Article

Lung Cancer Recurrence Risk Prediction through Integrated Deep Learning Evaluation

Peng Huang [1,2,3,*], Peter B. Illei [3,4], Wilbur Franklin [5], Pei-Hsun Wu [6,7,8], Patrick M. Forde [1,3,9], Saeed Ashrafinia [10,11], Chen Hu [1,2,3], Hamza Khan [12], Harshna V. Vadvala [10], Ie-Ming Shih [3,4], Richard J. Battafarano [3,12], Michael A. Jacobs [3,10,13], Xiangrong Kong [2,14], Justine Lewis [15], Rongkai Yan [10], Yun Chen [16], Franck Housseau [1,3,9], Arman Rahmim [10,11,17], Elliot K. Fishman [1,3,10,12], David S. Ettinger [1,3], Kenneth J. Pienta [1,3], Denis Wirtz [1,3,6,7,8], Malcolm V. Brock [3,12], Stephen Lam [17,†] and Edward Gabrielson [3,4,†]

1   Department of Oncology, Johns Hopkins University, Baltimore, MD 21205, USA
2   Department of Biostatistics, Johns Hopkins University, Baltimore, MD 21205, USA
3   The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, MD 21205, USA
4   Department of Pathology, Johns Hopkins University, Baltimore, MD 21287, USA
5   Department of Pathology, University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA
6   Johns Hopkins Physical Sciences Oncology Center, Baltimore, MD 21218, USA
7   Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, MD 21218, USA
8   Institute for NanoBioTechnology, Johns Hopkins University, Baltimore, MD 21218, USA
9   Bloomberg–Kimmel Institute for Cancer Immunotherapy, Johns Hopkins University, Baltimore, MD 21205, USA
10 Department of Radiology, Johns Hopkins University, Baltimore, MD 21218, USA
11 Department of Electrical Engineering, Johns Hopkins University, Baltimore, MD 21218, USA
12 Department of Surgery, Johns Hopkins University, Baltimore, MD 21218, USA
13 Department of Diagnostic and Interventional Imaging, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA
14 Department of Ophthalmology, Johns Hopkins University, Baltimore, MD 21218, USA
15 Intensive Care Unit, Howard University College of Medicine, Washington, DC 20059, USA
16 Department of Mechanical Engineering, Johns Hopkins University, Baltimore, MD 21218, USA
17 BC Cancer Research Institute, University of British Columbia, Vancouver, BC V5Z 1L3, Canada
*   Correspondence: phuang12@jhmi.edu; Tel.: +1-410-502-0944
†   These authors contributed equally to this work.

Simple Summary: Few significant advances have been made over recent decades in predicting lung cancer progression risk after complete surgical removal of tumor in stage IA non-small-cell lung cancers (NSCLCs). Although several biomarkers have shown some predictive value, it is unclear whether these markers add value to traditional TNM staging. We developed an integrated deep learning evaluation (IDLE) score to combine patient's preoperative lung CT image findings and postoperative pathologic assessment and found that this score can better predict cancer progression risk than TNM staging and tumor grade. Improved predictive value of the IDLE score was primarily due to the complementary use of tumor measurements in CT images from an entire lung as well as microscopic tissue characteristics. Our findings suggest that integrating measurements from different aspects of tumor morphology is more robust for increasing prediction accuracy than building on the measurements of similar aspects of tumor morphology.

Abstract: *Background*: Prognostic risk factors for completely resected stage IA non-small-cell lung cancers (NSCLCs) have advanced minimally over recent decades. Although several biomarkers have been found to be associated with cancer recurrence, their added value to TNM staging and tumor grade are unclear. *Methods:* Features of preoperative low-dose CT image and histologic findings of hematoxylin- and eosin-stained tissue sections of resected lung tumor specimens were extracted from 182 stage IA NSCLC patients in the National Lung Screening Trial. These features were combined to predict the risk of tumor recurrence or progression through integrated deep learning evaluation (IDLE). Added values of IDLE to TNM staging and tumor grade in progression risk prediction and risk stratification were evaluated. *Results:* The 5-year AUC of IDLE was 0.817 ± 0.037 as compared

to the AUC = 0.561 ± 0.042 and 0.573 ± 0.044 from the TNM stage and tumor grade, respectively. The IDLE score was significantly associated with cancer recurrence ($p < 0.0001$) even after adjusting for TNM staging and tumor grade. Synergy between chest CT image markers and histological markers was the driving force of the deep learning algorithm to produce a stronger prognostic predictor. *Conclusions:* Integrating markers from preoperative CT images and pathologist's readings of resected lung specimens through deep learning can improve risk stratification of stage 1A NSCLC patients over TNM staging and tumor grade alone. Our study suggests that combining markers from nonoverlapping platforms can increase the cancer risk prediction accuracy.

Keywords: postoperative-stage IA NSCLC; artificial intelligent; biomarker; computer-aided diagnosis; tumor grade

1. Introduction

With the implementation of lung cancer screening using low-dose computed tomography worldwide, the proportion of stage IA lung cancers has increased to ≥50% compared to <10% in clinically diagnosed patients [1–3]. Stage IA non-small-cell lung cancer (NSCLC) patients are primarily treated with surgery alone, and patients with tumors that are completely resected with negative margins typically do not receive additional treatment. However, despite complete resection of the original tumor, cancer progression (recurrence, metastasis, or death) occurs in about 20% of patients with stage IA NSCLCs [4], and 70–90% of those who progress die from their lung cancers [5–11]. Successes in improving the outcomes of patients with IB or higher-stage resectable NSCLCs using neoadjuvant or adjuvant targeted therapy or immunotherapy alone or in combination with chemotherapy [12–14] lead to an increased interest in identifying the subset of stage IA NSCLC patients who may benefit from these therapies to improve the cure rate further. Currently, there are no validated prognostic markers to guide adjuvant therapy for patients with resected stage IA NSCLCs.

Early identification of stage IA NSCLCs that have a high risk of progression has an important role for guiding physicians to proactively treat patients who may benefit from early cancer treatment to offset development of metastatic disease. While a number of promising blood or radiological biomarkers are found to be associated with NSCLC progression, none of them has shown significant added value over the well-established TNM staging and tumor grade criteria for stratification of lung cancer recurrence and mortality risk [15,16]. Since cancer progression is a complex process that involves multiple factors before and after surgery, we investigated integrated deep learning evaluation (IDLE) of tumor macro- and micromorphological characteristics for the potential to add value to TNM staging and tumor grade in identifying those stage IA NSCLCs that are at a high risk of cancer progression.

Our report below is organized as follows. Section 2 describes how the study samples were selected, the method used to process images and extract features, the deep learning architecture, the statistical methods to evaluate predictors, and how we decoded the deep learning network's black box to discover synergies between the lung low-dose computed tomography (LDCT) image markers and surgical tissue characteristics. The results are presented in Section 3. Study implications are discussed in Section 4. Finally, our conclusions are summarized in Section 5. Throughout the paper, we interchangeably used the terms "feature" and "variable" to refer to a risk factor under investigation, and we used "progression" to refer to the competing event of lung cancer recurrence, metastasis, and lung cancer-associated death.

2. Materials and Methods

2.1. Study Sample

The National Lung Screening Trial (NLST) recruited 53,454 participants at a high risk of lung cancer between August 2002 and April 2004 from 10 clinical centers of the

American College of Radiology Imaging Network (ACRIN) and 23 clinical centers of the Lung Screening Study group (LSS) [17]. The participants were randomized to receive either three annual LDCT screenings (T0, T1, T2) or annual X-ray screenings. From the surgical resection specimens of newly diagnosed tumors at the ACRIN centers, tissue blocks containing a formalin-fixed paraffin-embedded tumor, lymph nodes, and adjacent and distant lung tissue were forwarded to a central NLST pathology core laboratory. Multiple slides were then cut from each block, stained with hematoxylin and eosin (H&E), and imaged using an Aperio ScanScope. We received preoperative LDCT screening images, H&E section images, pathologist's readings, and postoperative follow-up data from the NLST through several material transfer agreements. For this study, we selected NLST patients who met all the following criteria: (1) received primary lung tumor surgery; (2) had pathologist's readings collected by the NLST; (3) had pathological stage IA NSCLC diagnosis according to the American Joint Committee on Cancer's TNM staging system (8th edition) [18]; (4) had a largest pathological invasive tumor size no greater than 30 mm; and (5) had primary surgeries performed within 2 years after their last LDCT screening dates in the NLST study. A total of 182 patients met our selection criteria (Figure 1). From the pathology slides of these patients, 1076 pathologist-annotated regions of interest (ROIs) from 477 H&E images were identified, and 182 last preoperative LDCT lung screening images within 2 years of the primary tumor surgery were analyzed.
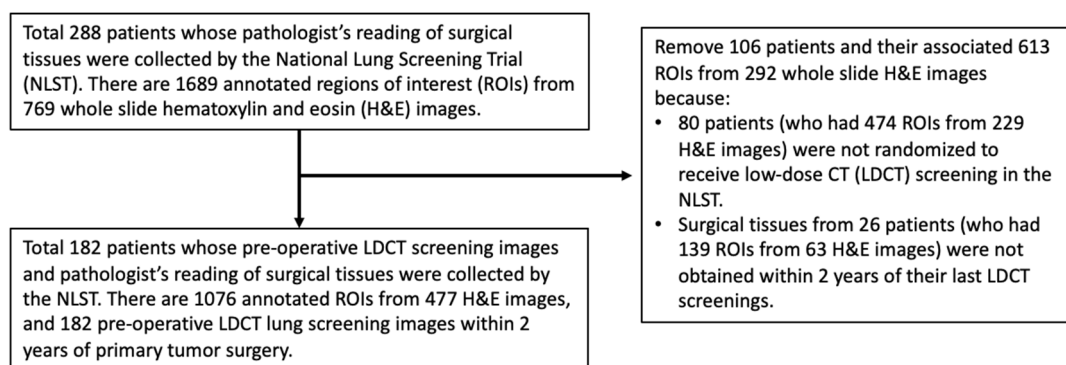
Figure 1. Study sample selection. Stage IA patients were selected from the National Lung Screening Trial who received primary tumor surgery within two years after their last LDCT screening date.

2.2. Tissue Image Feature Extraction

Images of hematoxylin- and eosin (H&E)-stained sections from the National Lung Screening Trial (NLST) were stored on a server located at the University of Colorado. All H&E images were reviewed, and the regions of interest were annotated by the NLST reference pathologist (W.F.). These ROIs were selected from invasive tumor regions, premalignant regions (if present), and nontumor regions, and the predetermined features for each ROI were entered into an online database. All slides were reviewed by multiple pathologists to confirm accuracy of pathology readings, and the operation of the central NLST pathology core laboratory was summarized by Patz et al. [19]. We extracted tissue features from pathologist's readings in all annotated ROIs associated with each individual patient. Five histologic subtype variables were created (Table S1) to classify tumors by diagnosis, which included adenocarcinoma in situ (AIS), invasive adenocarcinoma, carcinoid tumor, large-cell carcinoma, and squamous cell carcinoma. Table S1 lists all tissue features extracted from the annotated ROIs and surgical parameters.

2.3. Preoperative CT Image Feature Extraction

The CT scanner protocol parameters used in the NLST were 120 kVp, 40–80 mAs, detector collimation of 0.5–2.5 mm (for one data channel). All LDCT images were reconstructed through resampling and interpolation to have the same 0.5 mm × 0.5 mm × 0.5 mm voxel size. For the present study, we first segmented the whole lung to remove voxels from the

bone and chest wall. All noncalcified lesions with diameters $\geq$ 4 mm were included in the analysis. Surgically resected tumors were linked to pathologic diagnosis, and lesions without pathological diagnosis were treated as indeterminant. To segment a tumor (or a lesion), we first defined its gross volume using a 3D cube that covered the entire tumor with at least 1 cm margin. Next, we applied a median filter to remove image noise before applying morphological operations to define tumor's surface voxels. The median voxel intensity value from the adjacent normal lung tissue was subtracted from both tumor and peritumoral voxel values to normalize voxel intensities. We then added 300 to all voxel intensity values to make them non-negative, thus enabling radiomics energy features to quantify more effectively the lesion intensity distribution.

Similar to our prior published work [20], we extracted LDCT image radiomics and other texture features from intra-tumor, peri-tumor, and extra-tumor volumes of interest (VOIs) (Table S2). A nested approach was used to extract voxel intensity spatial distribution. For this approach, we first used all voxels within the segmented lesion to define its weighted center as follows:

$$C_{100} = (c_x, c_y, c_z) = \left(\frac{\sum_{i,j,k} x_{ijk} d_{ijk}}{\sum_{i,j,k} d_{ijk}}, \frac{\sum_{i,j,k} y_{ijk} d_{ijk}}{\sum_{i,j,k} d_{ijk}}, \frac{\sum_{i,j,k} z_{ijk} d_{ijk}}{\sum_{i,j,k} d_{ijk}} \right) \quad (1)$$

where the summation is over all tumor voxels with coordinates $(x_{ijk}, y_{ijk}, z_{ijk})$ and d_{ijk} is the voxel intensity value at location $(x_{ijk}, y_{ijk}, z_{ijk})$. The standard deviation of the tumor voxel spatial distribution was calculated by:

$$LocSd_{100} = \sqrt{ \frac{\sum_{ijk} d_{ijk}^2 \left[\left(x_{ijk} - c_x \right)^2 + \left(y_{ijk} - c_y \right)^2 + \left(z_{ijk} - c_z \right)^2 \right]}{\sum_{ijk} d_{ijk}^2} } \quad (2)$$

Since $LocSd_{100}$ could be affected by tumor volume, we used $LocSd_{100}V = LocSd_{100}/V$ to normalize $LocSd_{100}$ where V is the tumor volume. We repeated the same formula using 50% and 20% of the voxels with the highest voxel intensity values to obtain the corresponding weighted centers (C_{50} and C_{20}), $LocSd_{50}V$, and $LocSd_{20}V$. Distances between the centers were calculated by $d_{50} = \| C_{50} - C_{100} \|_2$ and $d_{20} = \| C_{20} - C_{100} \|_2$, where $\| \cdot \|_2$ is the Euclidean distance metric.

The 3D radiomics features were extracted using formulas from Aerts et al. [21]. We used fixed bin size 100 (in Hounsfield units) to calculate energy, root-mean-square (RMS), entropy, and uniformity features. The second-order gray-level co-occurrence matrices (GLCMs) were calculated by setting directions θ = 0, 45, 90, and 135 in sagittal, transverse, and coronal planes, respectively. This resulted in 13 GLCMs from directions defined by the following vectors:

$$\begin{array}{c} (1,0,0),\ (0,1,0),\ (1,1,0),\ (-1,1,0),\ (1,0,1),\ (1,1,1),\ (1,-1,1), \\ (0,0,1),\ (0,1,1),\ (0,-1,1),\ (-1,0,1),\ (-1,1,1),\ (-1,-1,1). \end{array} \quad (3)$$

For each of these 13 directions, the 22 GLCM texture features described in Table S2 were extracted. Their 3D texture feature values were then calculated using their average values over all 13 directions. For the second-order gray-level run-length matrix (GLRLM) feature extraction, we set the voxel distance parameter to d = 2, 3, 4, and 5, respectively, and used the 13 directions from (3) to calculate 52 (13 $\times$ 4) GLRLMs—one for each direction-by-distance combination. Eleven GLRLM texture features described in Table S2 were extracted from each GLRLM. Their 3D texture values were calculated using their average values over all 52 direction-by-distance combinations.

We used following features to compare intensity distributions between the voxels within the segmented lesions (or tumors) and the voxels from the peritumoral region:

$$\text{Mean intensity ratio} = \frac{\text{mean voxel intensity within the tumor}}{\text{mean voxel intensity within the peritumoral region}} \tag{4}$$

$$\text{Quantile ratio } R(q) = \frac{q\text{th voxel intensity quantile within the tumor}}{q\text{th voxel intensity quantile within the peritumoral region}} \tag{5}$$

where $q = 50$ and 90, respectively. A total of 173 features from each LDCT image were extracted. The procedure to process images and extract features is illustrated in Figure 2.
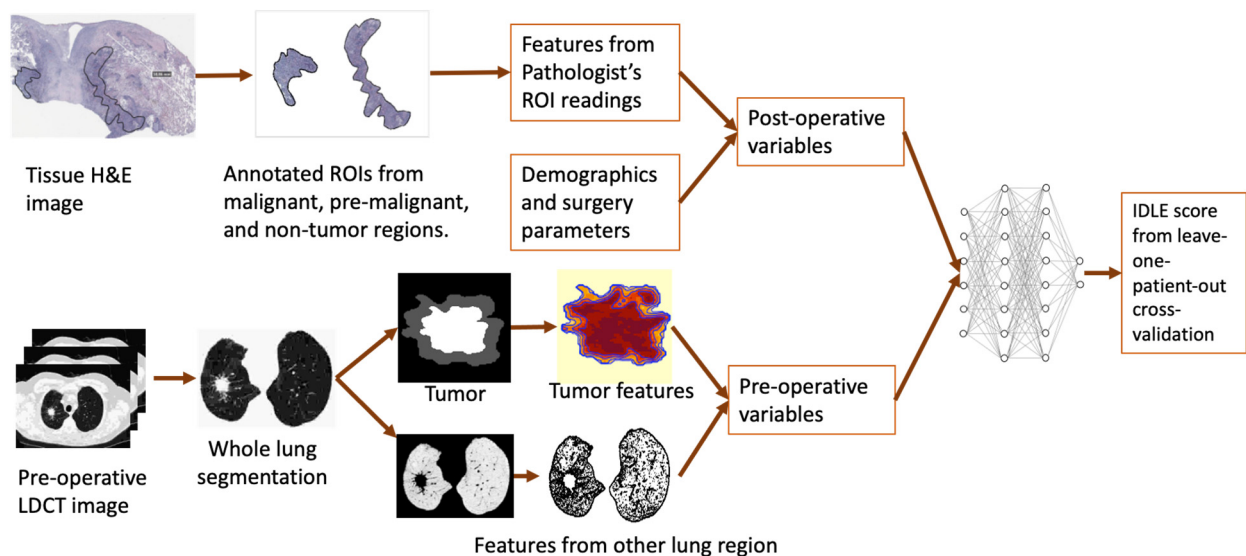
Figure 2. Study procedure. Postoperative variables include surgical tissue features extracted from pathologist's readings, patient demographics, and surgical parameters. Preoperative variables include CT image features extracted from the tumor region, the peritumoral region, other area of the lung, and the time interval between preoperative LDCT lung scanning and surgery. All these variables were used as input variables in the leave-one-patient-out cross-validation IDLE score computation.

2.4. Prediction Algorithm Development

We aimed to study the value of combining preoperative lung LDCT image texture features and histologic findings of resected tumors to predict the lung cancer progression risk using integrated deep learning evaluation (IDLE). The primary endpoint was progression-free survival defined as the time since the date of the initial primary tumor surgery to the date of lung cancer recurrence, metastasis, or lung cancer-related death, whichever came first. Patients who died from causes other than lung cancer or who were alive without progression were censored at the last contact date.

Although convolutional neural network (CNN) is a powerful tool for computer vision and image processing [22,23], a large sample size is generally required for a CNN to outperform handcrafted features [24]. For this reason, we chose to use a multilayer perceptron (MLP) neural network with handcrafted features because: (1) our study sample was not sufficiently large for CNN algorithm development; (2) it is difficult to decode the deep learning black box from automatically created CNN image features to identify the driving force that leads to the final predicted score.

We used the following input variables for the IDLE (Figure 2): (1) patient demographics at the time of surgery, (2) surgery type (sublobar resection or lobectomy), (3) residual disease after surgery (R0 or R1), (4) lymphadenectomy received, (5) surgical tissue-associated features (listed in Supplementary Table S1), (6) preoperative LDCT lung image features (listed in Supplementary Table S2) in different anatomic lung locations, and (7) interval

(in days) between the preoperative LDCT lung screening and surgery. We used MLP with two hidden layers and one last layer similar to our prior publication [25]. The first hidden layer activation functions were created from input variables, and the second hidden layer activation functions were created from the first hidden-layer variables with properly selected weights. The cross-entropy loss function with an L_2 penalty parameter was used in feature selection and weight optimization in both hidden layers. The last layer used a random survival forest with input variables and weights coming from the second hidden layer. The final output of the network was from the random survival forest-predicted value normalized to between 0 and 1. The IDLE score was calculated as a predicted risk score through the leave-one-patient-out cross-validation method.

2.5. Statistical Methods

Summary statistics were used to compare demographics and clinical characteristics between the patients with and without progression. Two-sample *t*-test was used for continuous variables, and Fisher's exact test was used for categorical variables. All tests were two-sided.

The prediction accuracy evaluation criteria included 5-year and 10-year area under the time-dependent ROC curve (AUC), time-dependent positive predictive value (PPV), time-dependent negative predictive value (NPV), and the hazard ratio (HR) of progression-free survival between high- and low-risk subgroups. The cutoff value of IDLE that maximizes the hazard ratio was selected to define its high-risk subgroup. The standard deviation of the AUC was computed using inverse probability of censoring weighted estimators in 500 bootstrap simulations. The time-dependent PPV was computed using Bayes' rule:

$$\text{PPV}(t) = \frac{(\text{time} - \text{dependent sensitivity at time } t) \times (1 - S(t))}{\text{Positive test probability}} \tag{6}$$

where $S(t)$ is the Kaplan–Meier estimate of progression-free survival probability at year t. The time-dependent NPV was calculated similarly. The added value of the IDLE scores to TNM staging and tumor grade was further evaluated through the multivariate Cox proportional hazards model adjusting for age at surgery, chemotherapy, and radiotherapy received.

2.6. Decoding the Deep Learning Black Box

To discover how features were used inside the deep learning black box, we took a slightly different approach from the leave-one-patient-out cross-validation method used by IDLE. We re-built IDLE networks using the same IDLE input variables but with all patients included (Figure S1A). We tracked how the input variables were processed within the network. We ranked the input features by the number of times they were selected across hidden layers, so the top ranked feature was defined as the most frequently selected variable. We then dichotomized these variables using the thresholds that maximize their hazard ratios in the univariate logrank test, with time to cancer progression as the endpoint. We repeated the same analysis using time to local recurrence and time to distant metastasis as the endpoint respectively. To find how deep learning enhanced features and how these features were integrated, we combined the input variables with all the hidden-layer features within the network and calculated their values using all 182 patients. Univariate tests were performed to examine how well each of these features was associated with cancer progression status using the *t*-test (for continuous features) and the chi-square test (for discrete features) to compare patients with and without lung cancer progression. Top 25 features with the smallest *p*-values from the *t*-test or the chi-square test were selected to construct a heatmap to visualize how these features worked together to boost prediction accuracy. We calculated *z*-scores for each feature using

$$z = \frac{(\text{feature value}) - (\text{mean feature value across all patients})}{\text{standard deviation}} \tag{7}$$

and plotted these *z*-scores on a heatmap.

To further evaluate synergy between preoperative LDCT image features and tissue H&E image features, we used exactly the same method for constructing an IDLE predictor to derive two additional deep learning predictors: the LDCT feature-derived predictor that did not use H&E features, and the H&E feature-derived predictor that did not use LDCT features. Their prediction accuracies were compared to IDLE through time-dependent ROC analyses.

3. Results

3.1. Study Sample Characteristics

The demographics and clinical characteristics of the selected 182 patients are summarized in Table 1. During 12 years follow up, cancer progression was observed in 54 patients. Age at surgery, smoke pack-years, tumor location, surgery type, surgically removed lesion size, largest invasive tumor size, and TNM staging did not differ between the patients with and without cancer progression (all $p > 0.10$). However, in 39% (21/54) of the patients who progressed, cancers were diagnosed more than 6 months after their last LDCT screening date, compared to 21% (27/128) in the patients who did not progress ($p = 0.0167$). There was no statistical difference in the waiting time from the date of cancer diagnosis to the date of surgery between the patients with and without progression ($p = 0.4042$), but patients who progressed had a longer time interval between the last preoperative LDCT screening date and the surgery date ($p = 0.0468$). Surgery was the initial treatment for all patients, except for three patients who received chemotherapy or radiotherapy two to three months prior to surgery. Eighty-four percent (153/182) of patients received lobectomy, and 97% (177/182) of patients had no residual disease after surgery. Ninety percent (164/182) of patients also received lymphadenectomy with negative results.

Table 1. Summary of the study sample. Cancer progression is defined as a competing event from any of the following events during the 12-year follow-up after the primary tumor surgery: lung cancer recurrence, metastasis, or lung cancer-related death.

		No Progression N = 128	Progression N = 54	p [1]
Cancers diagnosed 6 months after the last LDCT screening date		27	21	0.0167
Lung cancer-related death		0	45	
Age at surgery		64.7 ± 4.9	65.9 ± 4.8	0.1067
Female, N (%)		58 (45%)	24 (44%)	1.0
Smoke pack-years		66 ± 29	72 ± 41	0.2889
Days from the last LDCT screening to the date of lung surgery		177 ± 210	267 ± 299	0.0468
Surgery type	Sublobar resection Lobectomy	21 107	8 46	1.0
Lymphadenectomy	N (%)	115 (90%)	49 (91%)	1.0
Residual disease after surgery	R0 R1	124 4	53 1	1.0
Surgically removed lesion size (mm)		19.7 ± 13.9	20.6 ± 12.6	0.6713
Largest invasive tumor size (mm)		11.4 ± 6.8	12.9 ± 6.4	0.1604
Pathological cancer stage (TNM, 8th edition)	IA1 (T1a) IA1 (T1b) IA1 (T1c)	50 63 15	19 30 5	1.0
Highest tumor grade from all the ROIs	1 = well-differentiated 2 = moderately differentiated 3 = poorly differentiated 4 = undifferentiated Undetermined (GX)	34 53 35 5 1	4 30 14 4 2	0.0163

[1] Two-sample *t*-test was used for continuous variables, and Fisher's exact test was used for categorical variables. All tests were two-sided.

Each patient in our cohort had 2–3 digitally imaged H&E tumor sections, and a total of 477 H&E whole-slide images were available from all 182 patients. Within these H&E-stained slides, 1076 pathologist-annotated regions of interest (ROIs) were used to extract tissue features. Fifty-five patients had two or more histological subtypes reported in their H&E sections, including 53 patients with both adenocarcinoma and AIS; one patient with adenocarcinoma, squamous carcinoma, and AIS; and one patient with squamous carcinoma and AIS. The remaining 127 patients had only one histology subtype diagnosed, including 56 patients with adenocarcinoma, 15 patients with squamous carcinoma, 42 patients with large-cell carcinoma, and 14 patients with AIS. There was no statistical difference in the histological subtype distribution between the patients with and without progression ($p = 0.1508$), but patients who progressed did generally have higher tumor grades ($p = 0.0163$).

The IDLE scores were developed through leave-one-patient-out cross-validation prediction. We found that patients who progressed had higher IDLE scores and a higher tumor grade than patients who did not progress. Figure 3 illustrates preoperative tumor LDCT image texture maps and the corresponding predicted IDLE scores from eight cancers. Patients shown in panels A, B, C, and D survived for over 10 years without lung cancer progression, whereas patients shown in panels E, F, G, H died from lung cancer within 3 years after surgery. All eight of these patients received primary tumor surgeries within one month of their cancer diagnoses. However, lung cancers were diagnosed 330, 233, and 250 days after the last low-dose CT screening dates for patients C, E, and G, respectively. We note that aggressive tumors (from patients E, F, G, and H) had substantially different LDCT image texture maps as compared to nonaggressive tumors (from patients A, B, C, and D) in voxel intensity and skewness patterns.
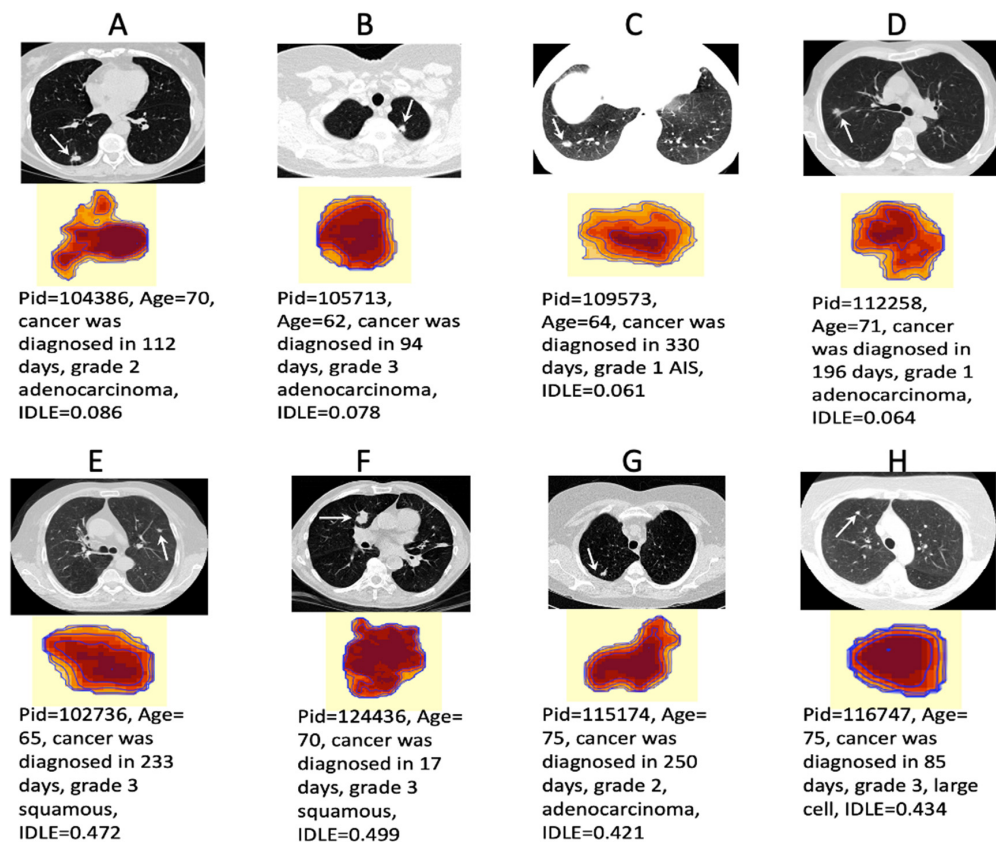

Figure 3. Examples of LDCT tumor texture maps and IDLE scores from four patients with nonaggressive tumors (**A–D**) who survived for over 10 years without cancer progression after the surgery and four patients with aggressive tumors (**E–H**) who died within 3 years after the surgery. Pid = patient ID number in the NLST.

3.2. Added Values of the IDLE Scores to TNM Staging and Tumor Grade

Markedly increased prediction accuracy was observed with the application of IDLE scores. The 5-year time-dependent AUC of IDLE was 0.817 ± 0.037, which was significantly higher than the corresponding AUCs of 0.561 ± 0.042 and 0.573 ± 0.044 from TNM staging and tumor grade (Figure 4A). Similarly, the 10-year time-dependent AUC of IDLE was 0.792 ± 0.039 as compared to the AUCs of 0.507 ± 0.041 and 0.569 ± 0.045 from TNM staging and tumor grade (Figure 4D). The 5-year and 10-year time-dependent NPVs for TNM staging were evaluable only when sensitivities were less than 0.740 and 0.645, respectively, since higher sensitivities were not reached. After fixing the common 5-year and 10-year time-dependent sensitivities in the range of 60–95% for all predictors, IDLE resulted in uniformly higher positive predictive value and negative predictive value than TNM staging and tumor grade (Figure 4B,C,E,F). IDLE also better separated high risk subgroup from the low one (HR = 5.643, $p < 0.0001$) than TNM staging (T1b: HR = 1.319, $p = 0.3454$; T1c: HR = 0.914, $p = 0.8593$; T1a: reference) and tumor grade (HR = 1.200, $p = 0.5319$) (Figure 4G–I).
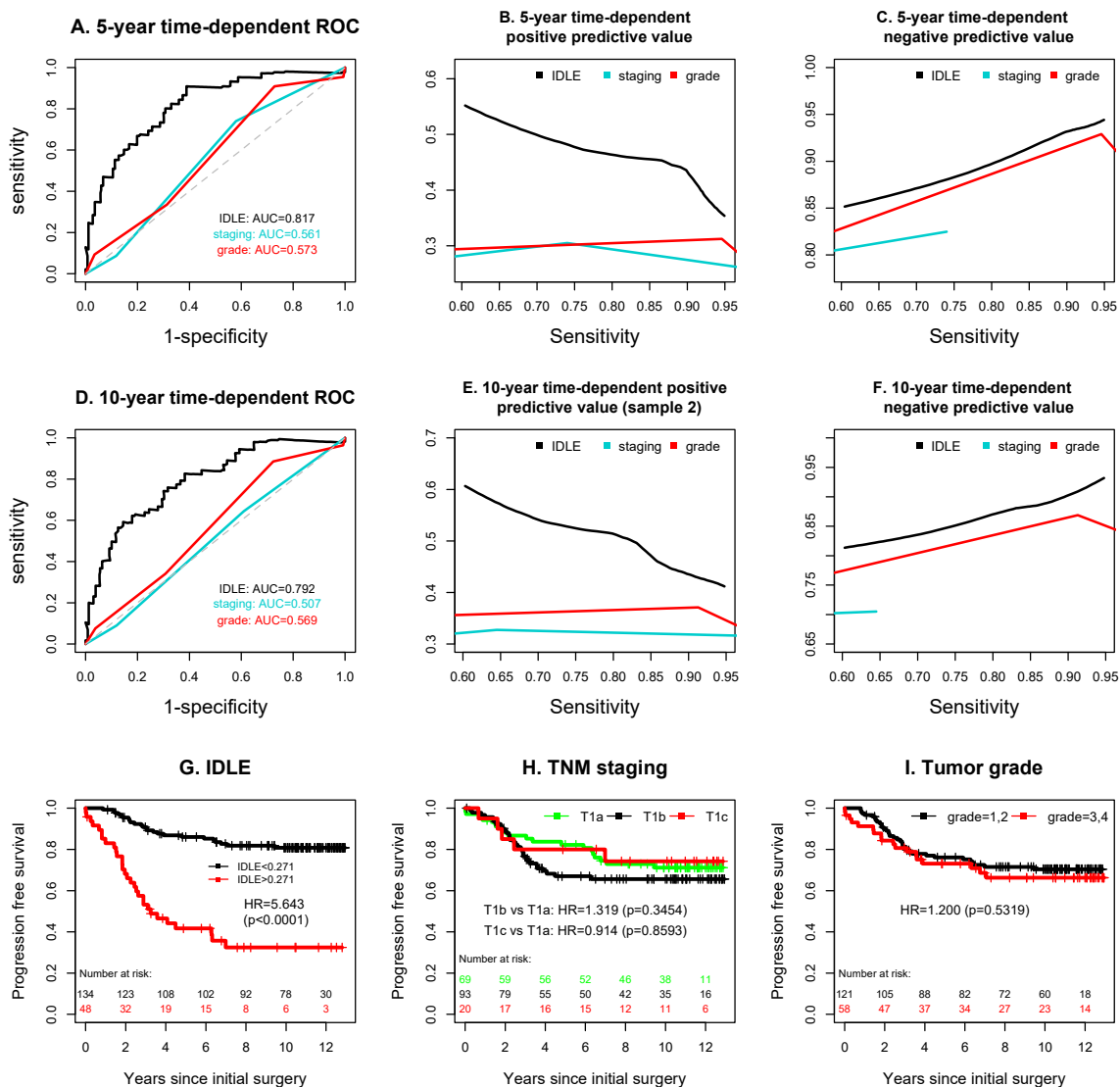
Figure 4. Prediction accuracy comparison of IDLE, TNM staging, and tumor grade. Negative predictive values for TNM staging were evaluable only in a part of sensitivity regions since higher sensitivities were not reached. The patients with undetermined (GX) tumor grade were excluded from the Kaplan–Meier curves in plot I.

In the multivariate proportional hazards regression model that included IDLE, TNM staging, tumor grade, and the potential confounding variables (age at surgery, chemotherapy, and radiotherapy received either before or after the surgery) (Table 2), both TNM staging and tumor grade lost statistical association with cancer progression (*p* > 0.63) when IDLE was included in the model, while the association of IDLE with progression was highly significant (HR = 5.671, 95% CI = 3.1650–10.1605, *p* < 0.0001). Thus, IDLE has a stronger association with cancer progression than TNM staging and tumor grade in this multivariate analysis.

Table 2. Multivariate analysis to study the added value of IDLE.

	HR	95% CI	*p*
IDLE high	5.6708	(3.1650, 10.1605)	<0.0001
T1b [1]	0.8665	(0.4667, 1.6087)	0.6499
T1c [1]	0.7708	(0.2620, 2.2680)	0.6364
High grade	0.9818	(0.5440, 1.7720)	0.9513
Age at surgery	1.0318	(0.9738, 1.0934)	0.2888
Chemotherapy	0.6700	(0.2806, 1.5996)	0.3671
Radiotherapy	1.3959	(0.4064, 4.7945)	0.5963

[1] The reference is the T1a subgroup.

3.3. Synergy inside the Deep Learning Network Black Box

The most frequently selected IDLE input variables by the deep learning hidden layers were (1) tumor grade > 1, (2) tumor $LocSdV_{100}$ feature from the LDCT image, (3) tumor root-mean-square feature from the LDCT image, (4) age at surgery, (5) GLRLM skewness and kurtosis in the tumor's LDCT image, (6) the largest invasive tumor dimension from the H&E image, (7) histology subtype other than AIS, and (8) days between the last LDCT screening date and the surgery date. The logrank test showed that these variables were able to significantly separate patient's cancer progression risks with properly chosen cut points. Variables that were significantly associated with local recurrence or distant metastasis were only from the LDCT image features. Notably, tumor GLRLM skewness, solid or partially solid lesions in the right upper lobe, tumor region $LocSdV_{100}$, and the tumor pixel mean and standard deviation values were all positively associated with local recurrence (*p*-values ranged from 0.0147 to 0.0425). The tumor pixel mean and root-mean-square values were also positively associated with distant cancer progression (*p* = 0.0007 and 0.0393). These observations imply that higher recurrence was observed among patients with solid or partially solid tumors as compared to nonsolid tumors, and their tumor voxel intensity distribution was skewed with one or more spatial clusters of high-intensity voxels.

When ranking the combined deep learning input and hidden-layer variables using *p*-values from the univariate *t*-test or the chi-square test to compare patients with and without cancer progression, all top 25 ranked features were from the hidden-layer variables that are composite functions of features in different platforms, including (1) preoperative LDCT image features, (2) surgical tissue features, (3) demographics, and (4) days between LDCT screening and surgery. Figure S1B shows that these hidden-layer variables, when working together, provided a better separation of patients with and without cancer progression than any single input variable, thus demonstrating strong synergy between the tumor's global features (from preoperative LDCT images) and the localized histologic features (from H&E images).

To further study the synergy, we used exactly the same method of constructing IDLE to calculate LDCT feature-predicted scores (by excluding H&E features from the IDLE) and H&E feature-predicted scores (by excluding LDCT features from the IDLE). The 5- and 10-year time-dependent AUCs were 0.673 ± 0.041 and 0.632 ± 0.043 for the LDCT feature-predicted scores and 0.638 ± 0.050 and 0.648 ± 0.047 for the tissue H&E feature-predicted scores (Figure S2): they were all significantly lower than the AUCs from the IDLE. The time-dependent positive predictive values and negative predicted values (Figure S2) of

these two individual scores were also uniformly lower than the corresponding integrated scores from the IDLE in Figure 4A–F.

4. Discussion

In this study, we showed that integrating diverse prognostic variables—global tumor features from preoperative LDCT images and localized histologic features from tissue H&E images—through a deep learning algorithm can identify aggressive stage IA NSCLCs better than TNM staging and tumor grade. Since IDLE provided uniformly higher ROC curve, positive predictive values, and negative predictive values than TNM staging and tumor grade, it has the potential to better identify patients with a high risk of cancer progression right after the primary surgery, and to select patients for early treatment. In this study, we demonstrated that the IDLE score added value to TNM staging and tumor grade and was significantly associated with cancer progression even after adjusting for the TNM stage and tumor grade (Table 2). These results imply that IDLE contains cancer progression risk information that is not available from the TNM stage and tumor grade.

We identified the driving force of synergy between the global LDCT image features and the local tissue features inside the deep learning black box. These hidden-layer variables, when integrated together, were able to accurately differentiate the patients who had cancer progression from those who did not progress (Figure S1B). Interestingly, none of the individual input variables was among the top 25 ranked variables used by the deep learning network, implying that no individual input feature was sufficiently strong as a single marker. Even when combining multiple image features of LDCT or multiple morphological features of histopathology, the prediction accuracies of these separate approaches were substantially lower than the prediction accuracy of IDLE, which integrated both LDCT image features and histology image features (Figure S2 and Figure 4).

Why does the integration of LDCT features with histology features provide a better predictive accuracy? While standardized tissue histopathological features provide insight into tumor-localized properties, histology of the tumor alone has limited ability to quantify how the tumor interacts with the global environment in the lung. Since LDCT image features quantify the tumor's global morphology while H&E image features quantify the tumor's local morphology, tumor characteristics from these two feature platforms are generally nonoverlapping. Our study suggests that integrating features from these nonoverlapping platforms produces a much stronger predictor than adding multiple features from a singular platform, which likely has overlapping information. Recently, a number of molecular biomarkers have been reported to have prognostic value in early lung cancer, and some of these markers also provide targets for treatment [26–31]. It is likely that integrating molecular biomarkers with IDLE may further improve prediction accuracy. To facilitate such an investigation, we make our IDLE data publicly available at GitHub with the participants' NLST identification numbers included.

Interestingly, the number of days between the last LDCT image scanning date and the surgery date was negatively associated with progression-free survival, and this variable was one of the top input features used by the deep learning network. A longer interval between the LDCT screening date and the surgery date was primarily due to the delay in cancer diagnosis. Among the 54 patients who progressed, 21 patients had delayed cancer diagnosis (Table 1), and even though they were still in stage IA at the time of surgery, they had a higher tumor grade than the other patients. Timely diagnosis of aggressive tumors is evidently important to improve the lung cancer screening efficacy. Our recent work in DeepLR [25] (available at www.caced.jhu.edu) provides estimates of lung cancer incidence risk, tumor aggressiveness, and suggests screening intervals that could help earlier diagnosis of aggressive tumors.

TNM staging and tumor grade have long been recognized to be robust prognostic markers of cancer progression for postoperative NSCLC patients. Attempts to improve this clinical gold standard through computer-aided prediction models are sparse. D'Antonoli et al. showed that using a combination of tumoral and peritumoral radiomic features in

preoperative CT with TNM staging outperformed TNM staging alone in patients with stage IA–IIB [32]. Wang et al. showed that the top nuclear morphometric features from H&E tissue microarray images were associated with cancer recurrence in patients with resected early-stage NSCLCs [33]. Other authors have used either CT features or histopathological features to identify patients with early-stage NSCLC who are at risk of local recurrence or distant metastasis [16,34,35]. Our study represents the first attempt that combines both LDCT and histopathology features to improve the prediction of cancer recurrence or progression.

A limitation that we recognize in our study is that the sample size was relatively small, and the results were only cross-validated. Further prospective validation in a larger sample is needed. Recognizing this limitation and the need for further refinement of the IDLE score, our study does support our original hypothesis that nonoverlapping tumor morphological features from different platforms can improve lung cancer progression risk stratification over TNM staging and tumor grade.

5. Conclusions

Low-dose computed tomography (LDCT) screening for high-risk individuals increases the detection of stage IA NSCLCs [2,3]. The current standard of care is surgery alone. However, stage I NSCLCs treated in this manner have diverse survival outcomes [36–40]. Identifying the subset of patients with potentially curable lung cancer who are at a high risk of tumor recurrence or progression who would benefit from adjuvant therapy is critical for improving the outcome of these patients. Our study supports the use of diverse nonoverlapping tumor measures in lung cancer risk evaluation. The findings of our study have potential to not only help treating oncologists improve clinical management of early-stage NSCLCs, but also to help investigators to better define patient selection criteria and determine what measurements to collect in future clinical trial designs.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Field, J.K.; Duffy, S.W.; Baldwin, D.R.; Brain, K.E.; Devaraj, A.; Eisen, T.; Green, B.A.; Holemans, J.A.; Kavanagh, T.; Kerr, K.M.; et al. The UK Lung Cancer Screening Trial: A pilot randomised controlled trial of low-dose computed tomography screening for the early detection of lung cancer. *Health Technol. Assess.* **2016**, *20*, 1–146. [CrossRef] [PubMed]
2. Tammemagi, M.C.; Schmidt, H.; Martel, S.; McWilliams, A.; Goffin, J.R.; Johnston, M.R.; Nicholas, G.; Tremblay, A.; Bhatia, R.; Liu, G.; et al. PanCan Study T. Participant selection for lung cancer screening by risk modelling (the Pan-Canadian Early Detection of Lung Cancer [PanCan] study): A single-arm, prospective study. *Lancet Oncol.* **2017**, *18*, 1523–1531. [CrossRef]
3. De Koning, H.J.; van der Aalst, C.M.; de Jong, P.A.; Scholten, E.T.; Nackaerts, K.; Heuvelmans, M.A.; Lammers, J.J.; Weenink, C.; Yousaf-Khan, U.; Horeweg, N.; et al. Reduced Lung-Cancer Mortality with Volume CT Screening in a Randomized Trial. *N. Engl. J. Med.* **2020**, *382*, 503–513. [CrossRef]
4. Rami-Porta, R.; Crowley, J.J.; Goldstraw, P. The revised TNM staging system for lung cancer. *Ann. Thorac. Cardiovasc. Surg.* **2009**, *15*, 4–9.
5. Wang, C.; Wu, Y.; Shao, J.; Liu, D.; Li, W. Clinicopathological variables influencing overall survival, recurrence and post-recurrence survival in resected stage I non-small-cell lung cancer. *BMC Cancer* **2020**, *20*, 150. [CrossRef]
6. Ujiie, H.; Kadota, K.; Chaft, J.E.; Buitrago, D.; Sima, C.S.; Lee, M.C.; Huang, J.; Travis, W.D.; Rizk, N.P.; Rudin, C.M.; et al. Solid Predominant Histologic Subtype in Resected Stage I Lung Adenocarcinoma Is an Independent Predictor of Early, Extrathoracic, Multisite Recurrence and of Poor Postrecurrence Survival. *J. Clin. Oncol.* **2015**, *33*, 2877–2884. [CrossRef] [PubMed]
7. Su, S.; Scott, W.J.; Allen, M.S.; Darling, G.E.; Decker, P.A.; McKenna, R.J.; Meyers, B.F. Patterns of survival and recurrence after surgical treatment of early stage non-small cell lung carcinoma in the ACOSOG Z0030 (ALLIANCE) trial. *J. Thorac. Cardiovasc. Surg.* **2014**, *147*, 747–752. [CrossRef] [PubMed]
8. Shimada, Y.; Saji, H.; Yoshida, K.; Kakihana, M.; Honda, H.; Nomura, M.; Usuda, J.; Kajiwara, N.; Ohira, T.; Ikeda, N. Prognostic factors and the significance of treatment after recurrence in completely resected stage I non-small cell lung cancer. *Chest* **2013**, *143*, 1626–1634. [CrossRef]
9. Hung, J.J.; Jeng, W.J.; Hsu, W.H.; Wu, K.J.; Chou, T.Y.; Hsieh, C.C.; Huang, M.H.; Liu, J.S.; Wu, Y.C. Prognostic factors of postrecurrence survival in completely resected stage I non-small cell lung cancer with distant metastasis. *Thorax* **2010**, *65*, 241–245. [CrossRef]
10. Hung, J.J.; Hsu, W.H.; Hsieh, C.C.; Huang, B.S.; Huang, M.H.; Liu, J.S.; Wu, Y.C. Post-recurrence survival in completely resected stage I non-small cell lung cancer with local recurrence. *Thorax* **2009**, *64*, 192–196. [CrossRef]
11. Nakagawa, T.; Okumura, N.; Ohata, K.; Igai, H.; Matsuoka, T.; Kameyama, K. Postrecurrence survival in patients with stage I non-small cell lung cancer. *Eur. J. Cardiothorac. Surg.* **2008**, *34*, 499–504. [CrossRef] [PubMed]
12. Forde, P.M.; Chaft, J.E.; Smith, K.N.; Anagnostou, V.; Cottrell, T.R.; Hellmann, M.D.; Zahurak, M.; Yang, S.C.; Jones, D.R.; Broderick, S.; et al. Neoadjuvant PD-1 Blockade in Resectable Lung Cancer. *N. Engl. J. Med.* **2018**, *378*, 1976–1986. [CrossRef] [PubMed]
13. Felip, E.; Altorki, N.; Zhou, C.; Csoszi, T.; Vynnychenko, I.; Goloborodko, O.; Luft, A.; Akopov, A.; Martinez-Marti, A.; Kenmotsu, H.; et al. Adjuvant atezolizumab after adjuvant chemotherapy in resected stage IB-IIIA non-small-cell lung cancer (IMpower010): A randomised, multicentre, open-label, phase 3 trial. *Lancet* **2021**, *398*, 1344–1357. [CrossRef]
14. Wu, Y.L.; Tsuboi, M.; He, J.; John, T.; Grohe, C.; Majem, M.; Goldman, J.W.; Laktionov, K.; Kim, S.W.; Kato, T.; et al. Osimertinib in Resected EGFR-Mutated Non-Small-Cell Lung Cancer. *N. Engl. J. Med.* **2020**, *383*, 1711–1723. [CrossRef]
15. Qiu, B.; Guo, W.; Zhang, F.; Lv, F.; Ji, Y.; Peng, Y.; Chen, X.; Bao, H.; Xu, Y.; Shao, Y.; et al. Dynamic recurrence risk and adjuvant chemotherapy benefit prediction by ctDNA in resected NSCLC. *Nat. Commun.* **2021**, *12*, 6770. [CrossRef] [PubMed]
16. Huang, Y.; Liu, Z.; He, L.; Chen, X.; Pan, D.; Ma, Z.; Liang, C.; Tian, J.; Liang, C. Radiomics Signature: A Potential Biomarker for the Prediction of Disease-Free Survival in Early-Stage (I or II) Non-Small Cell Lung Cancer. *Radiology* **2016**, *281*, 947–957. [CrossRef] [PubMed]
17. National Lung Screening Trial Research Team; Aberle, D.R.; Adams, A.M.; Berg, C.D.; Black, W.C.; Clapp, J.D.; Fagerstrom, R.M.; Gareen, I.F.; Gatsonis, C.; Marcus, P.M.; et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N. Engl. J. Med.* **2011**, *365*, 395–409. [CrossRef] [PubMed]
18. Detterbeck, F.C.; Boffa, D.J.; Kim, A.W.; Tanoue, L.T. The Eighth Edition Lung Cancer Stage Classification. *Chest* **2017**, *151*, 193–203. [CrossRef]
19. Patz, E.F., Jr.; Caporaso, N.E.; Dubinett, S.M.; Massion, P.P.; Hirsch, F.R.; Minna, J.D.; Gatsonis, C.; Duan, F.; Adams, A.; Apgar, C.; et al. National Lung Cancer Screening Trial American College of Radiology Imaging Network Specimen Biorepository originating from the Contemporary Screening for the Detection of Lung Cancer Trial (NLST, ACRIN 6654): Design, intent, and availability of specimens for validation of lung cancer biomarkers. *J. Thorac. Oncol.* **2010**, *5*, 1502–1506. [CrossRef]
20. Huang, P.; Park, S.; Yan, R.; Lee, J.; Chu, L.C.; Lin, C.T.; Hussien, A.; Rathmell, J.; Thomas, B.; Chen, C.; et al. Added Value of Computer-aided CT Image Features for Early Lung Cancer Diagnosis with Small Pulmonary Nodules: A Matched Case-Control Study. *Radiology* **2017**, *286*, 286–295. [CrossRef]

21. Aerts, H.J.; Velazquez, E.R.; Leijenaar, R.T.; Parmar, C.; Grossmann, P.; Carvalho, S.; Bussink, J.; Monshouwer, R.; Haibe-Kains, B.; Rietveld, D.; et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat. Commun.* **2014**, *5*, 4006. [CrossRef] [PubMed]

22. Anwar, S.M.; Majid, M.; Qayyum, A.; Awais, M.; Alnowami, M.; Khan, M.K. Medical Image Analysis using Convolutional Neural Networks: A Review. *J. Med. Syst.* **2018**, *42*, 226. [CrossRef] [PubMed]

23. Khan, A.; Sohail, A.; Zahoora, U.; Qureshi, A.S. A survey of the recent architectures of deep convolutional neural networks. *Artif. Intell. Rev.* **2020**, *53*, 5455–5516. [CrossRef]

24. Lin, W.; Hasenstab, K.; Moura Cunha, G.; Schwartzman, A. Comparison of handcrafted features and convolutional neural networks for liver MR image adequacy assessment. *Sci. Rep.* **2020**, *10*, 20336. [CrossRef] [PubMed]

25. Huang, P.; Lin, C.T.; Li, Y.; Tammemagi, C.M.; Brock, M.; Atkar-Khattra, S.; Xu, Y.; Hu, P.; Mayo, J.; Schmidt, H.; et al. Prediction of lung cancer risk at follow-up screening with low-dose CT: A training and validation study of a deep learning method. *Lancet Digit. Health* **2019**, *1*, e353–e362. [CrossRef]

26. Passiglia, F.; Bronte, G.; Castiglia, M.; Listi, A.; Calo, V.; Toia, F.; Cicero, G.; Fanale, D.; Rizzo, S.; Bazan, V.; et al. Prognostic and predictive biomarkers for targeted therapy in NSCLC: For whom the bell tolls? *Expert Opin. Biol. Ther.* **2015**, *15*, 1553–1566. [CrossRef]

27. Dejima, H.; Iinuma, H.; Kanaoka, R.; Matsutani, N.; Kawamura, M. Exosomal microRNA in plasma as a non-invasive biomarker for the recurrence of non-small cell lung cancer. *Oncol. Lett.* **2017**, *13*, 1256–1263. [CrossRef]

28. Tang, Y.; Qiao, G.; Xu, E.; Xuan, Y.; Liao, M.; Yin, G. Biomarkers for early diagnosis, prognosis, prediction, and recurrence monitoring of non-small cell lung cancer. *Onco Targets Ther.* **2017**, *10*, 4527–4534. [CrossRef]

29. Kanaoka, R.; Iinuma, H.; Dejima, H.; Sakai, T.; Uehara, H.; Matsutani, N.; Kawamura, M. Usefulness of Plasma Exosomal MicroRNA-451a as a Noninvasive Biomarker for Early Prediction of Recurrence and Prognosis of Non-Small Cell Lung Cancer. *Oncology* **2018**, *94*, 311–323. [CrossRef]

30. Camidge, D.R.; Doebele, R.C.; Kerr, K.M. Comparing and contrasting predictive biomarkers for immunotherapy and targeted therapy of NSCLC. *Nat. Rev. Clin. Oncol.* **2019**, *16*, 341–355. [CrossRef]

31. Bodor, J.N.; Boumber, Y.; Borghaei, H. Biomarkers for immune checkpoint inhibition in non-small cell lung cancer (NSCLC). *Cancer* **2020**, *126*, 260–270. [CrossRef] [PubMed]

32. Akinci D'Antonoli, T.; Farchione, A.; Lenkowicz, J.; Chiappetta, M.; Cicchetti, G.; Martino, A.; Ottavianelli, A.; Manfredi, R.; Margaritora, S.; Bonomo, L.; et al. CT Radiomics Signature of Tumor and Peritumoral Lung Parenchyma to Predict Nonsmall Cell Lung Cancer Postsurgical Recurrence Risk. *Acad. Radiol.* **2020**, *27*, 497–507. [CrossRef] [PubMed]

33. Wang, X.; Janowczyk, A.; Zhou, Y.; Thawani, R.; Fu, P.; Schalper, K.; Velcheti, V.; Madabhushi, A. Prediction of recurrence in early stage non-small cell lung cancer using computer extracted nuclear features from digital H&E images. *Sci. Rep.* **2017**, *7*, 13543. [CrossRef] [PubMed]

34. Corredor, G.; Wang, X.; Zhou, Y.; Lu, C.; Fu, P.; Syrigos, K.; Rimm, D.L.; Yang, M.; Romero, E.; Schalper, K.A.; et al. Spatial Architecture and Arrangement of Tumor-Infiltrating Lymphocytes for Predicting Likelihood of Recurrence in Early-Stage Non-Small Cell Lung Cancer. *Clin. Cancer Res.* **2019**, *25*, 1526–1534. [CrossRef] [PubMed]

35. Sun, F.; Ma, K.; Yang, X.; Li, M.; Shi, Y.; Zhan, C.; Jiang, W.; Wang, Q. A nomogram to predict prognosis after surgery in early stage non-small cell lung cancer in elderly patients. *Int. J. Surg.* **2017**, *42*, 11–16. [CrossRef] [PubMed]

36. Khullar, O.V.; Liu, Y.; Gillespie, T.; Higgins, K.A.; Ramalingam, S.; Lipscomb, J.; Fernandez, F.G. Survival After Sublobar Resection versus Lobectomy for Clinical Stage IA Lung Cancer: An Analysis from the National Cancer Data Base. *J. Thorac. Oncol.* **2015**, *10*, 1625–1633. [CrossRef] [PubMed]

37. Yang, C.J.; Kumar, A.; Klapper, J.A.; Hartwig, M.G.; Tong, B.C.; Harpole, D.H., Jr.; Berry, M.F.; D'Amico, T.A. A National Analysis of Long-term Survival Following Thoracoscopic Versus Open Lobectomy for Stage I Non-small-cell Lung Cancer. *Ann. Surg.* **2019**, *269*, 163–171. [CrossRef]

38. Ijsseldijk, M.A.; Shoni, M.; Siegert, C.; Seegers, J.; van Engelenburg, A.K.C.; Tsai, T.C.; Lebenthal, A.; Ten Broek, R.P.G. Oncological Outcomes of Lobar Resection, Segmentectomy, and Wedge Resection for T1a Non-Small-Cell Lung Carcinoma: A Systematic Review and Meta-Analysis. *Semin. Thorac. Cardiovasc. Surg.* **2020**, *32*, 582–590. [CrossRef]

39. Winckelmans, T.; Decaluwe, H.; De Leyn, P.; Van Raemdonck, D. Segmentectomy or lobectomy for early-stage non-small-cell lung cancer: A systematic review and meta-analysis. *Eur. J. Cardiothorac. Surg.* **2020**, *57*, 1051–1060. [CrossRef]

40. Chan, M.V.; Huo, Y.R.; Cao, C.; Ridley, L. Survival outcomes for surgical resection versus CT-guided percutaneous ablation for stage I non-small cell lung cancer (NSCLC): A systematic review and meta-analysis. *Eur. Radiol.* **2021**, *31*, 5421–5433. [CrossRef]