ARTIFICIAL INTELLIGENCE

End-to-End Non–Small-Cell Lung Cancer Prognostication Using Deep Learning Applied to Pretreatment Computed Tomography

Felipe Soares Torres, MD, PhD[1]; Shazia Akbar, PhD[2]; Srinivas Raman, MD[3]; Kazuhiro Yasufuku, MD, PhD[4]; Carola Schmidt, MD[2]; Ahmed Hosny, PhD[5,6]; Felix Baldauf-Lenschen, BA[2]; and Natasha B. Leighl, MD, MMSc[7,8]

PURPOSE Clinical TNM staging is a key prognostic factor for patients with lung cancer and is used to inform treatment and monitoring. Computed tomography (CT) plays a central role in defining the stage of disease. Deep learning applied to pretreatment CTs may offer additional, individualized prognostic information to facilitate more precise mortality risk prediction and stratification.

METHODS We developed a fully automated imaging-based prognostication technique (IPRO) using deep learning to predict 1-year, 2-year, and 5-year mortality from pretreatment CTs of patients with stage I-IV lung cancer. Using six publicly available data sets from The Cancer Imaging Archive, we performed a retrospective five-fold cross-validation using pretreatment CTs of 1,689 patients, of whom 1,110 were diagnosed with non–small-cell lung cancer and had available TNM staging information. We compared the association of IPRO and TNM staging with patients' survival status and assessed an Ensemble risk score that combines IPRO and TNM staging. Finally, we evaluated IPRO's ability to stratify patients within TNM stages using hazard ratios (HRs) and Kaplan-Meier curves.

RESULTS IPRO showed similar prognostic power (concordance index [C-index] 1-year: 0.72, 2-year: 0.70, 5-year: 0.68) compared with that of TNM staging (C-index 1-year: 0.71, 2-year: 0.71, 5-year: 0.70) in predicting 1-year, 2-year, and 5-year mortality. The Ensemble risk score yielded superior performance across all time points (C-index 1-year: 0.77, 2-year: 0.77, 5-year: 0.76). IPRO stratified patients within TNM stages, discriminating between highest- and lowest-risk quintiles in stages I (HR: 8.60), II (HR: 5.03), III (HR: 3.18), and IV (HR: 1.91).

CONCLUSION Deep learning applied to pretreatment CT combined with TNM staging enhances prognostication and risk stratification in patients with lung cancer.

INTRODUCTION

Lung cancer remains the leading cause of cancer death in North America and worldwide.[1] The TNM staging system is used to classify the anatomic extent of cancerous tissue. This system helps to discriminate between patients into distinct groups, called TNM stages,[2] and informs management of patients with cancer.[3] In patients with lung cancer, TNM staging is a key prognostic factor, driving treatment and monitoring decisions.[4] Radiologic imaging, particularly computed tomography (CT), plays a central role in defining the stage of disease. Analysis of CTs currently relies upon manual localization, classification, and measurement of lesions and is subject to interobserver and intraobserver variability.[5-9] More precise prognostication could help clinicians make personalized treatment decisions that can guide de-escalation and intensification strategies to optimize outcomes in patients with cancer.

The adoption of deep learning techniques in radiologic research have shown promise in performing image interpretation tasks including detection, classification, and segmentation with close-to-expert performance.[10-12] Beyond common image interpretation tasks, convolutional neural networks (CNNs), a form of deep learning, are able to identify and quantify complex features in images that are not readily discernible to the naked eye. A few studies have used CNNs to derive mortality risk prediction in patients with lung cancer[13-15]; however, most of this work relies upon manual steps, such as segmenting the primary lesion,[15] or placing seed points[16] or bounding boxes[17] over regions of interest. Prior imaging-based prognostication (IPRO) research has shown that imaging features extracted not just from the tumor itself but also from the surrounding tissue[18-20] and body composition[21-23] can provide additional prognostic insight. A fully automated approach, in which a system would analyze the entire thorax in a CT, may complement traditional TNM staging of patients with lung cancer and provide greater prognostic power in a standardized manner that does not require manual steps subject to interobserver variability.

ASCO

JCO® Clinical Cancer Informatics

CONTEXT

Key Objective

Lung cancer is the leading cause of cancer death worldwide. TNM staging derived from computed tomography (CT) plays a central role in determining a lung cancer prognosis. This study explores deep learning's ability to automatically analyze the entire thorax of a patient with lung cancer and generate an imaging-based prognostication score from a pretreatment CT.

Knowledge Generated

When tested on 1,110 patients with stage I-IV non–small-cell lung cancer with known survival outcomes, imaging-based prognostication complements TNM staging by enhancing prognostication and risk stratification.

Relevance

Prognostic insight beyond TNM staging can be automatically derived from the thorax using deep learning applied to pretreatment CT scans of patients with non–small-cell lung cancer. More accurate prognostication has the potential to inform personalized treatment decisions.

In this retrospective study, we propose an end-to-end deep learning approach in which the entire thorax of individual patients with lung cancer is automatically evaluated to generate an IPRO score. Using publicly available pretreatment CTs split across a five-fold validation, we assess how IPRO compares with and complements TNM staging for purposes of 1-year, 2-year, and 5-year mortality risk predictions in the withheld validation set. Furthermore, we evaluate IPRO's ability to stratify patients across and within TNM stages. Finally, we review the distribution of known prognostic clinical variables including age, sex, TNM stage, and histology across IPRO's risk deciles and quantify the amount of attention placed on lung lesions.

METHODS

Data

We retrospectively reviewed publicly available pretreatment CTs of patients with lung cancer that also contained survival outcomes. Imaging data and associated clinical information were obtained from six data sets[24-29] made available in The Cancer Imaging Archive (Data Supplement). A total of 1,689 patients were selected who had a biopsy-confirmed non–small-cell lung cancer diagnosis, survival information, and at least one pretreatment axial CT. Patients diagnosed with small-cell lung cancer were excluded. Mortality and CT acquisition dates were used to compute survival time and status at specified censoring dates (ie, 1 year, 2 years, and 5 years). Cases that were lost to follow-up before a given censoring date were excluded from training and validation (Data Supplement).

Given that some patients had multiple pretreatment CTs, we limited the validation to only the final (ie, most recent) pretreatment CT to assess the performance of IPRO and TNM staging. The Data Supplement provides an overview of the distribution of TNM stages and survival status among the 5-year validation data set, which contained 1,110 patients (579 alive and 531 deceased) with a median age of 64 years (range: 43-88 years) and in which 62% were male.

The median time between date of image acquisition and last follow-up was 5.0 years. Additional data selection and preprocessing details are provided in the Extended Methods section in the Data Supplement.

IPRO Framework

The proposed IPRO framework consists of a thorax localizer and a three-dimensional convolutional neural network (3DCNN) that extracts imaging features automatically along the axial, sagittal, and coronal directions, simultaneously (Fig 1). The thorax localizer consists of an algorithm that limits the model input to a 3D space (36 cm × 36 cm × 36 cm in size) centered on the lungs, thus excluding features outside of the thorax (eg, abdomen) and outside of the body (eg, CT scanner table; see the Data Supplement for full description of the thorax localizer model). Input size was determined by assessing segmentation masks from the thorax localizer and selecting the largest bounding box encompassing the lungs, suitable for the given data set. The automatically identified thorax region is then fed into the 3DCNN, which outputs probability scores between 0 and 1 indicating 1-year, 2-year, and 5-year mortality for a given CT.

The architecture of our 3DCNN is based on a widely adopted, state-of-the-art, neural network called InceptionNet.[30] This architecture enables features to be learned without being prone to overfitting, suitable for medical applications where individual data points tend to be large but the number of patients is few (compared with other computer vision tasks). To make the two-dimensional InceptionNet 3D, transfer learning is first applied to stabilize the network using ImageNet,[31] and then intermediate layers are duplicated in a new temporal dimension (ie, *z*-axis). The use of transfer learning in this manner has shown to be effective for preventing overfitting in medical applications.[32] The resulting architecture allows for entire 3D CT volumes to be fed into the 3DCNN without further modifications and incorporates rich features like volume of tumors and features in peritumoral tissue that span multiple CT slices, rather than just a single two-dimensional slice. Although this architecture was originally intended for video sequence
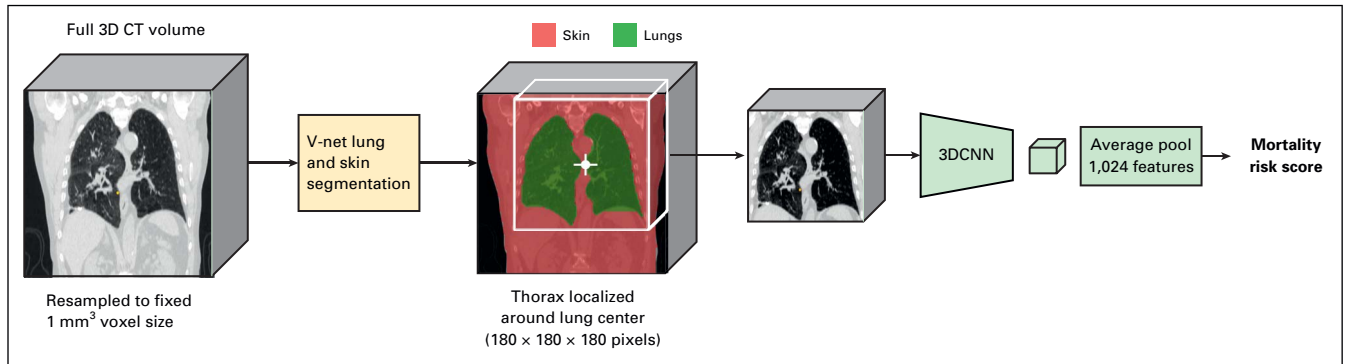
FIG 1. Proposed IPRO mortality risk prediction pipeline. During inference, lung and skin segmentation masks are extracted from a CT volume (left), after which a region encompassing the thorax is extracted (middle) and fed into the IPRO model (right). 3D, three-dimensional; CT, computed tomography; IPRO, imaging-based prognostication.

analysis,[33] to date, it has also been used to diagnose pancreatic cancer using magnetic resonance imaging.[34] We expand on related work, which detects and classifies malignancies in lung cancer screening chest CTs,[35] by predicting mortality risk for patients with confirmed lung cancer diagnoses and incorporating a wider range of pretreatment CTs from multiple data sets and sites.

Experimental Setup

To train and validate IPRO, we performed a five-fold cross validation[36] across six lung cancer data sets. This involved randomly splitting the data into five groups, while ensuring class balance based on survival status and TNM staging distribution. We then iteratively withheld each group for validation while training on the remaining four groups until each group was used for validation (Data Supplement). Models were trained to predict mortality as posterior probabilities between 0 (low-risk) and 1 (high-risk) at time t, given 3D CT volumes, where $t = 1$, 2, or 5 years. We report the average performance for each model across five folds and the standard deviation between folds in the Data Supplement.

To compare the prognostic power of IPRO to that of TNM staging, generalized linear regression models were trained using solely TNM staging information in the same 5-fold cross-validation to predict t-year mortality. The glm library in R was used for training and predicting regression models on seven TNM subtypes (IA, IB, IIA, IIB, IIIA, IIIB, and IV).

Finally, the Ensemble models that combined IPRO and TNM staging were generated by training a linear regression model per fold, where the inputs were TNM staging and IPRO mortality risk scores at time t.

Statistical Analysis

To compare risk scores with survival status at time t, we report concordance index (C-index) and area under the receiver operating characteristic curve (AUC). Pearson r^2 correlations between IPRO scores and time-to-event from date of CT acquisition are also reported. Statistical significance between models was assessed using a two-sample two-tailed t-test.

IPRO was used to stratify patients with lung cancer, similar to an approach adopted by Becker et al,[37] where Kaplan-Meier curves were generated per risk group. Each group is defined as a subset of the patients in the validation set sorted by ascending IPRO mortality risk scores. To quantify differences between predicted highest- and lowest-risk groups defined as quintiles (ie, 20%) or deciles (ie, 10%) of the patients with either the highest or lowest IPRO scores, we used the coxph function to report hazard ratio (HR) and log-rank P-values. All statistical analyses were performed in R.

Model Understanding

As the black box nature of CNNs limits our ability to understand why it made a particular outcome prediction, we attempt to understand the algorithm by exploring associations between the outcome predictions and known prognostic clinical variables like age, sex, TNM stage, and histology across IPRO's risk deciles. We also show Gradient-weighted Class Activation Mapping (GradCAM)[38] activation maps as a visual explanation to indicate on which anatomical regions within the thorax IPRO placed attention to generate its mortality risk prediction. Such visualizations, although preliminary, offer insight into a subset of the features learned in the 3DCNN and attempt to shine a light into the black box of deep learning–based predictions. We defined our regions of interest by volumetrically segmenting measurable primary lesions in a subset of 358 CTs in the validation set. For each CT scan in this subset, we then quantify the model's average attention placed on the primary lesions and compare this to the model's average attention throughout the entire thorax. Additional information about the lesion segmentation and attention quantification is provided in the Extended Methods section in the Data Supplement.

RESULTS

Mortality Risk Prediction

IPRO showed evidence of similar prognostic power compared with that of TNM staging in predicting 1-year, 2-year, and 5-year mortality (Fig 2). Although the average C-index was lower for IPRO compared with TNM for 2- and 5-year
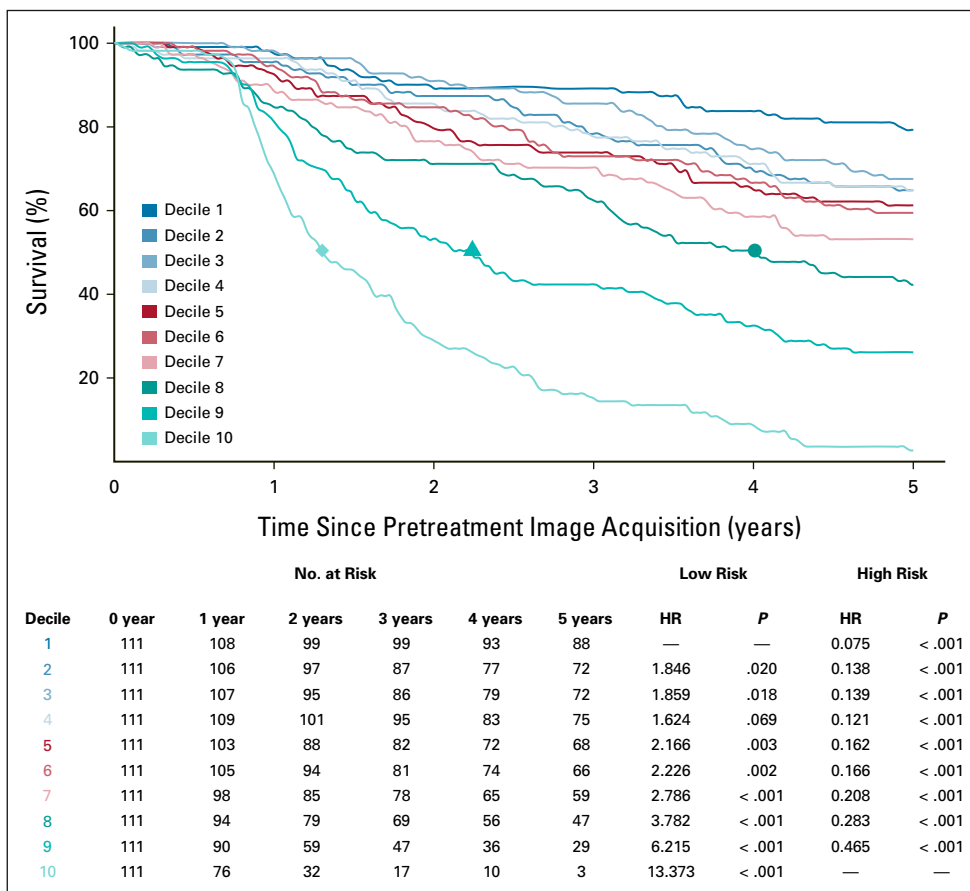
FIG 2. C-index scores for the three models (TNM, IPRO, and Ensemble) that predict lung cancer mortality risk at 1 year, 2 years, and 5 years. Error bars denote standard deviation between folds. C-index, concordance index; IPRO, imaging-based prognostication.
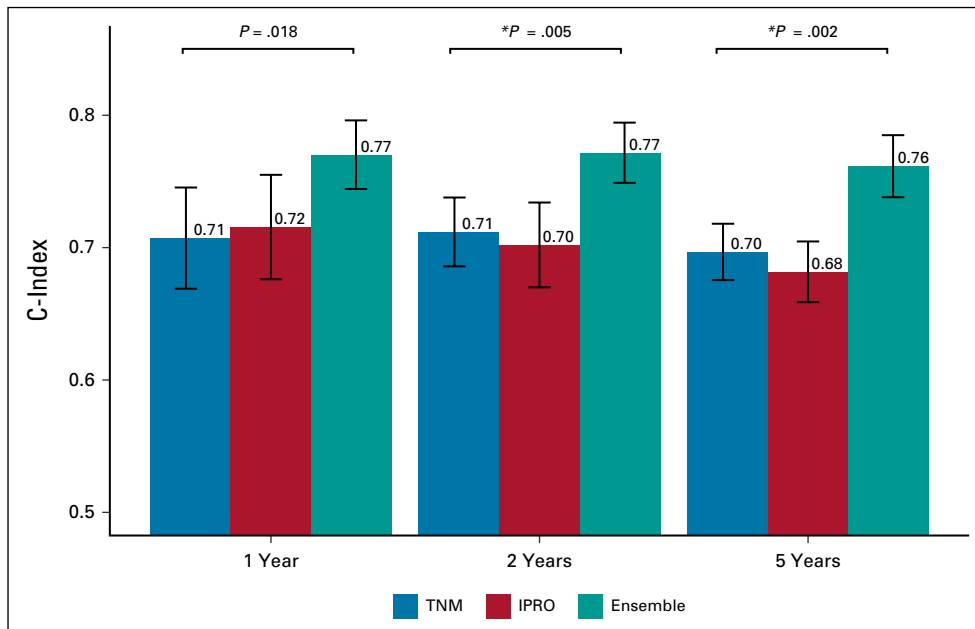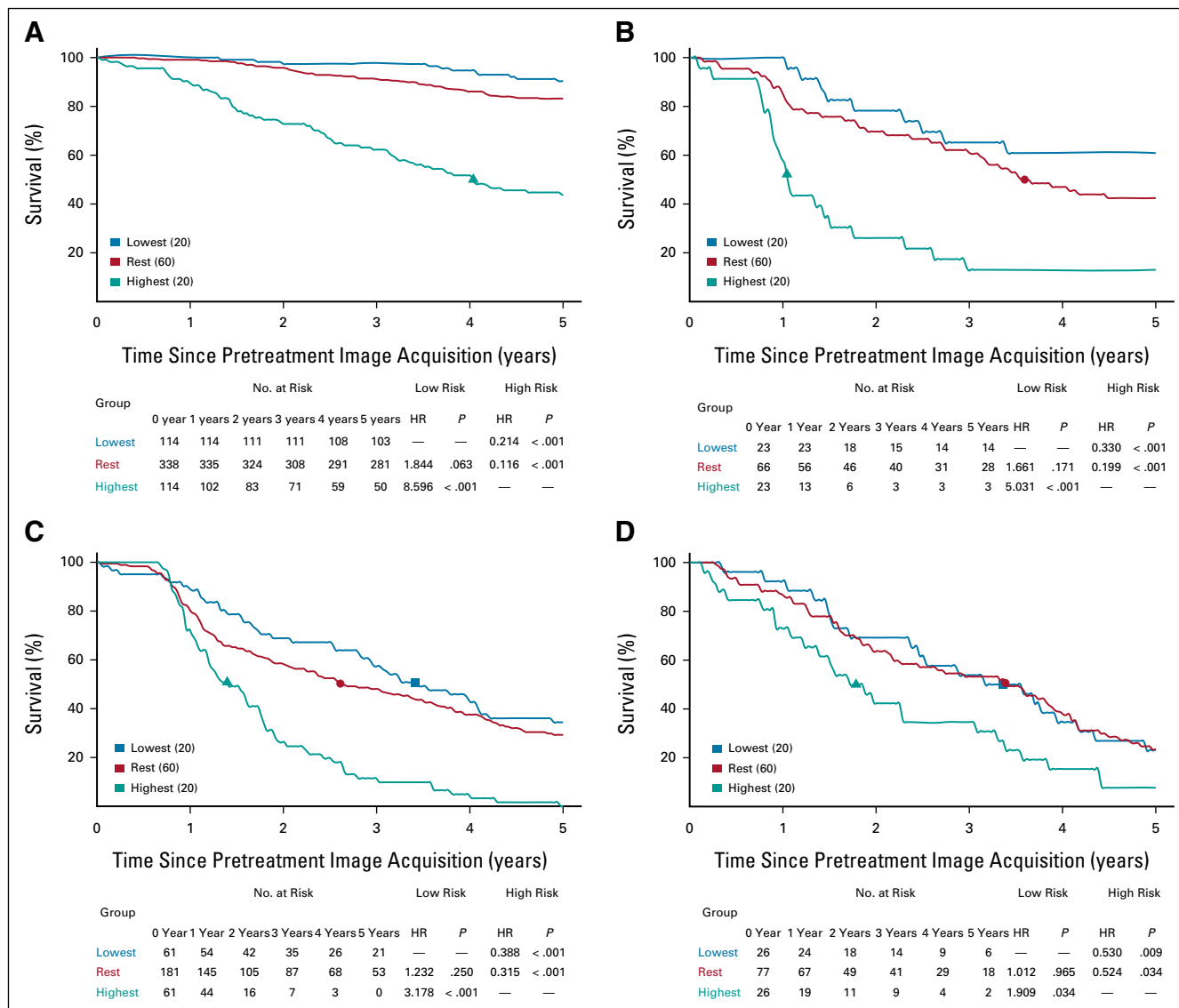
mortality, standard deviations suggest performance is similar for both TNM and IPRO (Data Supplement). When stage IV patients were removed from the validation set, average C-index for IPRO increased to 0.72 for 5-year mortality risk assessment (Data Supplement). The Ensemble model, which combines IPRO and TNM staging information, yielded significantly superior prognostic performance at 2-year and 5-year time intervals ($P \leq .01$) when

FIG 3. Kaplan-Meier curves for 5-year IPRO mortality risk deciles for all TNM stage groups. HRs, *P*-values, and number of patients at risk per year are provided for each decile. HR, hazard ratio; IPRO, imaging-based prognostication.

			No. at Risk				Low Risk		High Risk	
Decile	0 year	1 year	2 years	3 years	4 years	5 years	HR	P	HR	P
1	111	108	99	99	93	88	—	—	0.075	< .001
2	111	106	97	87	77	72	1.846	.020	0.138	< .001
3	111	107	95	86	79	72	1.859	.018	0.139	< .001
4	111	109	101	95	83	75	1.624	.069	0.121	< .001
5	111	103	88	82	72	68	2.166	.003	0.162	< .001
6	111	105	94	81	74	66	2.226	.002	0.166	< .001
7	111	98	85	78	65	59	2.786	< .001	0.208	< .001
8	111	94	79	69	56	47	3.782	< .001	0.283	< .001
9	111	90	59	47	36	29	6.215	< .001	0.465	< .001
10	111	76	32	17	10	3	13.373	< .001	—	—

FIG 4. Stage-specific Kaplan-Meier curves for 5-year IPRO mortality risk quintiles: (A) stage I, (B) stage II, (C) stage III, and (D) stage IV. IPRO, imaging-based prognostication.

compared with that of TNM alone (Fig 2). The Data Supplement summarizes the results across metrics including C-index, area under the curve, and Pearson r^2.

Patient Stratification

Kaplan-Meier curves in Figure 3 show risk stratification by IPRO deciles of all patients with lung cancer (stages I-IV) included in the 5-year validation. HRs between each decile and the highest risk group (ie, decile 10) were statistically significant ($P < .001$). HRs between each decile and the lowest risk decile (ie, decile 1) were statistically significant for deciles ≥ 6. Kaplan-Meier curves illustrating the 1-year and 2-year IPRO deciles are shown in the Data Supplement.

We assessed IPRO's ability to stratify patients within each TNM stage via high-risk and low-risk quintiles (Fig 4).

Stage I patients in the highest risk IPRO quintile had a 8.6-fold (95% CI, 4.5 to 16.3; $P < .001$) increased 5-year mortality hazard compared with stage I patients in the lowest-risk quintile. Similarly, in stage II and stage III, patients in the highest-risk IPRO quintile had a 5.0-fold (95% CI, 2.3 to 11.2; $P < .001$) and 3.2-fold (95% CI, 2.1 to 4.8; $P < .001$) increased 5-year mortality hazard compared with stage II and stage III patients in the lowest risk quintile, respectively. Across all TNM stages, the weakest patient stratification existed for stage IV patients where the highest risk IPRO quintile had a 1.9-fold (95% CI, 1.1 to 3.5; $P = .033$) increased 5-year mortality hazard compared with stage IV patients in the lowest risk quintile. Kaplan-Meier curves by TNM stage illustrating the 1-year and 2-year IPRO quintiles are shown in the Data Supplement.

TABLE 1. Distribution of Known Prognostic Factors by IPRO Risk Decile Including Age, Sex, TNM Stage and Histology

IPRO Risk Decile	Median Age	Sex[a] (Male/Female)	Stage				Histology[a]			
			I	II	III	IV	SqCC	AC	LCC	Other
1 (low)	62	35/76	82	5	11	13	12	51	3	45
2	64	52/59	68	3	22	18	33	36	4	38
3	63	50/59	70	7	18	16	23	45	3	38
4	63	62/48	61	7	27	16	30	41	5	34
5	63	76/35	65	7	22	17	24	54	4	29
6	64	73/35	58	15	23	15	27	45	6	30
7	66	90/16	50	15	31	15	31	44	5	26
8	65	87/23	53	15	30	13	34	41	4	31
9	68	75/30	42	23	40	6	34	34	13	24
10 (high)	68	73/36	17	15	79	0	40	20	26	23

Abbreviations: AC, adenocarcinoma; IPRO, Image-based prognostication; LCC, large cell carcinoma; SCC, small cell carcinoma; SqCC, squamous cell carcinoma.

[a]Excludes 20 patients that are missing sex and histology information and 120 patients with SCC.

Model Understanding

To further explore IPRO's 5-year mortality predictions, we assessed the distribution of known prognostic variables including age, sex, TNM stage, and histology across the IPRO risk deciles (Table 1). Comparing the characteristics of patients IPRO deemed lowest risk (decile 1) with those deemed highest risk (decile 10), the median age increases from 62 to 68 years and the sex composition shifts from 31.5% male in the lowest-risk patients to 67.0% male in the highest-risk patients. The most common histologic subtype in patients comprising the lowest-risk decile was adenocarcinoma (46%), whereas squamous cell carcinoma (37%) and large cell carcinoma (24%) accounted for the majority of highest-risk patients. Patients with lung cancer diagnosed as TNM stages I and II account for 78.4% of patients in the lowest-risk decile but only 29.4% of patients in the highest-risk decile.

GradCAM activation maps indicated that IPRO learned to place outsized attention on primary lesions. On average, 54% more attention was placed on primary lesions (0.245) compared with the average attention throughout the thorax (0.159), which was statistically significant ($P < .001$). Finally, we reviewed GradCAM activation maps to qualitatively assess on which anatomical regions within the thorax IPRO placed attention to generate the 5-year mortality risk prediction. In Figure 5, three representative sample cases are provided depicting areas that received the greatest attention (red) and the least attention (blue). Hand-drawn white ellipses (not available to IPRO) denote areas containing primary lesions.

DISCUSSION

Staging classification systems enable physicians to communicate information about an individual tumor or group of tumors in a standardized way.[39] The anatomic extent of disease is a major factor affecting prognosis and informs the appropriate treatment selection. We demonstrate that deep learning may provide additional prognostic insight based on both known and unknown features present in pretreatment CTs in a quantifiable and continuous manner. Our end-to-end fully automated framework, IPRO, was designed to ingest CTs of varying sources and imaging protocols and automatically analyze the 3D region encompassing the thorax. IPRO predicted mortality consistently and accurately at 1-year, 2-year, and 5-year time intervals, and generated similar prognostic performance to TNM staging, while also stratifying patients across and within TNM stages. By combining IPRO with TNM, the Ensemble model showed improved performance across all time intervals, suggesting that IPRO-derived imaging features may complement human-derived features.

Subject to further prospective validation, IPRO could have the potential to guide treatment intensification and de-escalation strategies. Akin to molecular biomarkers that have revolutionized the approach to systemic therapies in lung cancer, the use of prognostic imaging biomarkers may help personalize treatment decisions. For example, stage I patients deemed high risk have a prognosis similar to stage II patients and may benefit from the addition of neoadjuvant and/or adjuvant systemic therapy, which is currently not the standard of care. TNM staging captures the extent of disease in the entire body, including metastatic disease, whereas IPRO explores only the chest. This likely explains why IPRO does not stratify stage IV patients as well as stage I-III patients (Fig 4). Although IPRO may be better suited for early-stage lung cancer prognostication,[40] where tumor burden is expected to be localized within the chest, we demonstrated that combining IPRO with TNM provided a better prognostic prediction in all three time points. Given the variability and the limited data set used in this study, we
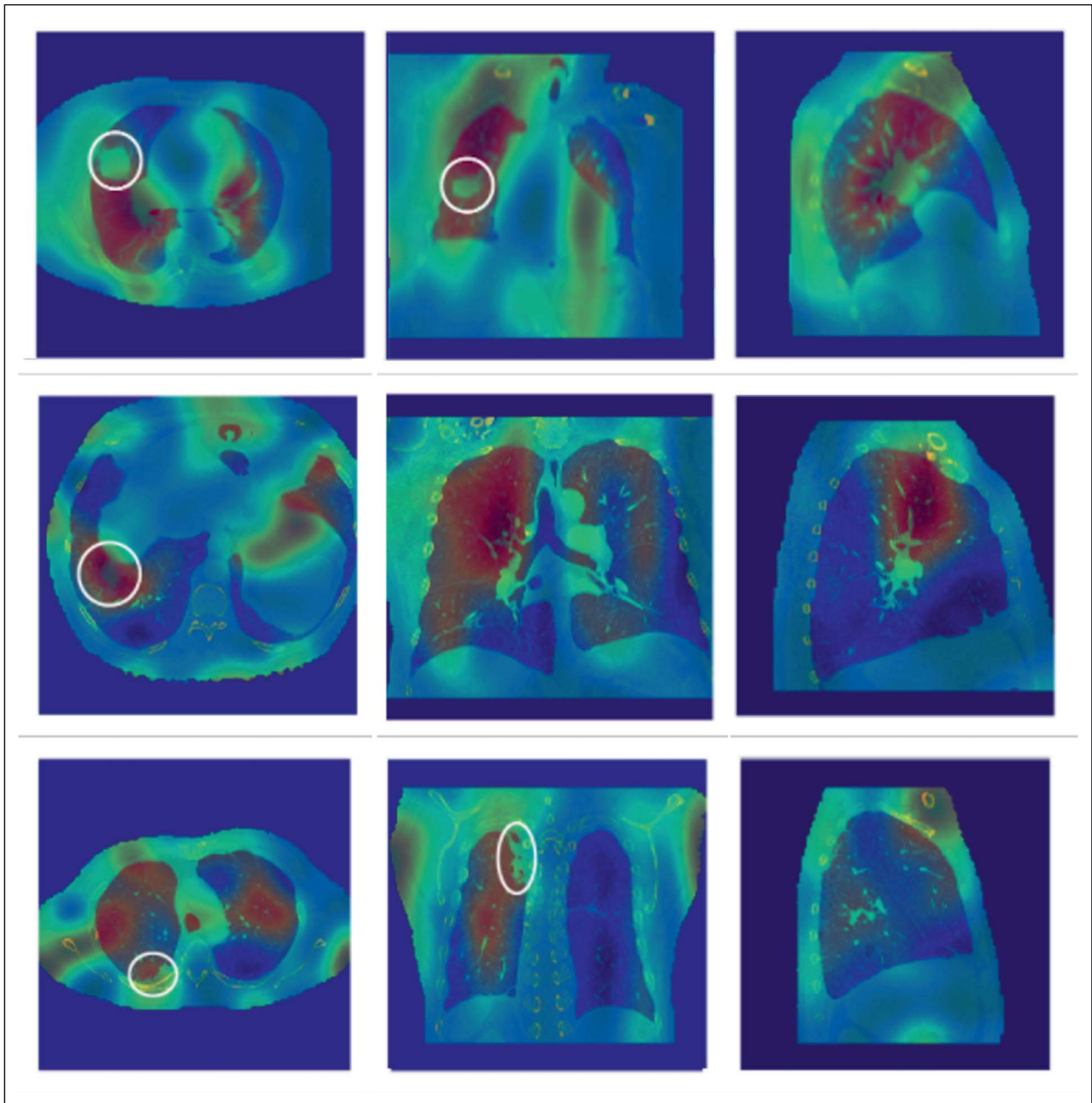
FIG 5. GradCAM attention heatmap for patients in the highest IPRO mortality risk group (decile 10). Red denotes areas that received greatest attention from the model. The white circle denotes the region of the primary tumor. GradCAM, Gradient-weighted Class Activation Mapping; IPRO, imaging-based prognostication.

expect that, when examined using a larger data set, performance may increase.

In our model development, we enabled IPRO to consider prognostic features across the entire thorax, eliminating the need for radiologists to manually annotate regions of interest, such as primary tumors. Manual annotation is a time-consuming process, requires radiologic expertise, is subject to inter-reader variability,[41,42] and enforces the assumption that only annotated regions of interest are correlated with outcomes. However, not defining regions of interest as the model input limits our ability to understand

what features IPRO considered to generate predictions. In reviewing regions of the CT volume that received the greatest attention by IPRO (Fig 5), our preliminary results indicate that IPRO gravitated toward tissue comprising primary lesions, suggesting that IPRO learned that this tissue has prognostic value. Peritumoral areas also received attention (Fig 5), suggesting that such areas may hold additional prognostic insight.[18-20] Evaluating distributions of known prognostic variables such as age and sex for patients across IPRO's predicted risk groups (Table 1) revealed that patients predicted to be in the highest-risk group (decile

10) were on average age 6 years older and mostly male compared with those predicted to be in the lowest-risk group (decile 1). Histology subtypes in decile 10 were also more likely to exhibit large cell carcinoma and squamous cell carcinoma subtypes, consistent with findings from previous studies that these subtypes are associated with worse outcomes.[43,44]

Prognostic features unrelated to cancerous tissue, such as coronary artery calcification, size of the heart, body composition, or pulmonary emphysema, may have been learned and considered by the model. This, however, remains speculative given the black box nature of deep neural networks, and further developments in machine learning techniques are required to explore this hypothesis. We are currently training and evaluating region-specific 3DCNNs to better understand the anatomic origins of IPRO's predictions.

Although IPRO shows great potential, further improvements may prove advantageous for lung cancer prognostication. Treatment decisions are usually based on the extent of disease in addition to patient, tumor, and environmental factors.[39] It is yet to be determined how IPRO will perform when additional clinical variables like pulmonary function and treatment type are introduced into the model. Future work will explore how clinical variables and longitudinal imaging can be incorporated to further refine prognostic accuracy and derive actionable insights.

We note there are some limitations to this study. First, this is a retrospective study, and more work is needed to validate predictions on larger, independent data sets, as well as in prospective studies. Second, our models were validated on pretreatment CTs and, therefore, treatment decisions were not considered. The impact of treatment on survival is certainly a factor to be considered in future studies, particularly given the recent advances in systemic treatment for advanced lung cancer. Also, we used inconsistent clinical and pathologic TNM staging as a comparator, which is agnostic in relation to comorbidities and treatment effectively applied. Third, a significant proportion of our patients were from the National Lung Screening Trial,[24] and therefore, more heavily skewed toward earlier-stage cancers and patients age 55-74 years who have 30 or more pack-years of smoking history. Finally, clinical TNM staging was not available for all CTs in the available data sets, limiting the total number of patients included in this study.

In conclusion, deep learning applied to pretreatment CTs combined to TNM staging may enhance prognostication and risk stratification in patients with lung cancer. Prospective testing and validation of the IPRO model in a large population of patients with lung cancer is warranted.

AFFILIATIONS

[1]Joint Department of Medical Imaging, Toronto General Hospital, Department of Medical Imaging, University of Toronto, Toronto, ON, Canada

[2]Altis Labs, Inc, Toronto, ON, Canada

[3]Princess Margaret Cancer Centre, Department of Radiation Oncology, University of Toronto, Toronto, ON, Canada

[4]Division of Thoracic Surgery, University Health Network and University of Toronto, Toronto, ON, Canada

[5]Artificial Intelligence in Medicine (AIM) Program, Mass General Brigham, Harvard Medical School, Boston, MA

[6]Department of Radiation Oncology, Dana Farber Cancer Institute and Brigham and Women's Hospital, Boston, MA

[7]Department of Medical Oncology and Hematology, Princess Margaret Cancer Centre, University Health Network, Toronto, ON, Canada

[8]Department of Medicine, University of Toronto, Toronto, ON, Canada

CORRESPONDING AUTHOR

Felipe Soares Torres, MD, PhD, Toronto General Hospital, 585 University Ave, 1PMB 274, Toronto, ON M5G2N2, Canada; e-mail: felipe.torres@uhn.ca.

DISCLAIMER

The statements contained herein are solely those of the authors and do not represent or imply concurrence or endorsement by NCI. The results published here are in whole or part based upon data generated by the TCGA Research Network: http://cancergenome.nih.gov/.

EQUAL CONTRIBUTION

F.S.T. and S.A. contributed equally to this work.

AUTHOR CONTRIBUTIONS

Conception and design: Felipe Soares Torres, Shazia Akbar, Felix Baldauf-Lenschen

Financial support: Felix Baldauf-Lenschen

Administrative support: Felix Baldauf-Lenschen

Collection and assembly of data: Shazia Akbar, Carola Schmidt, Felix Baldauf-Lenschen

Data analysis and interpretation: Felipe Soares Torres, Shazia Akbar, Srinivas Raman, Kazuhiro Yasufuku, Srinivas Raman, Kazuhiro Yasufuku, Carola Schmidt, Ahmed Hosny, Felix Baldauf-Lenschen, Natasha B. Leighl

Manuscript writing: All authors

Final approval of manuscript: All authors

Accountable for all aspects of the work: All authors

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/cci/author-center.

ACKNOWLEDGMENT

REFERENCES

1. Eisenhauer EA, Therasse P, Bogaerts J, et al: New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). Eur J Cancer 45:228-247, 2009

2. Yoon SH, Kim KW, Goo JM, et al: Observer variability in RECIST-based tumour burden measurements: A meta-analysis. Eur J Cancer 53:5-15, 2016

3. William WN, Pataer A, Kalhor N, et al: Computed tomography RECIST assessment of histopathologic response and prediction of survival in patients with resectable non–small-cell lung cancer after neoadjuvant chemotherapy. J Thorac Oncol 8:222-228, 2013

4. Choi HS, Jeong BK, Jeong H, et al: Application of the new 8th TNM staging system for non-small cell lung cancer: Treated with curative concurrent chemoradiotherapy. Radiat Oncol. 12:122, 2017

5. Travis WD, Asamura H, Bankier AA, et al: The IASLC lung cancer staging project: Proposals for coding T categories for subsolid nodules and assessment of tumor size in part-solid tumors in the forthcoming eighth edition of the TNM classification of lung cancer. J Thorac Oncol 11:1204-1223, 2016

6. Detterbeck FC, Nicholson AG, Franklin WA, et al: The IASLC lung cancer staging project: Summary of proposals for revisions of the classification of lung cancers with multiple pulmonary sites of involvement in the forthcoming eighth edition of the TNM classification. J Thorac Oncol 11:639-650, 2016

7. Carter BW, Lichtenberger III JP, Benveniste MK, et al: Revisions to the TNM staging of lung cancer: Rationale, significance, and clinical application. RadioGraphics 383:374-391, 2018

8. Fonseca A, Detterbeck FC: How many names for a rose: Inconsistent classification of multiple foci of lung cancer due to ambiguous rules. Lung Cancer 85:7-11, 2014

9. van Riel SJ, Sánchez C, Bankier AA: Observer variability for classification of pulmonary nodules on low-dose CT images and its effect on nodule management. Radiology 277:863-871, 2015

10. Lui H, Li L, Wormstone M, et al: Development and validation of a deep learning system to detect glaucomatous optic neuropathy using fundus photographs. JAMA Ophthalmol 137:1353-1360, 2019

11. Esteva A, Kuprel B, Novoa RA, et al: Dermatologist-level classification of skin cancer with deep neural networks. Nature 542:115-118, 2017

12. Yoo H, Kim KH, Singh R, et al: Validation of a deep learning algorithm for the detection of malignant pulmonary nodules in chest radiographs. JAMA Netw Open. 3:e2017135, 2020

13. Baek S, He Y, Allen BG, et al: Deep segmentation networks predict survival of non-small cell lung cancer. Scientific Rep 9, 17286, 2019

14. Afshar P, Mohammadi A, Tyrrell PN, et al: DRTOP: Deep learning-based radiomics for the time-to-event outcome prediction in lung cancer. Scientific Rep 10, 12366, 2020

15. Hosny A, Parmar C, Coroller TP, et al: Deep learning for lung cancer prognostication: A retrospective multi-cohort radiomics study. PLoS Med 15: e1002711, 2018

16. Wang S, Liu Z, Chen X, et al: Unsupervised deep learning features for lung cancer overall survival analysis. 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, July 18-21, 2018

17. Kim H, Goo JM, Lee KH, et al: Preoperative CT-based deep learning model for predicting disease-free survival in patients with lung adenocarcinomas. Radiology 296: 192764, 2020

18. Dou TH, Coroller TP, van Griethuysen JJM, et al: Peritumoral radiomics features predict distant metastasis in locally advanced NSCLC. PLoS One 13: e0206108, 2018

19. Pérez-Morales J, Tunali I, Stringfield O, et al: Peritumoral and intratumoral radiomic features predict survival outcomes among patients diagnosed in lung cancer screening. Scientific Rep 10:1-15, 2020

20. She Y, Jin Z, Wu J, et al: Development and validation of a deep learning model for non–small cell lung cancer survival. JAMA Netw Open. 3:e205842, 2020

21. Guo H, Kruger H, Wang G, et al: Knowledge-based analysis for mortality prediction from CT images. IEEE J Biomed Health Inform 24:457-464, 2020

22. Martin L, Birdsell L, Macdonald N, et al: Cancer cachexia in the age of obesity: Skeletal muscle depletion is a powerful prognostic factor, independent of body mass index. J Clin Oncol 31:1539-1547, 2013

23. Prado CMM, Lieffers JR, McCargar LJ, et al: Prevalence and clinical implications of sarcopenic obesity in patients with solid tumours of the respiratory and gastrointestinal tracts: A population-based study. Lancet Oncol 9:629-635, 2008

24. National Lung Screening Trial Research Team, Aberle DR, Berg CD, et al: The National Lung Screening Trial: Overview and study design. Radiology 258:243-253, 2011

25. Aerts H, Wee L, Rios Velazquez E, et al: Data from NSCLC-radiomics [data set]. Cancer Imaging Arch, 2019. doi 10.7937/K9/TCIA.2015.PF0M9REI

26. Bakr S, Gevaert O, Echegaray S, et al: Data for NSCLC radiogenomics collection. Cancer Imaging Arch, 2017. doi 10.7937/K9/TCIA.2017.7hs46erv

27. Grove O, Berglund AE, Schabath MB, et al: Data from: Quantitative computed tomographic descriptors associate tumor shape complexity and intratumor heterogeneity with prognosis in lung adenocarcinoma. Cancer Imaging Arch, 2015. doi 10.7937/K9/TCIA.2015.A6V7JIWX

28. Kirk S, Lee Y, Kumar P, et al: Radiology data from The Cancer Genome Atlas Lung Squamous Cell Carcinoma [TCGA-LUSC] collection. Cancer Imaging Arch, 2016. doi 10.7937/K9/TCIA.2016.TYGKKFMQ

29. Albertina B, Watson M, Holback C, et al: Radiology data from The Cancer Genome Atlas Lung Adenocarcinoma [TCGA-LUAD] collection. The Cancer Imaging Arch, 2016. doi 10.7937/K9/TCIA.2016.JGNIHEP5

30. Deng J, Dong W, Socher R, et al: ImageNet: A large-scale image database. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Miami, FL, June 20-25, 2009

31. Szegedy A, Liu W, Jia Y, et al: Going deeper with convolutions. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, June 7-12, 2015

32. Shin H-C, Roth HR, Gao M, et al: Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans Med Imaging 35:1285-1298, 2016

33. Carreira J, Zisserman A: Quo vadis, action recognition? A new model and the kinetics dataset. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, July 21-26, 2017

34. LaLonde R, Tanner I, Nikiforaki K, et al: INN: Inflated neural networks for IPMN diagnosis, in Shen D, Yap P-T, Liu T, et al (eds): Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. Lecture Notes in Computer Science, Volume 11768, 2019, pp 101-109

35. Ardila D, Kiraly AP, Bharadwaj S, et al: End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nat Med 25:954-961, 2019

36. Stone M: Cross-validatory choice and assessment of statistical predictions. J R Stat Soc 36:111-147, 1974

37. Becker T, Weberpals J, Jegg AM, et al: An enhanced prognostic score for overall survival of patients with cancer derived from a large real-world cohort. Ann Oncol 31:1561-1568, 2020

38. Selvaraju RR, Cogswell M, Das A, et al: Grad-CAM: Visual explanations from deep networks via gradient-based localization. IEEE International Conference on Computer Vision (ICCV), Venice, Italy, October 22-29, 2017

39. Detterbeck F: Stage classification and prediction of prognosis: Difference between accountants and speculators. J Thorac Oncol. 8:820-822, 2013

40. Torres FS, Akbar S, Raman S, et al: Automated imaging-based stratification of early stage lung cancer patients prior to receiving surgical resection using deep learning applied to CTs. J Clinic Oncol 39, 2021 (suppl 15; abstr 1552)

41. McErlean A, Panicek DM, Zabor EC, et al: Intra- and interobserver variability in CT measurements in oncology. Radiology 269:451-459, 2013

42. Singh S, Pinsky P, Fineberg NS, et al: Evaluation of reader variability in the interpretation of follow-up CT scans at lung cancer screening. Radiology 259: 263-270, 2011

43. Iyoda A, Haroshima K, Moriya Y, et al: Prognostic impact of large cell neuroendocrine histology in patients with pathologic stage Ia pulmonary non–small cell carcinoma. J Thorac Cardiovasc Surg 132:312-315, 2006

44. Fukui T, Taniguchi T, Kawaguchi K, et al: Comparisons of the clinicopathological features and survival outcomes between lung cancer patients with adenocarcinoma and squamous cell carcinoma. Gen Thorac Cardiovasc Surg 63:507-513, 2015