchers upon request, please contact the authors for details.

Another possible improvement may be made by employing higher dimensional data – studies have recently shown that features of the three-dimensional radiation dose distribution can be correlated to lymphopenia in thoracic radiotherapy patients [20,31], indicating that the upper mediastinum might be an important region for the development of RIL. Given larger training cohorts, we are therefore planning to employ higher dimensional dose distribution information, similar to other emerging approaches [12], to improve prediction performance in future studies.

Conclusion

In this retrospective study, we developed a novel neural-network-based ensemble model to predict radiation-induced lymphopenia at the end of CCRT and compared the performance of the proposed model to those of existing approaches. Our results show that this type of compact convolutional structure can abstract dosimetric information from multiple organ DVHs very efficiently and combine them with important clinical and patient factors. We also demonstrated that ensemble modeling succeeds in stabilizing model variance even when randomizing initial weights at every iteration. Grad-CAM based network analysis helped explain which areas of the DVH are most impactful for the development of lymphopenia and provided valuable insights on how to prevent RIL using patient-specific modifications of the radiotherapy treatment plan. Models such as the one proposed here are crucial for the emerging concept of "lymphocyte-sparing radiotherapy", which seeks to minimize the adverse impact of RT on the immune system in the context of radiotherapy-immunotherapy combination regimen.

Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Yejin Kim: Methodology, Software, Investigation, Visualization, Writing – original draft. **Ibrahim Chamseddine:** Methodology, Software. **Yeona Cho:** Conceptualization, Resources, Data curation. **Jin Sung Kim:** Conceptualization, Resources, Data curation. **Radhe Mohan:** Conceptualization, Resources, Data curation. **Nadya Shusharina:** Resources, Data curation, Writing – review & editing. **Harald Paganetti:** Conceptualization, Methodology, Supervision. **Steven Lin:** Conceptualization, Resources, Data curation. **Hong In Yoon:** Conceptualization, Writing – review & editing, Supervision. **Seungryong Cho:** Conceptualization, Writing – review & editing, Supervision. **Clemens Grassberger:** Conceptualization, Writing – review & editing, Supervision.

Funding

This work was supported in part by the National Cancer Institute (P01 CA 261699) and the National Research Foundation of Korea (Grant NRF2020R1A2C2011959).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neo.2023.100889.

References

[1] A.A. Thai, B.J. Solomon, L.V. Sequist, J.F. Gainor, R.S. Heist, Lung cancer, Lancet 398 (10299) (2021) 535–554, doi:10.1016/S0140-6736(21)00312-3.
[2] W.J. Curran, et al., Sequential vs concurrent chemoradiation for stage iii non-small cell lung cancer: randomized phase III trial RTOG 9410, J. Natl. Cancer Inst. 103 (19) (2011) 1452–1460, doi:10.1093/jnci/djr325.
[3] A. Balasubramanian, J. Onggo, A. Gunjur, T. John, S. Parakh, Immune checkpoint inhibition with chemoradiotherapy in stage iii non–small-cell lung cancer: a systematic review and meta-analysis of safety results, Clin. Lung Cancer 22 (2) (2021) 74–82, doi:10.1016/j.cllc.2020.10.023.
[4] S.J. Antonia, et al., Durvalumab after chemoradiotherapy in stage III non–small-cell lung cancer, N. Engl. J. Med. 377 (20) (2017) 1919–1929, doi:10.1056/nejmoa1709937.
[5] W.T. Turchan, S.P. Pitroda, R.R. Weichselbaum, Combined radio-immunotherapy: An opportunity to increase the therapeutic ratio of oligometastasis-directed radiotherapy, Neoplasia (United States) 27 (C) (2022) 100782, doi:10.1016/j.neo.2022.100782.
[6] V. Verma, C.B. Simone, M. Werner-Wasik, Acute and late toxicities of concurrent chemoradiotherapy for locally-advanced non-small cell lung cancer, Cancers (Basel) 9 (9) (2017), doi:10.3390/cancers9090120.
[7] R.W. Byhardt, et al., Response, toxicity, failure patterns, and survival in five radiation therapy oncology group (RTOG) trials of sequential and/or concurrent chemotherapy and radiotherapy for locally advanced non-small-cell carcinoma of the lung, Int. J. Radiat. Oncol. Biol. Phys. 42 (3) (1998) 469–478, doi:10.1016/S0360-3016(98)00251-X.
[8] B.M. Lee, H.K. Byun, J. Seong, Significance of lymphocyte recovery from treatment-related lymphopenia in locally advanced pancreatic cancer, Radiother. Oncol. 151 (2020) 82–87, doi:10.1016/j.radonc.2020.07.026.
[9] R. Thomas, G. Al-Khadairi, J. Decock, Immune checkpoint inhibitors in triple negative breast cancer treatment: promising future prospects, Front. Oncol. 10 (2021) 1–17 February, doi:10.3389/fonc.2020.600573.
[10] Gaber Plavc, et al., Challenges in combining immunotherapy with radiotherapy in recurrent/metastatic head and neck cancer, Cancers 12 (11) (2020) 3197 30 Oct., doi:10.3390/cancers12113197.
[11] G.Y. Ku, et al., Single-institution experience with ipilimumab in advanced melanoma patients in the compassionate use setting, Cancer 116 (2010) 1767–1775.
[12] A.E. Marciscano, et al., Elective nodal irradiation attenuates the combinatorial efficacy of stereotactic radiation therapy and immunotherapy, Clin. Cancer Res.Clincanres. 3427 (2017) (2018), doi:10.1158/1078-0432.ccr-17-3427.
[13] L.R.G. Pike, et al., The impact of radiation therapy on lymphocyte count and survival in metastatic cancer patients receiving PD-1 immune checkpoint inhibitors, Int. J. Radiat. Oncol. Biol.Phys. 103 (2019) 142–151.
[14] C. Tang, et al., Lymphopenia association with gross tumor volume and lung V5 and its effects on non-small cell lung cancer patient outcomes, Int. J. Radiat. Oncol. Biol.Phys. 89 (2014) 1084–1091.
[15] J.L. Campian, X. Ye, M. Brock, S.A. Grossman, Treatment-related lymphopenia in patients with stage III non-small-cell lung cancer, Cancer Investig. 31 (2013) 183–188.
[16] Wang Jing, Ting Xu, Lirong Wu, Pablo B. Lopez, Clemens Grassberger, Susannah G. Ellsworth, Radhe Mohan, et al., Severe radiation-induced lymphopenia attenuates the benefit of durvalumab after concurrent chemoradiotherapy for non-small cell lung cancer, JTO Clin. Res. Rep. (2022) 100391.
[17] C. Friedes, et al., Association of severe lymphopenia and disease progression in unresectable locally advanced non-small cell lung cancer treated with definitive chemoradiation and immunotherapy, Lung Cancer (2021), doi:10.1016/j.lungcan.2021.01.022.
[18] C Zhu, SH Lin, X Jiang, et al., A novel deep learning model using dosimetric and clinical information for grade 4 radiotherapy-induced lymphopenia prediction, Phys. Med. Biol. 65 (3) (2020) 035014 Published 2020 Feb 4, doi:10.1088/1361-6560/ab63b6.
[19] TE Kroese, J Jairam, JP Ruurda, et al., Severe lymphopenia acquired during chemoradiotherapy for esophageal cancer: Incidence and external validation of a prediction model, Radiother. Oncol. 163 (2021) 192–198, doi:10.1016/j.radonc.2021.08.009.
[20] Y. Cho, et al., Lymphocyte dynamics during and after chemo-radiation correlate to dose and outcome in stage III NSCLC patients undergoing maintenance immunotherapy, Radiother. Oncol. 168 (2022) 1–7, doi:10.1016/j.radonc.2022.01.007.
[21] Kim, Y., Chamseddine, I., Paganetti, H., Cho, Y., Yoon, H.I., Kim, J.S., … Grassberger, C. (2022, February 25). Prediction model for stage III non-small cell lung cancer patients treated with radiotherapy. 10.17605/OSF.IO/TXZ2P
[22] Z. Liao, et al., Bayesian adaptive randomization trial of passive scattering proton therapy and intensity-modulated photon radiotherapy for locally advanced non–small-cell lung cancer, J. Clin. Oncol. 36 (18) (2018) 1813–1822, doi:10.1200/JCO.2017.74.0720.
[23] S. Cui, R.K. Ten Haken, I. El Naqa, Integrating multiomics information in deep learning architectures for joint actuarial outcome prediction in non-small cell lung cancer patients after radiation therapy, Int. J. Radiat. Oncol. Biol. Phys. 110 (3) (2021) 893–904, doi:10.1016/j.ijrobp.2021.01.042.
[24] J. Tolles, W.J. Meurer, Logistic regression: Relating patient characteristics to outcomes, JAMA - J. Am. Med. Assoc. 316 (5) (2016) 533–534, doi:10.1001/jama.2016.7653.
[25] S. Wongvibulsin, K.C. Wu, S.L. Zeger, Clinical risk prediction with random forests for survival, longitudinal, and multivariate (RF-SLAM) data analysis, BMC Med. Res. Methodol. 20 (1) (2019) 1–14, doi:10.1186/s12874-019-0863-0.

[26] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: visual explanations from deep networks via gradient-based localization, Int. J. Comput. Vis. 128 (2) (2020) 336–359, doi:10.1007/s11263-019-01228-7.
[27] S. Vivekanandan, et al., The impact of cardiac radiation dosimetry on survival after radiation therapy for non-small cell lung cancer, Int. J. Radiat. Oncol. Biol. Phys. 99 (1) (2017) 51–60, doi:10.1016/j.ijrobp.2017.04.026.
[28] S. Ebrahimi, et al., Radiation-induced lymphopenia risks of photon versus proton therapy for esophageal cancer patients, Int. J. Part. Ther. 8 (2) (2021) 17–27, doi:10.14338/IJPT-20-00086.
[29] S.G. Ellsworth, et al., Lymphocyte depletion rate as a biomarker of radiation dose to circulating lymphocytes during fractionated partial-body radiotherapy, Adv. Radiat. Oncol. 100959 (2022), doi:10.1016/j.adro.2022.100959.
[30] S. Qian, et al., Are my deep learning systems fair? An empirical study of fixed-seed training, Adv. Neural Inf. Process. Syst. 36 (2021) 30211–30227 NeurIPS.
[31] S. Monti, T. Xu, R. Mohan, Z. Liao, G. Palma, L. Cella, Radiation-induced esophagitis in non-small-cell lung cancer patients: voxel-based analysis and NTCP modeling, Cancers (Basel) 14 (7) (2022) 1–14, doi:10.3390/cancers14071833.