



OPEN

A new model using deep learning to predict recurrence after surgical resection of lung adenocarcinoma

Pil-Jong Kim¹, Hee Sang Hwang², Gyuheon Choi², Hyun-Jung Sung², Bokyoung Ahn², Ji-Su Uh², Shinkyoo Yoon³, Deokhoon Kim², Sung-Min Chun², Se Jin Jang² & Heounjeong Go²✉

This study aimed to develop a deep learning (DL) model for predicting the recurrence risk of lung adenocarcinoma (LUAD) based on its histopathological features. Clinicopathological data and whole slide images from 164 LUAD cases were collected and used to train DL models with an ImageNet pre-trained efficientnet-b2 architecture, densenet201, and resnet152. The models were trained to classify each image patch into high-risk or low-risk groups, and the case-level result was determined by multiple instance learning with final FC layer's features from a model from all patches. Analysis of the clinicopathological and genetic characteristics of the model-based risk group was performed. For predicting recurrence, the model had an area under the curve score of 0.763 with 0.750, 0.633 and 0.680 of sensitivity, specificity, and accuracy in the test set, respectively. High-risk cases for recurrence predicted by the model (HR group) were significantly associated with shorter recurrence-free survival and a higher stage (both, $p < 0.001$). The HR group was associated with specific histopathological features such as poorly differentiated components, complex glandular pattern components, tumor spread through air spaces, and a higher grade. In the HR group, pleural invasion, necrosis, and lymphatic invasion were more frequent, and the size of the invasion was larger (all, $p < 0.001$). Several genetic mutations, including *TP53* ($p = 0.007$) mutations, were more frequently found in the HR group. The results of stages I-II were similar to those of the general cohort. DL-based model can predict the recurrence risk of LUAD and identify the presence of the *TP53* gene mutation by analyzing histopathologic features.

Keywords Deep learning, Lung adenocarcinoma, Recurrence, Histopathology, Pathology image

Lung cancer is the leading cause of cancer morbidity and mortality worldwide, and the incidence of lung adenocarcinoma (LUAD) is still increasing^{1,2}. Currently, locoregional treatment such as surgical resection or radiation therapy is recommended as standard treatment in stages I–II LUAD, except for some cases of stage IIB showing invasive growth³. However, postoperative recurrence is frequent even after complete resection of lung cancer, and the prognosis is generally poor even with salvage treatment⁴. Therefore, predicting the risk of recurrence of lung cancer patients would be very useful when selecting the adjuvant treatment plan.

One of the key factors correlated with recurrence is tumor histology. Of note, a new international association for the study of lung cancer (IASLC) grading system for invasive LUAD has been validated with improved recurrence-free and overall survival discrimination. Tumor spread through air spaces (STAS), a novel invasive pattern of non-small cell lung cancer (NSCLC), has been demonstrated in many studies to be strongly correlated with recurrence after resection, especially in stage I cancers^{5,6} but the concept has been criticized because of the difficulty to discriminate the artifacts associated with specimen handling⁷. In addition, various histopathologic features, such as pathologic TNM stage, tumor size, solid and micropapillary patterns, resection margin status, invasion of blood vessels and/or pleura, and tumor microenvironment have a significant correlation with patient prognosis⁸. However, a detailed histopathologic examination of lung cancer is very difficult and laborious, making it vulnerable to error. According to the results of a previous study, the reproducibility of the current IASLC grading system is good, but not very high, even for expert pathologists⁹.

¹School of Dentistry and Dental Research Institute, Seoul National University, Seoul, Republic of Korea. ²Department of Pathology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea. ³Department of Oncology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea. ✉email: damul37@amc.seoul.kr

Recent advances in digital pathology could help solve this problem. Developments in machine learning (ML)-based image analysis techniques, especially in deep learning (DL), have shown that they can assist with diagnoses, identify novel features, and predict patients' outcomes¹⁰. Research into ML-based histological analysis of lung cancer has mainly dealt with segmentation of tumor boundaries and the classification of tumor types^{11–13}. Several studies tried to predict patient outcomes by automatic histological analyses of histomorphometric features^{14,15} and tumor microenvironment features¹⁶. Recent studies showed that DL-based analysis of images, not histomorphometric features, could predict the recurrence of LUAD^{17,18}. They had meaningful predictive performance, but they lacked analyses about the relationship between the models' output and other histopathologic parameters¹⁹.

In this study, we aimed to develop a new DL-based model to predict the recurrence of LUAD, and then we investigated the results in the context of histopathological parameters and tumoral genetic aberrations.

Materials and methods

Clinicopathological data acquisition

Clinical, pathological, and genomic data were retrieved from a previously reported cohort²⁰. It consists of 164 cases of lung adenocarcinoma that were surgically resected from January 2015 to December 2015. Their data were retrospectively retrieved at Asan Medical Center (AMC), Seoul, Republic of Korea^{20,21}. The pathological data were reviewed by pulmonary pathologists (HSH and BA). Patients' pathological diagnoses were established in line with the World Health Organization (WHO) criteria⁸, IASLC guideline⁹ and the 8th edition of the American Joint Committee on Cancer (AJCC) Cancer Staging Manual²². Tumor samples were subjected to targeted next-generation sequencing (NGS) using the AMC OncoPanel version 4, a custom cancer panel encompassing the entire exome area or mutation hotspot regions of 334 cancer-related genes and intron area of fusion hotspots of the *ALK*, *EGFR*, *NTRK1*, *RET*, *ROS1*, and *BRAF* genes²⁰. The inclusion and exclusion criteria for patients are summarized in Fig. 1.

Image data preparation for training the deep learning model

One representative hematoxylin & eosin (H&E)-stained slide was selected from each case by manual review blinded to clinical and pathological information. The slides were scanned with a 3D Hitech Panoramic 250 Flash II (Budapest, Hungary) scanner at 20× magnification and a resolution of 0.221 μm per pixel. Whole slide images (WSIs) were exported in mrxs format. Four expert pathologists (GC, HJS, JSU, and HG) annotated the boundaries of the tumor site using QuPath 0.3.0 (<https://qupath.github.io>). It was reconfirmed in all images that the annotation results correctly indicated the tumor site.

For developing the DL model, image patches (256 × 256 pixels) were randomly extracted from the annotated tumor area with an average of 100 patches per non-recurrent case and 148 patches per recurrent case to balance the data size between the recurrent and non-recurrent groups. In total, 19,188 patches were retrieved. They

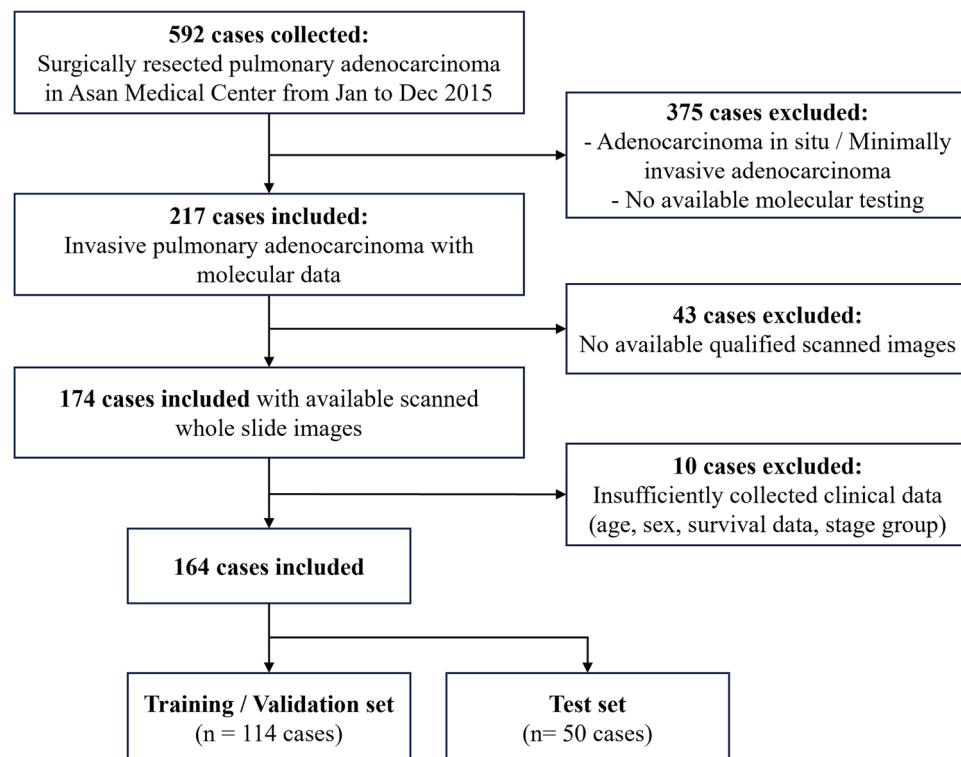


Figure 1. Flowchart of inclusion and exclusion of the patient cohort.

were randomly divided into independent training and test sets at a ratio of 7:3. Cross validation method with fivefold was involved in train procedure. The images were normalized with the training set data in each channel.

Training method of the deep learning model

Due to the complexity and small size of this study's data set, a lightweight network with fewer parameters was suitable because it requires less training time and achieves a performance comparable to other networks. To decide suitable DL model, we compared efficientnet-b2, densenet201 and resnet152. After comparing these DL modeling's accuracy metrics in cross validation, we chose efficientnet-b2 architecture as our classifier, considering its special design for improving accuracy and efficiency through AutoML and model scaling with a verified ability to accomplish classification tasks with high accuracy while using a relatively small number of parameters (~7 million)²³. The model network used ImageNet based pre-trained initialization of weights and was trained with cross-entropy as the loss function. The model parameters were updated by Adam optimizer with 0.9 β_1 and 0.999 β_2 ²⁴. The network was trained with a batch size of 256 and an initial learning rate of 1e-6. The model parameters were iteratively updated to decrease the cross entropy. The model was saved when the least loss of cross-entropy was obtained in the validation set and then it was used for further evaluation and manipulation.

The input data were individual tumor image patches. Ground truth was the status of tumor recurrence of the case from which the image patch was extracted. During model training, data augmentation was applied to improve its robustness: flipping, translation, rotation, and color augmentations, including random contrast (multiplication by 0.5–1.5), brightness (multiplication by 0.65–1.35), hue (addition by -32 to 32) and value (addition by -32 to 32). The DL network was developed with the PyTorch framework (version 1.11.0) on a dual NVIDIA GeForce RTX 3090 under the Python (version 3.8) environment.

Performance evaluation of the model

The model classified each image patch into low-risk (LR) or high-risk (HR) groups according to the output (the model-based feature). Image patches classified as the HR group and extracted from a case with actual tumor recurrence were considered true positive, and vice versa. A case-level output was determined by multiple instance learning with 2-layer preNN and 1-layer afterNN with final FC layer's 1408 features from a model from all patches²⁵.

The average value of the model-based features of the extracted patches. A confusion matrix was used to illustrate the performance of the trained model on the training, validation, and testing set with 4 categorical results [true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN)]. Besides, additional parameters, including sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and F1-score were calculated to obtain a comprehensive performance measure of the results. The 95% confidence intervals (CIs) of sensitivity, specificity, PPV and NPV were calculated to estimate the corresponding variability²⁶. To validate its clinical performance, recurrence-free survival (RFS) rates by risk group were compared using the Kaplan–Meier method and the log-rank test.

Clinicopathological analysis and statistical methods

We analyzed the associations between the model-based results at the case-level and the pathological parameters. Whole data (164 cases) were used in this analysis because this was not for validating the model's performance but for acquiring insights into the model's interpretation. The proportion of poorly differentiated (PD) components and complex glandular pattern (CGP) were evaluated by eyeballing by expert pulmonary pathologists. PD components include solid, micropapillary, cribriform, and CGP. CGP include fused glands with irregular borders and single cells infiltrating desmoplastic patterns⁹. Differences between continuous variables in two groups were evaluated by Student's t-test. Differences in frequencies of categorical variables were estimated by a chi-square test with correction. All statistical evaluations were performed with R version 4.2.1 (The R Foundation for Statistical Computing, Vienna, Austria). *p* value < 0.05 was considered statistically significant.

Ethical approval

This study was conducted according to the ethical guidelines of the Declaration of Helsinki. All studies involving patients were examined and approved by the Institutional Review Board of Asan Medical Center (IRB approval number: 2018-1198). The requirement for written informed consent was waived by IRB of Asan Medical Center because of the retrospective nature of the study and use of deidentified data.

Results

Risk prediction performance of the model

Efficientnet-b2, densenet201, and resnet152 were compared based on cross-validation accuracy at the patch level, and as a result, efficientnet-b2 was chosen as the final learning architecture (Supplementary Table S1). The model performance at the patch-level and case-level were summarized in Table 1. At the patch-level, the model achieved a sensitivity of 70.7% and a specificity of 46.0%. The F1 score was 0.6332 and the accuracy was 58.5%. The area under the curve (AUC) of the receiver operating curves (ROC) in the training, and test sets were 0.622 and 0.604, respectively (Fig. 2A,B). At the selected threshold, 26 of the 50 cases were classified as the HR group in test set. The sensitivity was 75.0% and the specificity was 63.3%. The F1 score was 0.6522 and the accuracy was 68%. The AUC in the training, test sets were 0.796, 0.763, respectively (Fig. 2C,D).

The predicted HR groups were significantly associated with shorter RFS, even when the data were confined to stage I-II cases (Fig. 3A–C). The mean (\pm standard deviation [SD]) RFS was significantly shorter in the HR group (*p* < 0.001): HR group, 855.71 days (\pm 547.83), LR group, 1178.57 days (\pm 521.26). The mean overall survival (OS) was also shorter in the HR group, but the difference was not significant (*p* = 0.143).

	AUC	F1	PPV	NPV	Sensitivity	Specificity	Accuracy
Patch-level performance							
Training set (n = 13,368)	0.622	0.6332	0.5732	0.6051	0.7072	0.4602	0.5852
Test set (n = 5820)	0.604	0.6300	0.5422	0.6335	0.7518	0.4033	0.5722
Case-level performance							
Training set (n = 114)	0.796	0.6526	0.6596	0.7463	0.6458	0.7576	0.7105
Test set (n = 50)	0.763	0.6522	0.5769	0.7917	0.7500	0.6333	0.6800

Table 1. Classification performance of the model. *PPV* positive predictive value, *NPV* negative predictive value.

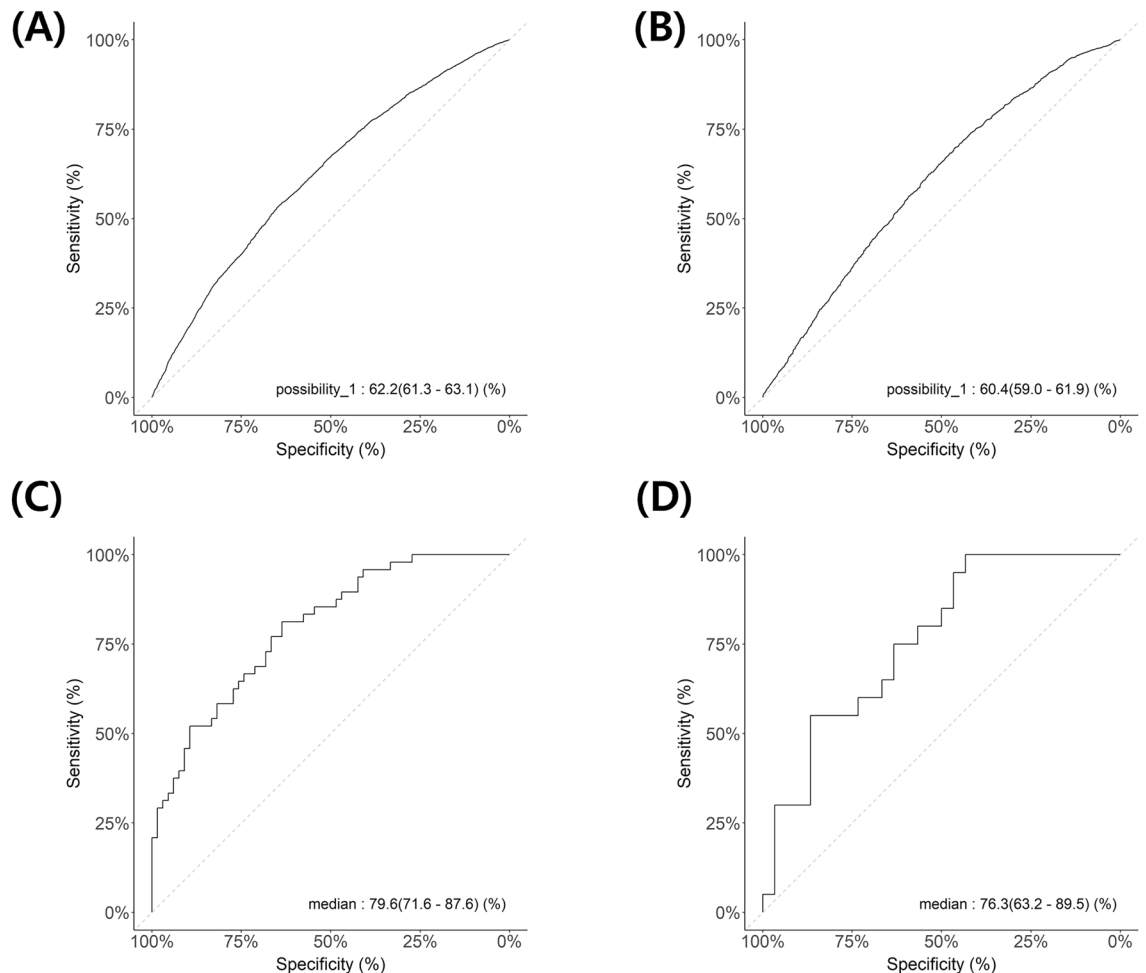


Figure 2. Receiver operating curves of the model at the patch-level (A–C, A: training set, B: validation set, C: testing set) and at the case-level (D–F, D: training set, E: validation set, F: testing set).

Histopathologic features according to risk group and recurrence

Histopathologic comparisons between the HR and LR group are summarized in Table 2. The tumor invasion size was larger in the HR group ($p < 0.001$). The proportion of the predominant histologic type was different between the groups ($p < 0.001$). Cases in which lepidic, acinar and papillary types were predominant, considered well to moderately differentiated histologic subtypes²⁷, were more likely to be assigned to the LR group. In contrast, solid, micropapillary, mucinous and cribriform-predominant cases were only observed in the HR group. IASLC grades of the tumors were higher in the HR group ($p < 0.001$). The HR group had a higher proportion of PD and CGP components ($p < 0.001$, both). Necrosis, STAS, pleural invasion and lymphovascular invasion (LVI) were more common in the HR group ($p < 0.001$ for all comparisons, except STAS's $p = 0.003$). pT, pN and stage group tended to be higher in the HR group ($p < 0.001$ for all comparisons).

Class activation maps (CAMs) shown in Fig. 4 display representative image patches with the highest risk (Fig. 4A) and the lowest risk (Fig. 4B). Representative LR patches were composed of relatively monotonous cells

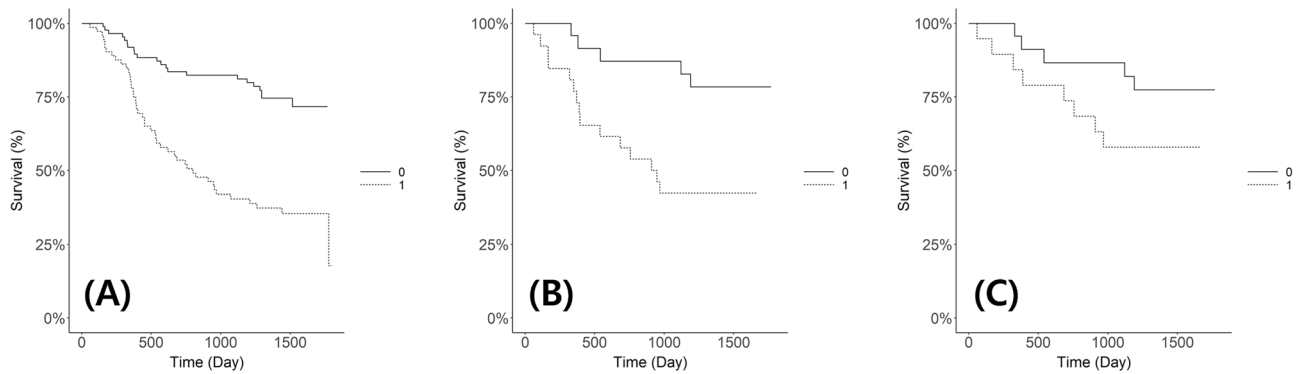


Figure 3. Kaplan–Meier estimation of recurrence-free survival. (A) General group, (B) testing set, (C) stages I–II in testing set.

with lepidic or papillary growth patterns, while the HR patches had tumor cells with pleomorphic nuclei and complex structures. At the case level, The WSIs classified under the HR group often exhibit pronounced cellular pleomorphism, solid structures, and overall poor histological differentiation (Fig. 5A,C). On the other hand, WSIs classified under the LR group predominantly include well-differentiated histologic features with minimal tumor cell pleomorphism, displaying lepidic patterns as shown in Fig. 5B,D.

Additionally, we compared histologic features between patients grouped by their status of actual tumor recurrence. These results are summarized in Supplementary Table S2. The mean tumor invasion size and proportions of PD and CGP components were significantly higher in the recurrence group. IASLC grade, necrosis, STAS, pleural invasion, LVI, pT, pN and stage group were significantly higher in the recurrence group. On the other hand, a predominant histologic type was not significantly associated with recurrence ($p = 0.923$), validating the performance of the IASLC grade.

Association with genomic alterations

NGS data from 163 cases were retrieved and the results are summarized in Table 3. Mutations in four genes were found in a significant number of patients: *CDKN2A*, *TP53*, *KRAS* and *EGFR*. The HR group was significantly associated with *TP53* alterations ($p = 0.007$) and in line with the model prediction, *TP53* alteration was significantly associated with cases of actual recurrence ($p < 0.001$). *ALK* translocation was found in 2 cases, all of which were assigned to the HR group.

Clinical and histopathological characteristics of stage I–II cases

Stage I–II cases were analyzed with more attention because this model could have a significant beneficial impact on these patients by guiding the selection of their adjuvant treatment. Stage I–II patients comprised 125 of the 164 cases (76.2%). Clinical and histopathological comparisons of the Stage I–II patients, when grouped by the model-based risk group and by actual recurrence status, revealed results similar to those of the all patients (Stages I–IV). Among the testing set data, 42 of 50 cases (84.0%) were Stages I–II and the HR group exhibited a significantly shorter RFS (Fig. 3C), validating its predictive performance in early-stage LUAD patients. OS was not significantly different. The detailed clinical and histopathological comparison data of this group are provided in Supplementary Tables S3 and S4.

Discussion

In this study, we developed a model to predict the risk of recurrence of LUAD by DL-based image analysis. This classification model showed good performance with high sensitivity, implying its potential usefulness as a screening tool. The model revealed an AUC of 0.763 in the testing set, which is better performance to the IASLC grade (an AUC of 0.690)⁹. The predicted risk groups were strongly correlated with histopathological features and several genetic mutations. Clinicopathologic results for stage I–II cases were virtually the same as those of the general group.

Pathological research typically sees strong AI model performance in areas where histological differences are easily recognized by pathologists. Unfortunately, in the case of LUAD, histological characteristics are diverse and complex, making it challenging for pathologists to discern differences easily. The present study was aimed an exploratory effort to determine if an AI model can successfully identify histological differences between recurrence and non-recurrence in early-stage lung cancer cases with partial resection—an unresolved challenge for pathologists. This study demonstrated the AI model's potential to predict recurrence in partially resected lung tissue, marking a significant achievement. If efforts to introduce more advanced models based on this research and develop algorithms explaining the model's decisions are attempted in future studies, it is anticipated that identifying patients in need of closer monitoring will become possible, leading to improved patient survival.

Lung cancer has various histological types such as LUAD, squamous cell carcinoma, and small cell lung cancer²⁸. Squamous cell carcinoma primarily originates in the central part of the lung, and when surgery is feasible, lobectomy is commonly performed. Therefore, this type of tumor is generally not considered a candidate for partial resection. In the case of small cell lung cancer, which also typically arises in the central region of the

Variables	Total	Low-risk (n=91)	High-risk (n=73)	p value
Age, year				
Mean (SD)	62.84 (10.67)	63.55 (10.65)	61.94 (10.69)	0.340
Sex				
Male	91 (55.5%)	45 (49.5%)	46 (63.0%)	0.114
Female	73 (44.5%)	46 (50.5%)	27 (37.0%)	
RFS, day				
Mean (SD)	1034.86 (555.42)	1178.57 (521.26)	855.71 (547.83)	< 0.001
OS, day				
Mean (SD)	1292.13 (483.20)	1341.74 (458.93)	1230.29 (508.24)	0.143
Tumor invasion size				
Mean (SD)	3.19 (1.76)	2.76 (1.43)	3.75 (1.97)	< 0.001
Predominant histologic type				
Lepidic	5 (3.0%)	5 (5.5%)	0 (0.0%)	< 0.001
Acinar	61 (37.2%)	38 (41.8%)	23 (31.5%)	
Papillary	43 (26.2%)	30 (33.0%)	13 (17.8%)	
Solid	33 (20.1%)	8 (8.8%)	25 (34.2%)	
Micropapillary	4 (2.4%)	2 (2.2%)	2 (2.7%)	
Cribiform	4 (2.4%)	1 (1.1%)	3 (4.1%)	
Mucinous	14 (8.5%)	7 (7.7%)	7 (9.6%)	
IASLC grade				
1	7 (4.3%)	6 (6.6%)	1 (1.4%)	< 0.001
2	51 (31.1%)	45 (49.5%)	6 (8.2%)	
3	106 (64.6%)	40 (44.0%)	66 (90.4%)	
PD component, %				
Mean (SD)	49.09 (38.60)	30.23 (34.66)	72.60 (29.34)	< 0.001
CGP component, %				
Mean (SD)	19.49 (25.16)	14.12 (20.93)	26.19 (28.34)	< 0.001
Necrosis				
Absent	100 (61.0%)	71 (78.0%)	29 (39.7%)	< 0.001
Present	64 (39.0%)	20 (22.0%)	44 (60.3%)	
STAS				
Absent	50 (30.5%)	37 (40.7%)	13 (17.8%)	0.003
Present	114 (69.5%)	54 (59.3%)	60 (82.2%)	
Pleural invasion				
Absent	105 (64.0%)	70 (76.9%)	35 (47.9%)	0.001
PL1	27 (16.5%)	10 (11.0%)	17 (23.3%)	
PL2	19 (11.6%)	8 (8.8%)	11 (15.1%)	
PL3	13 (7.9%)	3 (3.3%)	10 (13.7%)	
Lymphovascular invasion				
Absent	84 (51.2%)	62 (68.1%)	22 (30.1%)	< 0.001
Present	80 (48.8%)	29 (31.9%)	51 (69.9%)	
pT stage				
pT1	70 (42.7%)	52 (57.1%)	18 (24.7%)	< 0.001
pT2	58 (35.4%)	26 (28.6%)	32 (43.8%)	
pT3	26 (15.9%)	10 (11.0%)	16 (21.9%)	
pT4	10 (6.1%)	3 (3.3%)	7 (9.6%)	
pN stage				
pN0	117 (71.8%)	75 (83.3%)	42 (57.5%)	< 0.001
pN1	17 (10.4%)	3 (3.3%)	14 (19.2%)	
pN2	29 (17.8%)	12 (13.3%)	17 (23.3%)	
Stage group				
I	93 (56.7%)	64 (70.3%)	29 (39.7%)	< 0.001
II	33 (20.1%)	13 (14.3%)	20 (27.4%)	
III	36 (22.0%)	14 (15.4%)	22 (30.1%)	
IV	2 (1.2%)	0 (0.0%)	2 (2.7%)	

Table 2. Clinicopathological characteristics of patients according to the model-based risk group. *SD* standard deviation, *IASLC* International Association for the Study of Lung Cancer, *PD* poorly differentiated, *CGP* complex glandular pattern, *STAS* tumor spread through air spaces, *PL* level of pleural invasion (PL1, visceral pleural elastic layer; PL2, visceral pleural surface; PL3, parietal pleura and/or chest wall).

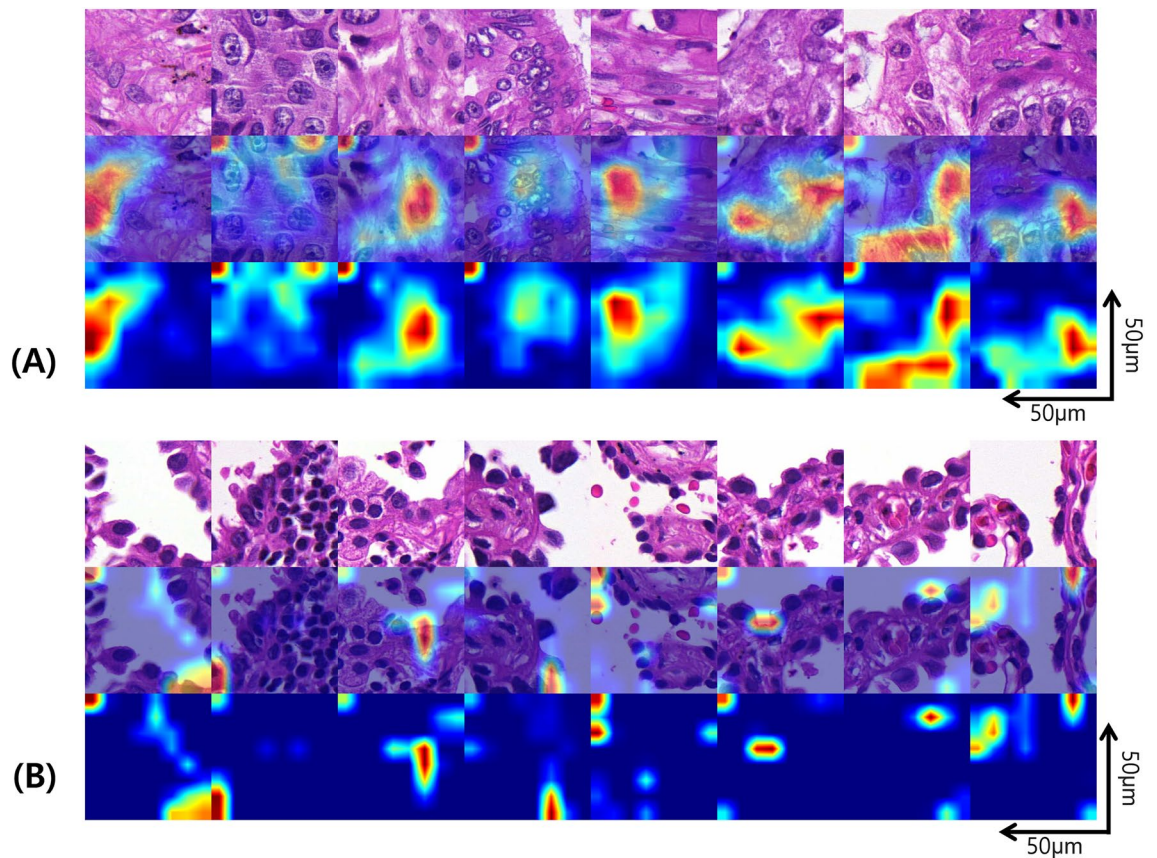


Figure 4. Class activation map of representative image patches. (A) Patches with the highest risk, (B) patches with the lowest risk.

lung, standard treatments include radiation therapy or chemotherapy. LUAD, the most common histological subtype at 38.5%, is experiencing a significant increase in incidence and is the most common subtype for which partial resection is performed²⁹. Considering the significant histological differences among these three types, we chose adenocarcinoma as the focus of our study to create a meaningful model, specifically predicting tumor recurrence after partial resection, for clinical practice. We anticipated that creating a model encompassing all three histological subtypes would be challenging due to their distinct characteristics. Additionally, considering the target application of the model, we judged that including all three tumors from a clinical perspective would not be suitable.

The model's output reflects histopathological features known to be associated with the tumor biology. The structural pattern is currently the most important factor in the histological subtyping of LUAD⁹. The HR group showed not only significantly higher proportions of PD and CGP components, but also more complex pattern in representative image patches than the LR group. Enlarged and pleomorphic nuclei in the HR patches are consistent with previous studies, which showed that nuclear size is more significantly associated with the prognosis than the nuclear to cytoplasm ratio (N/C ratio) in LUAD^{30,31}. In addition, we showed various histopathologic parameters like STAS, pleural invasion, and LVI were significantly associated with the HR group, although they might not be reflected in the patch-level evaluation of the model because they are usually observed in sparsely scattered areas around the tumor border. It suggests that the HR group has aggressive phenotype.

Detection of genomic alterations of LUAD by the DL-based model has been successful in previous studies^{19,32}. Our study also showed biological feature reflected by the model was its association with *TP53* alterations^{33,34}. *TP53* are tumor suppressor genes, and its mutations are known to be associated with tumor progression and poor prognosis^{33,34}. From the perspective of the tumor immune microenvironment, *TP53* alterations in LUAD have been reported to be associated with high infiltration of M0 macrophages and an immunosuppressive environment, along with *KRAS* mutations^{35,36}. These cases may have a high potential for the effectiveness of immune checkpoint inhibitors (ICIs). If the present model is tuned to more accurately predict *TP53* gene mutations, it could serve as a valuable screening test for selectively applying adjuvant ICI treatment, such as PD-L1 inhibitors in LUAD patients who have undergone partial resection at the early stages of *TP53* gene alterations.

This study and previous studies^{17,18} demonstrated the potential of DL-based risk prediction of LUAD using histopathological images. This study lies in the utilization of actual patient data, serving as the direct application target for the developed model, employing various DL architectures, and notably enhancing predictive power through the application of MIL. Moreover, the study not only confirmed the model's emphasis on distinguishing HR and LR recurrence groups by comparing detailed interpretations of a specialized pulmonary pathologist

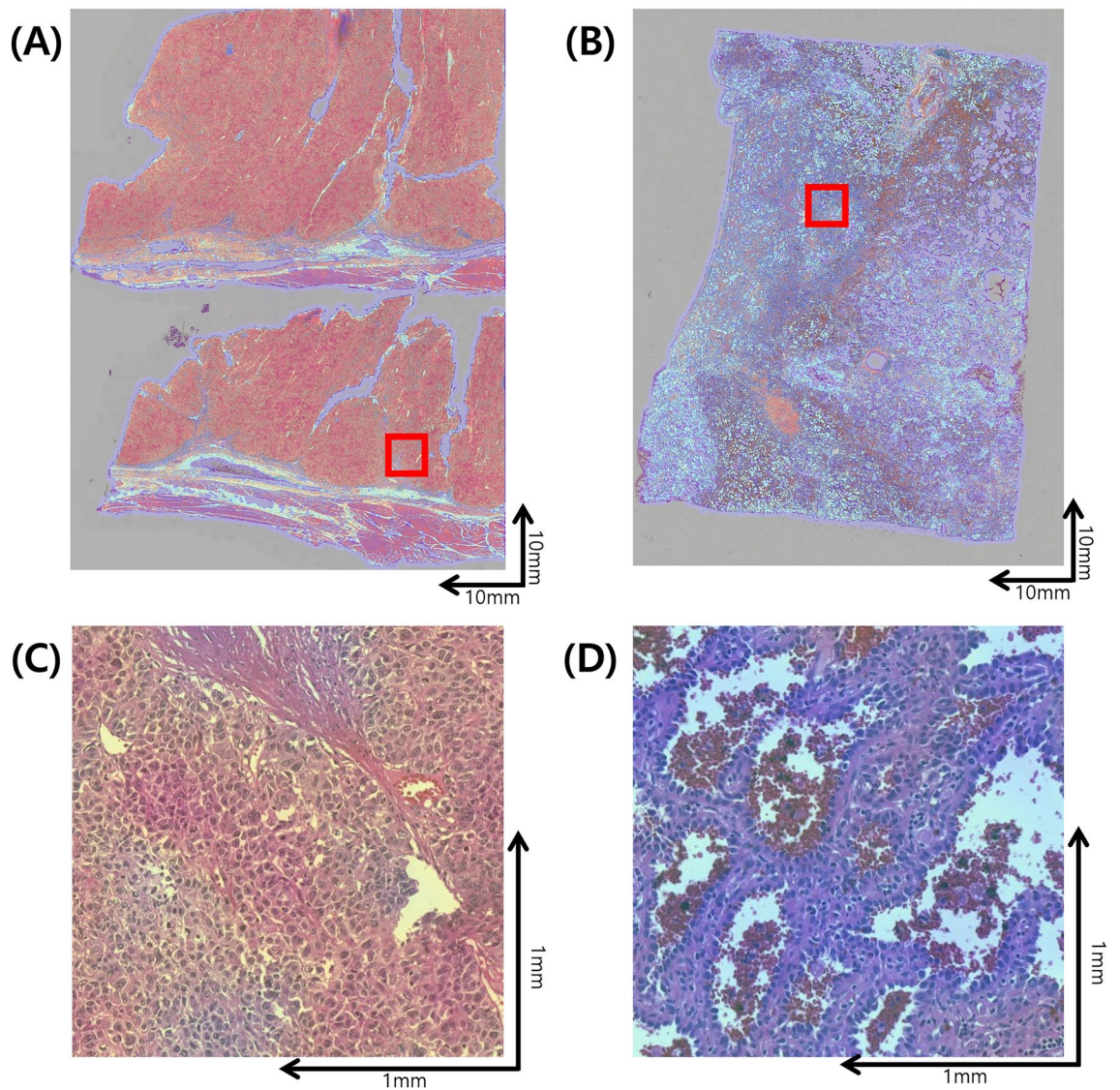


Figure 5. Heatmaps indicating tumor probability in whole slide images. (A) Whole slide image with the highest tumor probability classification. (B) Whole slide image with the lowest tumor probability classification. (C) Magnified image from slide (A) highlighting details. (D) Magnified image from slide (B) highlighting details.

and various cancer genetic variations but also elucidated the model's specific interpretability by highlighting its correlation with various histopathological findings and genetic changes currently crucial in LUAD pathology interpretation. The results from DL-based models were good but still suboptimal for clinical practice use. Insufficient data size, heterogenous histology of LUAD, confounding elements including epithelioid macrophages or lack of optimized DL architecture could limit the performance of histopathologic models. However, a study from the IASLC group showed that the power of histologic characteristics as a tool for prognosis prediction is limited⁹. A critical improvement could be achieved by a multidisciplinary approach, including clinical and genetic data along with histological features. Several studies have attempted such an approach^{37,38}, but they did not fully integrate pathological images into their models. Further studies are warranted.

In conclusion, the DL model showed good performance in recurrence prediction by analyzing histopathological images. The predicted risk group was associated with aggressive biological features. The model can provide useful information for the risk stratification and the selection of treatment of LUAD.

Genomic alterations	Low-risk (n = 91)	High-risk (n = 73)	p value
CDKN2A alteration			
Absent	77 (85.6%)	58 (79.5%)	0.413
Present	13 (14.4%)	15 (20.5%)	
TP53 alteration			
Absent	51 (56.7%)	25 (34.2%)	0.007
Present	39 (43.3%)	48 (65.8%)	
KRAS mutation			
Absent	78 (86.7%)	63 (86.3%)	1.000
Present	12 (13.3%)	10 (13.7%)	
EGFR mutation			
Absent	30 (33.3%)	34 (46.6%)	0.119
Present	60 (66.7%)	39 (53.4%)	
ALK translocation			
Absent	90 (100.0%)	71 (97.3%)	0.387
Present	0 (0.0%)	2 (2.7%)	
Genomic alterations	No recurrence (n = 96)	Recurrence (n = 68)	p value
CDKN2A alteration			
Absent	79 (82.3%)	56 (83.6%)	0.997
Present	17 (17.7%)	11 (16.4%)	
TP53 alteration			
Absent	56 (58.3%)	20 (29.9%)	<0.001
Present	40 (41.7%)	47 (70.1%)	
KRAS mutation			
Absent	84 (87.5%)	57 (85.1%)	0.831
Present	12 (12.5%)	10 (14.9%)	
EGFR mutation			
Absent	39 (40.6%)	25 (37.3%)	0.793
Present	57 (59.4%)	42 (62.7%)	
ALK translocation			
Absent	95 (99.0%)	66 (98.5%)	1.000
Present	1 (1.0%)	1 (1.5%)	

Table 3. Genomic alterations according to the model-based risk group and recurrence.

Data availability

Data will be made available on request to corresponding author and with the permission of the institutional review board of Asan Medical Center.

Received: 18 April 2023; Accepted: 12 March 2024

Published online: 16 March 2024

References

- Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer statistics, 2021. *CA Cancer J. Clin.* **71**, 7–33. <https://doi.org/10.3322/caac.21654> (2021).
- Yim, S. H. L. *et al.* Rise and fall of lung cancers in relation to tobacco smoking and air pollution: A global trend analysis from 1990 to 2012. *Atmos. Environ.* **269**, 118835. <https://doi.org/10.1016/j.atmosenv.2021.118835> (2022).
- Ettinger, D. S. *et al.* Non-small cell lung cancer, Version 3.2022, NCCN clinical practice guidelines in oncology. *J. Natl. Compr. Cancer Netw.* **20**, 497–530. <https://doi.org/10.6004/jnccn.2022.0025> (2022).
- Bowes, K. *et al.* Treatment patterns and survival of patients with locoregional recurrence in early-stage NSCLC: a literature review of real-world evidence. *Med. Oncol.* **40**, 1–8 (2023).
- Kadota, K. *et al.* Tumor spread through air spaces is an important pattern of invasion and impacts the frequency and location of recurrences after limited resection for small stage I lung adenocarcinomas. *J. Thorac. Oncol.* **10**, 806–814. <https://doi.org/10.1097/Jto.0000000000000486> (2015).
- Shih, A. R. & Mino-Kenudson, M. Updates on spread through air spaces (STAS) in lung cancer. *Histopathology* **77**, 173–180. <https://doi.org/10.1111/his.14062> (2020).
- Blaauwgeers, H., Russell, P. A., Jones, K. D., Radonic, T. & Thunnissen, E. Pulmonary loose tumor tissue fragments and spread through air spaces (STAS): Invasive pattern or artifact? A critical review. *Lung Cancer* **123**, 107–111. <https://doi.org/10.1016/j.lungcan.2018.07.017> (2018).
- Board, W. C. O. T. E. *Thoracic Tumours*, Vol. 5**** 5th edn. (International Agency for Research on Cancer, 2021).
- Moreira, A. L. *et al.* A grading system for invasive pulmonary adenocarcinoma: a proposal from the international association for the Study of Lung Cancer Pathology Committee. *J. Thorac. Oncol.* **15**, 1599–1610. <https://doi.org/10.1016/j.jtho.2020.06.001> (2020).
- Abels, E. *et al.* Computational pathology definitions, best practices, and recommendations for regulatory guidance: a white paper from the Digital Pathology Association. *J. Pathol.* **249**, 286–294. <https://doi.org/10.1002/path.5331> (2019).

11. Prabhu, S., Prasad, K., Robels-Kelly, A. & Lu, X. AI-based carcinoma detection and classification using histopathological images: A systematic review. *Comput. Biol. Med.* **142**, 105209. <https://doi.org/10.1016/j.compbiomed.2022.105209> (2022).
12. Sakamoto, T. *et al.* A narrative review of digital pathology and artificial intelligence: focusing on lung cancer. *Transl. Lung Cancer Res.* **9**, 2255–2276. <https://doi.org/10.21037/tlcr-20-591> (2020).
13. Chiu, H. Y., Chao, H. S. & Chen, Y. M. Application of artificial intelligence in lung cancer. *Cancers (Basel)* <https://doi.org/10.3390/cancers14061370> (2022).
14. Yu, K. H. *et al.* Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nat. Commun.* **7**, 12474. <https://doi.org/10.1038/ncomms12474> (2016).
15. Wang, X. *et al.* Prediction of recurrence in early stage non-small cell lung cancer using computer extracted nuclear features from digital H&E images. *Sci. Rep.* **7**, 13543. <https://doi.org/10.1038/s41598-017-13773-7> (2017).
16. Wang, S. *et al.* ConvPath: A software tool for lung adenocarcinoma digital pathological image analysis aided by a convolutional neural network. *EBioMedicine* **50**, 103–110. <https://doi.org/10.1016/j.ebiom.2019.10.033> (2019).
17. Wu, Z. *et al.* DeepLRHE: A deep convolutional neural network framework to evaluate the risk of lung cancer recurrence and metastasis from histopathology images. *Front. Genet.* **11**, 768. <https://doi.org/10.3389/fgene.2020.00768> (2020).
18. Shim, W. S. *et al.* DeepRePath: Identifying the prognostic features of early-stage lung adenocarcinoma using multi-scale pathology images and deep convolutional neural networks. *Cancers (Basel)* <https://doi.org/10.3390/cancers13133308> (2021).
19. Coudray, N. *et al.* Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat. Med.* **24**, 1559–1567. <https://doi.org/10.1038/s41591-018-0177-5> (2018).
20. Ahn, B. *et al.* Clinicopathologic and genomic features of high-grade pattern and their subclasses in lung adenocarcinoma. *Lung Cancer* **170**, 176–184 (2022).
21. Lee, G. *et al.* Blood vessel invasion predicts postoperative survival outcomes and systemic recurrence regardless of location or blood vessel type in patients with lung adenocarcinoma. *Ann. Surg. Oncol.* **28**, 7279–7290. <https://doi.org/10.1245/s10434-021-10122-x> (2021).
22. Amin, M. B., American Joint Committee on Cancer. & American Cancer Society. *AJCC cancer staging manual*. Eight edition / editor-in-chief, Mahul B. Amin, MD, FACP ; editors, Stephen B. Edge, MD, FACS and 16 others ; Donna M. Gress, RHIT, CTR - Technical editor ; Laura R. Meyer, CAPM - Managing editor. edn, (American Joint Committee on Cancer, Springer, 2017).
23. Tan, M. & Le, Q. In *International Conference on Machine Learning*, 6105–6114 (PMLR).
24. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (arXiv preprint) (2014).
25. Multiple instance learning model implemented in pytorch.
26. Saito, T. & Rehmsmeier, M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One* **10**, e0118432. <https://doi.org/10.1371/journal.pone.0118432> (2015).
27. Travis, W. D. *et al.* International association for the study of Lung Cancer/American Thoracic Society/European respiratory society international multidisciplinary classification of lung adenocarcinoma. *J. Thorac. Oncol.* **6**, 244–285. <https://doi.org/10.1097/JTO.0b013e318206a221> (2011).
28. Nicholson, A. G. *et al.* The 2021 WHO classification of lung tumors: impact of advances since 2015. *J. Thorac. Oncol.* **17**, 362–387 (2022).
29. Cruz, C. S. D., Tanoue, L. T. & Matthay, R. A. Lung cancer: epidemiology, etiology, and prevention. *Clin. Chest Med.* **32**, 605–644 (2011).
30. Nakazato, Y. *et al.* Nuclear grading of primary pulmonary adenocarcinomas: correlation between nuclear size and prognosis. *Cancer* **116**, 2011–2019. <https://doi.org/10.1002/ncr.24948> (2010).
31. von der Thusen, J. H. *et al.* Prognostic significance of predominant histologic pattern and nuclear grade in resected adenocarcinoma of the lung: potential parameters for a grading system. *J. Thorac. Oncol.* **8**, 37–44. <https://doi.org/10.1097/JTO.0b013e318276274e> (2013).
32. Mayer, C. *et al.* Direct identification of ALK and ROS1 fusions in non-small cell lung cancer from hematoxylin and eosin-stained slides using deep learning algorithms. *Mod. Pathol.* **35**, 1882–1887. <https://doi.org/10.1038/s41379-022-01141-4> (2022).
33. Jiao, X. D., Qin, B. D., You, P., Cai, J. & Zang, Y. S. The prognostic value of TP53 and its correlation with EGFR mutation in advanced non-small cell lung cancer, an analysis based on cBioPortal data base. *Lung Cancer* **123**, 70–75. <https://doi.org/10.1016/j.lungcan.2018.07.003> (2018).
34. Jao, K. *et al.* The prognostic effect of single and multiple cancer-related somatic mutations in resected non-small-cell lung cancer. *Lung Cancer* **123**, 22–29. <https://doi.org/10.1016/j.lungcan.2018.06.023> (2018).
35. Liu, Z. *et al.* Development and validation of an immune-related gene prognostic index for lung adenocarcinoma. *J. Thorac. Dis.* **15**, 6205 (2023).
36. Bischoff, P. *et al.* Outcome of first-line treatment with pembrolizumab according to KRAS/TP53 mutational status for non-squamous PD-L1 high ($\geq 50\%$) NSCLC in the German National Network Genomic Medicine Lung Cancer (nNGM). *J. Thorac. Oncol.* **20**, 20 (2023).
37. Lee, B. *et al.* DeepBTS: Prediction of recurrence-free survival of non-small cell lung cancer using a time-binned deep neural network. *Sci. Rep.* **10**, 1952. <https://doi.org/10.1038/s41598-020-58722-z> (2020).
38. Lui, N. *et al.* A new model using artificial intelligence to predict recurrence after surgical resection of stage I–II non-small cell lung cancer. *J. Clin. Oncol.* **39**, 8537–8537. https://doi.org/10.1200/JCO.2021.39.15_suppl.8537 (2021).

Author contributions

P.J.K. and H.G. designed research study. H.S.H., G.C., H.J.S., B.A. J.S.U., S.Y. and D.K. contributed to the collection of data. H.S.H., G.C., H.J.S., B.A. J.S.U., S.Y., D.K. S.M.C. and S.J.J. confirmation of pathologic slide and image. P.J.K. contributed to processing and manipulating the results. P.J.K. H.S.H., and H.G. contributed to interpreting the results, and writing the manuscript. All authors contributed to drafting the final manuscript and figures.

Funding

This research was supported by a Grant (Grant No.2019IE7028-1 and 2021IE0018-1) from the Asan Institute for Life Sciences, Asan Medical Centre, Seoul, Republic of Korea and by the National Research Foundation of Korea (NRF) grant funded by the Republic of Korea (MSIT) (Grant No. NRF-2020R1C1C1003834).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-56867-9>.

Correspondence and requests for materials should be addressed to H.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024