



## Research article

# Deep learning combined with radiomics may optimize the prediction in differentiating high-grade lung adenocarcinomas in ground glass opacity lesions on CT scans



Xing Wang<sup>a,1</sup>, Li Zhang<sup>b,1</sup>, Xin Yang<sup>d,1</sup>, Lei Tang<sup>e,1</sup>, Jie Zhao<sup>f</sup>, Gaoxiang Chen<sup>b</sup>, Xiang Li<sup>a</sup>, Shi Yan<sup>a</sup>, Shaolei Li<sup>a</sup>, Yue Yang<sup>a</sup>, Yue Kang<sup>g</sup>, Quanzheng Li<sup>c,\*\*</sup>, Nan Wu<sup>a,\*</sup>

<sup>a</sup> Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Department of Thoracic Surgery II, Peking University Cancer Hospital & Institute, Beijing, China

<sup>b</sup> Center for Data Science, Peking University, Beijing, China

<sup>c</sup> MGH/BWH Center for Clinical Data Science, Boston, MA 02115, USA

<sup>d</sup> Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Department of Pathology, Peking University Cancer Hospital & Institute, Beijing, China

<sup>e</sup> Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Department of Radiology, Peking University Cancer Hospital & Institute, Beijing, China

<sup>f</sup> Center for Data Science in Health and Medicine, Peking University, Beijing, China

<sup>g</sup> Linkdoc AI Research (LAIR), Building A, Sinosteel International Plaza, No.8 Haidian Street, Haidian District, Beijing, China

## ARTICLE INFO

## Keywords:

Deep learning  
Ground glass opacifications  
Micropapillary  
Non-small cell lung cancer (NSCLC)  
Radiomics

## ABSTRACT

**Purpose:** Adenocarcinoma (ADC) is the most common histological subtype of lung cancers in non-small cell lung cancer (NSCLC) in which ground glass opacifications (GGOs) found on computed tomography (CT) scans are the most common lesions. However, the presence of a micropapillary or a solid component is identified as an independent predictor of prognosis, suggesting a more extensive resection. The purpose of our study is to explore imaging phenotyping using a method combining radiomics with deep learning (RDL) to predict high-grade patterns within lung ADC.

**Methods:** Included in this study were 111 patients differentiated as having GGOs and pathologically confirmed ADC. Four different groups of methods were compared to classify the GGOs for the prediction of the pathological subtypes of high-grade lung ADCs in definitive hematoxylin and eosin stain, including radiomics with gray-level features, radiomics with textural features, deep learning method, and the RDL.

**Results:** We evaluated the performance of different models on 111 NSCLC patients using 4-fold cross-validation. The proposed RDL has achieved an overall accuracy of 0.913, which significantly outperforms the other methods ( $p < 0.01$ , analysis of variation, ANOVA). In addition, we also verified the generality and practical effectiveness of these models on an independent validation dataset of 28 patients. The results showed that our RDL framework with an accuracy of 0.966 significantly surpassed other methods.

**Conclusion:** High-grade lung ADC based on histologic pattern spectrum in GGO lesions might be predicted by the framework combining radiomics with deep learning, which reveals advantage over radiomics alone.

## 1. Introduction

Adenocarcinoma (ADC) is the most common histological subtype of lung cancer in most countries, accounting for almost half of all lung cancers. Lung ADC is notorious for its high chance of recurrence, even

for pathological stage I patients. Up to 40 % of them might have a recurrent disease at a later date [1–3]. Having such a high chance of recurrence indicates the necessity of improved prognostic markers to identify patients with high-grade early stage ADC, in which ground glass opacifications (GGOs) found on computed tomography (CT) scans

\* Corresponding author at: Department of Thoracic Surgery II, Peking University Cancer Hospital & Institute. No. 52, Fucheng Avenue, Haidian District, Beijing 100142, China.

\*\* Corresponding author.

E-mail addresses: [li.quanzheng@mgm.harvard.edu](mailto:li.quanzheng@mgm.harvard.edu) (Q. Li), [nanwu@bjmu.edu.cn](mailto:nanwu@bjmu.edu.cn) (N. Wu).

<sup>1</sup> X.W, L.Z X.Y and L.T contributed equally to this work.

are the most common lesions. Therefore the International Association for the Study of Lung Cancer (IASLC), the American Thoracic Society (ATS), and the European Respiratory Society (ERS) frequently revise the standards of classification on lung ADCs, which has been subtyped according to a new classification proposed in 2011 [4]. In this revised classification, invasive lung ADCs can be stratified according to malignant aggressiveness as indolent (ADC in situ, minimally invasive ADC, and lepidic predominant invasive ADC), intermediate (papillary and acinar predominant invasive ADC), and high-grade (solid and micropapillary predominant invasive ADC) based on a study of prognosis according to histological subtypes [5–9], which form a graded spectrum in terms of tumor malignancy.

Most GGOs on CT scans were diagnosed as early stage ADCs and were usually treated with limited resection. However, there has been increasing evidence suggesting that the characteristic histologic heterogeneity of lung ADCs also hints at the biologic diversity of individual tumors with respect to prognosis [10,11]. Meanwhile, the presence of a micropapillary or solid component (high-grade pattern) was identified as an independent predictor of local recurrence in patients treated with limited resection [12–14]. So the differentiating for the grade of malignancy have the potential to select candidates for at least lobular resection [15], thus the criteria of decision for less invasive surgery can be redefined with the additional histological subtype parameter besides whole tumor size, density or SUV (standard uptake value) on CT or PET (positron emission tomography)-CT scans [16]. However, the accurate classification remains difficult in radiology or frozen pathological analysis before or during surgery [17,18], so early-stage ADC presented as GGOs still faced potential recurrent risk in limited resection, so additional assistance were needed to facilitate the differentiating procedure of this subgroup.

Radiomics refers to the extraction and analysis of large amounts of advanced quantitative imaging features reflecting radiological spatial distributions within a tumor and has been expected to provide descriptive and predictive models relating image features to tumor phenotypes [19,20]. Studies show that imaging phenotyping using radiomics can predict micropapillary or solid patterns within lung ADC, however the reliable predicting model was still vacant. Recently, deep learning has achieved high success in the classification tasks of natural images [21–23]. Deep learning provides various high-level semantic information of an image (CT scan) that is different from image features extracted by radiomics. So we expect that deep learning is able to improve the predicting model of classic radiomics for the pathological types of GGOs.

A common approach to combine radiomics and deep learning models is to train the two models independently and to combine the results of the two models by voting or averaging strategies. For example, [24] and [25] use the mean probability output of radiomics and deep learning models, and achieve higher prediction accuracy of breast cancer than using one of the models alone [26]. [26] proposes to use an additional neural network to learn the weights between radiomics and deep learning, because different models may not contribute equally to the final result. The weighted average value of the output probability from radiomics and deep learning is applied to the classification of lung cancer.

Another approach is the feature-level fusion, which combines the different feature vectors extracted and computed by the two models into a new feature vector for the subsequent classification and prediction. This approach has been applied to tasks such as detection and classification of lung nodules [27,28], image attribute analysis of tumors [29], and prediction of cancer survival rates [30,31].

Although the existing methods have achieved satisfactory results in the classification of cancers, their applications are limited in the classification of certain sub-types of tumors. This is because fewer data are clinically available on each specific sub-type, and the differences in images between sub-types are more challenging to distinguish.

Inspired by the aforementioned research, we adopt the combined

model of radiomics and deep learning to identify high-grade ADC. Due to the heterogeneity inside tumors, single-scale models fail to effectively extract useful features of the tumors. So the proposed method uses the advanced multi-scale network architecture of feature extraction to explore imaging phenotyping and to predict high-grade patterns within lung ADC.

## 2. Material and methods

### 2.1. Patient eligibility

From January 2015 to June 2018, patients admitted to Peking University Cancer Hospital were enrolled in the study. Among them, patients differentiated as having GGOs by radiologists by NCCN criteria were collected according to the radiology report (DICOM data retrieved in Peking University Cancer Hospital). GGOs were defined as hazy opacity that does not obscure underlying bronchial structures or pulmonary vessels at computed tomography [32]. All cases diagnosed as ADC were collected and analyzed retrospectively.

### 2.2. Imaging acquisition

All patients underwent nonenhanced and arterial phase enhanced CT examinations with Discovery CT750 HD scanner (GE Medical Systems, Milwaukee, WI). All patients were scanned in the supine position, and the scan coverage was the whole chest. Nonionic contrast material was injected through the antecubital vein at a rate of 3.5 mL/sec (1.5 mL/kg of body weight, total of 90–120 mL, iohexol 300 mgI/mL [Omnipaque, GE Healthcare]). The arterial phase scans were carried out at 30 s after injection of the contrast medium. Imaging parameters for CT: helical mode at a tube voltage of 120 kVp, tube current of 300 mA, collimation thickness of 5 mm, reconstructed thickness of 0.625 mm, rotation speed of 0.6 s, helical pitch of 1.375, and CT dose index volume (CTDIvol) of 17.65 mGy. The Institutional Review Board of Peking University Cancer Hospital (Beijing, China) approved the study protocol and waived the need for written informed consent by each patient because of the retrospective nature of the study.

### 2.3. Pathological examination

Surgically resected specimens were fixed in formalin, embedded in paraffin, sectioned with a microtome, and stained with hematoxylin and eosin (H&E). All available H&E-stained tumor slides were independently reviewed by two different pathologists who were blinded to clinical outcomes (using an Leica DM3000 microscope with a standard eyepiece measuring 22 mm in diameter). The extent of all five growth patterns (lepidic, acinar, papillary, micropapillary, and solid) were semi-quantitatively recorded by the percentage present in the entire set of tumor slides in 5% increments.

Histological growth patterns were defined according to the 2015 World Health Organization classification of lung tumors as follows [4]:

- 1 Lepidic pattern: tumor cells with lepidic growth along alveolar walls and with no evidence of stromal, vascular, or pleural invasion;
- 2 Acinar pattern: invasive tumor arranged in acini and tubules and composed of cuboidal or columnar cells that resemble bronchial gland or bronchial-lining epithelial cells, including Clara cells;
- 3 Papillary pattern: invasive tumor composed of papillae structures with a fibrovascular core and complicated secondary and tertiary branches;
- 4 Micropapillary pattern: small papillary tufts composed of tumor cells with peripheral nuclei and no fibrovascular core (When air-space contains a micropapillary pattern, even if it is surrounded by lepidic or acinar structures, it should be classified as a micropapillary pattern);
- 5 Solid pattern: invasive tumor composed of nests or sheets of tumor

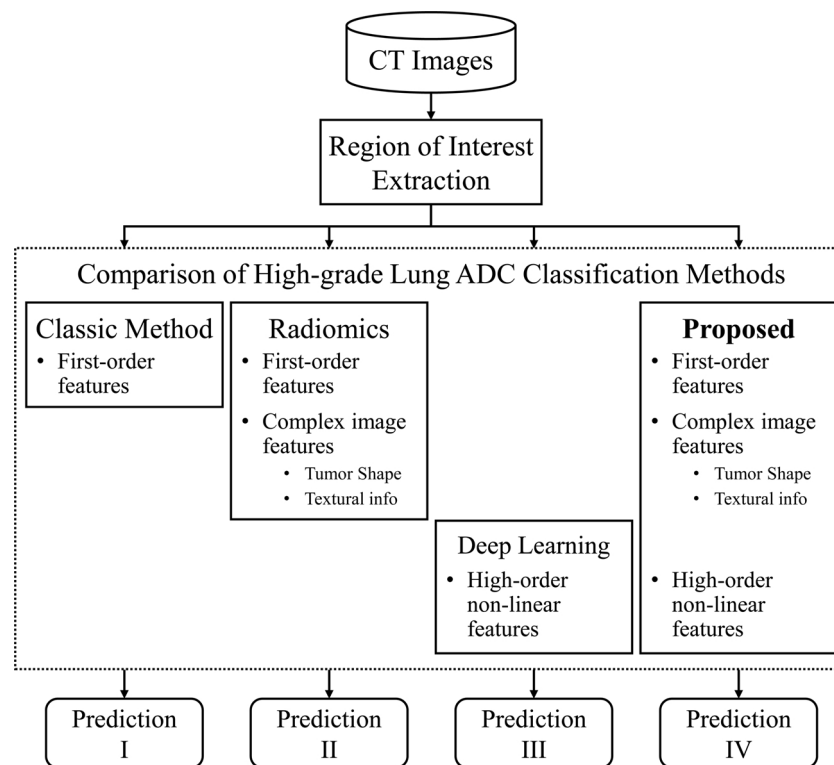


Fig. 1. Overview of the proposed method and the comparison of different high-grade lung ADC classification methods.

cells that lack acini, tubules, and papillae with mucin production.

Discrepancies between the two pathologists were later examined on a multiple-headed microscope, and the consensus was made after further discussion. The queried cases were reviewed by an experienced lung pathologist.

#### 2.4. Region of interest extraction

As shown in Fig. 1, four different groups of methods were compared, including the radiomics with gray-level features, the radiomics with textural features, the deep learning method, and the proposed RDL. The results of the pathological examination are treated as gold standards of the pathological subtypes of high-grade lung ADCs.

The three-dimensional (3D) morphology of the GGO was sketched by two experienced radiology experts (attending) based on its two-dimensional (2D) contour on each slice [33], using Osirix medical image workstation (Fig. 2). Based on the precise contour information, image processing techniques were employed to extract a specific 3D mask of each GGO, as well as its pseudo-color presentation in which different colors represented Hounsfield unit values inside the structure, accordingly, allowing an intuitional visualization for the study on each case. Visual implementation was programmed on Python platform based on Python 2D plotting library Matplotlib.

#### 2.5. Radiomics with gray-level features

For the radiomics with gray-level features, we used a K-nearest-neighbor (KNN) model to classify GGOs based on a small number of gradient-related features. Gradient operators, including Sobel filters (x- and y- directions), Robert filters (x- and y- directions) and Laplace operators were applied on the ground glass regions, respectively. Feature maps were generated and normalized to a range of 0–255 for analyses. We then calculated the histogram of the feature maps in terms of four bins with a size of 64 Gy levels, resulting in four adjacent intervals: [0, 64), [64, 128), [128, 192), and [192, 256). The numbers of

pixels in these intervals were used as inputs of the KNN model to classify GGOs. Thus, a total of 20 Gy-level features were selected for the KNN classification.

#### 2.6. Radiomics with textural features

Based on the region of interest extraction, we computed different types of radiomic features from the GGO regions to identify high-grade subtypes of lung ADCs. For each individual GGO region, we quantitatively extracted histograms and morphological features, as well as various textural features including gray co-occurrence matrix (GLCM), grayscale run length matrix (GLRLM), and grayscale size region matrices (GLSZM). A total of 107 commonly used radiomic features [34] were computed for further analyses (Table 1). We have exploited several machine learning approaches to classify GGOs, including logistic regression, Naive Bayes, decision tree and random forest. Based on the classification outputs, random forest showed the highest accuracy among all machine learning approaches. To further improve the accuracy of the model, we thus introduced a feature selection approach to the random forest model based on the information gain of each feature [35]. As shown in Table 2, 15 features were selected.

#### 2.7. Deep learning

Deep learning methods have achieved high success in multiple computer vision tasks, including image classification and image segmentation. In this paper, we attempted to classify pathological subtypes of GGOs using deep learning. We used a deep convolutional neural network to extract features from the original image. This network architecture was known as atrous spatial pyramid pooling (ASPP) network, which adopts a pyramidal convolution structure to extract multi-scale features [36]. As shown in Fig. 3, the ASPP network used four types of convolution kernels and integrated information of multiple scales to extract features of various sizes of GGOs.

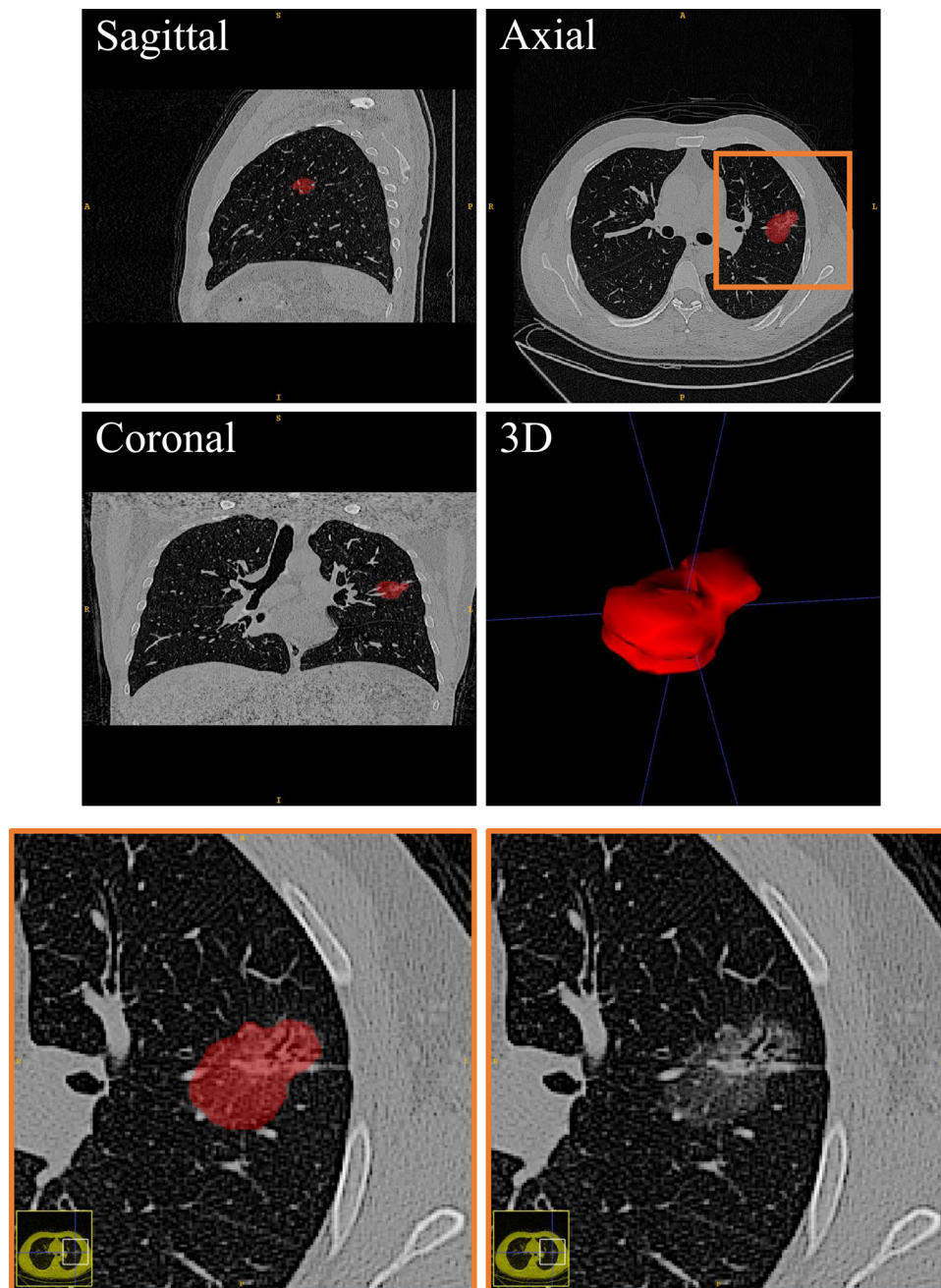


Fig. 2. Region of interest (ROI) for the ground glass opacifications on computed tomography scans.

## 2.8. Combining radiomics with deep learning

Deep neural networks extracted nonlinear features. Such features contained high-order semantic information and were usually different from radiomic features that generated by statistical models. Therefore, we attempted to combine the two types of features to generate an integrated feature sets of GGOs in CT scans, investigating whether the two types of features have complementary information of identifying the pathological subtypes of GGOs. As shown in Fig. 4, an autoencoding network is used to pre-train the deep neural network. The original image is projected on a low-order feature space (manifold), which is represented by a feature vector with 128 dimensions. Such high-order image information can be transformed into a feature vector with an arbitrary size. To balance the trade-off between radiomic features and deep-learning features, we choose 128 as the size of the deep-learning feature vector, which is close to the number of radiomic features. We

then constructed a multilayer neural network classifier and combined the 128 deep learning features with the 107 radiomic features to form a 235-dimensional feature vector. This new vector was then used to classify the pathological subtypes of GGO.

## 2.9. Network architectures and implementation

For the pure deep learning model, as shown in Fig. 3, the network backbone of deep neural networks was a convolutional neural network (CNN) with three stages. Each stage was consisting of a convolutional layer, a ReLU activation layer and a batch normalization layer. A pooling layer was added between different stages to adjust the receptive field of the model for better analyzing objects with different scales. We have also added the ASPP module to enhance multi-scale feature extraction. Due to the different GGO sizes of each patient, our deep learning model accepted input data of different sizes. We used the



**Table 1**  
All radiomic features for the prediction model.

First-Order Statistics		
10 Percentile	90 Percentile	Energy
Entropy	Interquartile Range	Kurtosis
Maximum	Mean Absolute Deviation	Mean
Median	Minimum	Range
Robust MAD	Root Mean Squared	Skewness
Total Energy	Uniformity	Variance
Shape Features		
Elongation	Flatness	Least Axis Length
Major Axis Length	Max2DDiameterColumn	Max2DDiameterRow
Max2DDiameterSlice	Max3DDiameter	Mesh Volume
Minor Axis Length	Sphericity	Surface Area
Surface Volume Ratio	Voxel Volume	
Gray Level Co-Occurrence Matrix		
Autocorrelation	Cluster Prominence	Cluster Shade
Cluster Tendency	Contrast	Correlation
Difference Average	Difference Entropy	Difference Variance
Inverse Difference (ID)	ID Moment	ID Moment
		Normalized
ID Normalized	Informational Correlation 1	Informational
		Correlation 2
Inverse Variance	Joint Average	Joint Energy
Joint Entropy	Max Correlation Coefficient	Max Probability
Sum Average	Sum Entropy	Sum Squares
Gray Level Run Length Matrix		
Non Uniformity (NU)	NU Normalized	Variance
High Run Emphasis	Long Run Emphasis	Long Run High
		Emphasis
Long Run Low	Low Run Emphasis	Run Entropy
Emphasis		
RL Non Uniformity	RL NU Normalized	Run Percentage
(NU)		
Run Variance	Short Run Emphasis	Short Run High
		Emphasis
Short Run Low		
Emphasis		
Gray Level Size Zone Matrix		
Non Uniformity (NU)	NU Normalized	Variance
High Zone Emphasis	Large Area Emphasis	Large Area High
		Emphasis
Large Area Low	Low Zone Emphasis	Size Zone NU
Emphasis		
Size Zone NU	Small Area Emphasis	Small Area High
Normalized		Emphasis
Small Area Low	Zone Entropy	Zone Percentage
Emphasis		
Zone Variance		
Gray Level Dependence Matrix		
Entropy	Non Uniformity (NU)	NU Normalized
Dependence Variance	Dependence NU	Variance
High Emphasis	Large Emphasis	Large High Emphasis
Large Low Emphasis	Low Emphasis	Small Emphasis
Small High Emphasis	Small Low Emphasis	
Neighboring Gray Tone Difference Matrix		
Busyness	Coarseness	Complexity
Contrast	Strength	

Average Pooling layer to integrate all deep learning features into a 128-dimensional feature vector, which represented high-order nonlinear image information extracted by the CNN for GGOs. Finally, we used the fully connected layer to classify such 128-dimensional feature vector to determine if the patient has a high-grade lung ADC.

To improve the training of the model that combined radiomics with deep learning (Fig. 3), we first pre-trained the deep neural network in an unsupervised way. We added a decoding CNN structure (green modules in Fig. 4) that restores the original image from the feature vector, forming an unsupervised training process with auto-encoding functionality [37]. We then used transfer learning [38] to import the unsupervised pre-trained network parameters into the trained model as the initial parameters. Furthermore, different from the pure depth learning model, the RDL model combined 128-dimensional deep learning features and 107-dimensional radiomic features to distinguish

**Table 2**  
Selected radiomic features based on information gain approach.

First-Order Statistics		
Robust MAD	Skewness	
Shape Features		
Elongation	Major Axis Length	Max2DDiameterColumn
Max2DDiameterRow	Max3DDiameter	Minor Axis Length
Gray Level Co-Occurrence Matrix		
Autocorrelation	Informational Correlation	Gray Level Nonuniformity
	1	
Correlation	Informational Correlation	
	2	
Gray Level Run Length Matrix		
Long Run Emphasis		
Neighboring Gray Tone Difference Matrix		
Strength		

the patients with high-grade lung ADC.

The optimization method of the above models used Adam algorithm [39]. The learning rate started from 0.0001 and decreased to one-tenth of the former every 30 epochs. Since the sizes of the model input data are different, we can only set the training batch size to 1. Model training and evaluation were run on pytorch0.4, using an NVIDIA GeForce GTX 1080ti GPU with 11GB memory.

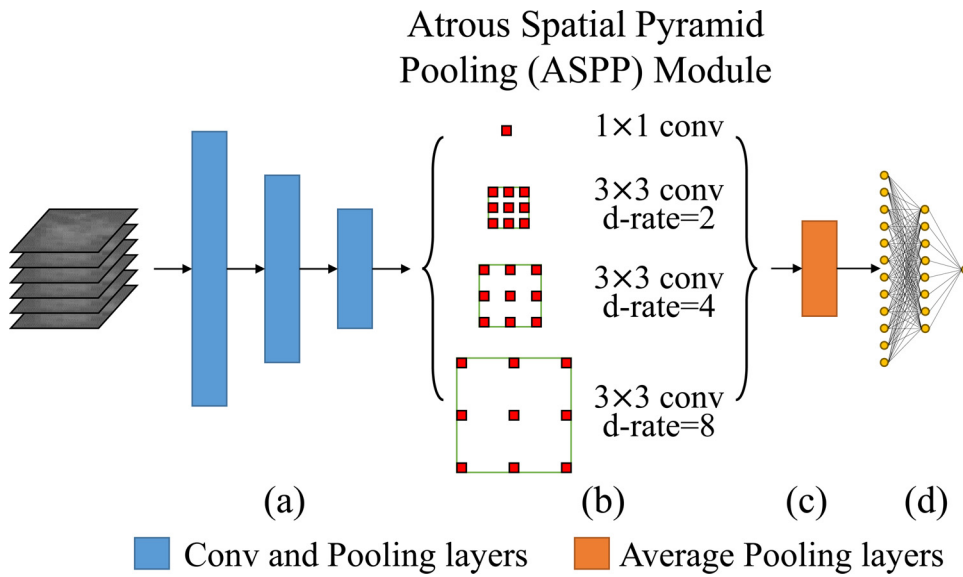
We trained all models based on the dataset of 111 NSCLC patients and evaluated them using a 4-fold cross-validation method. We also tested such models on an independent dataset of 28 patients to verify the generality and practical effectiveness of these models. The overall accuracy, sensitivity, and specificity of the classification were will be reported to measure the performance of all the models.

### 3. Results

In this work, 111 patients participated in the study (with a ratio being 0.48:1 of male to female), and 31 of them were diagnosed as having high-grade GGOs with micropapillary and solid patterns. Four different groups of methods to distinguish two classes of patients (0 for those without micropapillary/solid patterns and 1 for those with micropapillary/solid patterns) have been designed and evaluated. Such groups of methods include (1) radiomics with gray-level features, (2) radiomics with textural features, (3) deep learning, and (4) RDL. Fourfold cross-validation was used to compare these methods because of the limited size of the dataset.

We studied CT images of 111 patients with clinically confirmed lung ADC. The average age of these patients was 61.1 years (min: 35; max: 79), the proportion of female patients was 67.6 % (75 out of 111), and the patients with smoking habits accounted for 19.8 % (22 out of 111). After performing a pathological examination based on the patient's surgically resected specimens, we identified a total of 31 patients with micropapillary and/or solid patterns. We defined them as subjects with high-grade lung ADC in our dataset. For patients with high-grade lung ADC, the average age is 63.5 years, with a minimum of 40 and a maximum of 78, with a female patient ratio of 61.3 % (19 out of 31). Among them, patients with smoking habits accounted for 32.3 % (10 out of 31). We also tested the trained models on an independent dataset with 28 patients to verify the generalizability of the models. This dataset included 9 patients with micropapillary and/or solid patterns. The average age of these patients was 61.6 years (min: 46; max: 73), the proportion of female patients was 67.9 % (19 out of 28).

For the radiomics with gray-level features, 20 gradient-related features were examined. A two-tail Student *t*-test was applied to compare the feature values of the two classes. Five features with the lowest *p*-values are displayed (Table 3). Two features, the Laplace operator (second-order gradient) and the Sobel filter (first-order gradient), have statistically significant differences between classes 0 and 1. The accuracy, sensitivity, and specificity of the radiomics with gray-level



**Fig. 3.** Network structure of deep neural network with four major functional modules: (a) Convolutional and pooling layers (blue blocks) extract high-order semantic information. (b) ASPP module is used to calculating regional information of ground glass opacifications (GGOs) in multiple spatial scales. (c) A concatenation operation is used to aggregate such multi-scale information, and an average pooling layers is used to resample the feature maps to an appropriate size. (d) A fully-connected network is implemented for the final classification of pathological subtypes of GGOs.

features are 0.784, 0.911, 0.469, respectively. For the radiomics with textural features, Table 4 showed the classification outputs, and the random forest classifier reached the highest performance among the classic machine learning approaches with an accuracy of 0.802, a sensitivity of 0.899, and a specificity of 0.562. The information gain approach allows for the selection of 15 of the most effective features from the full set of features. The accuracy, sensitivity, and specificity of the random forest classifier after feature selection are then improved to 0.847, 0.937, and 0.625, respectively.

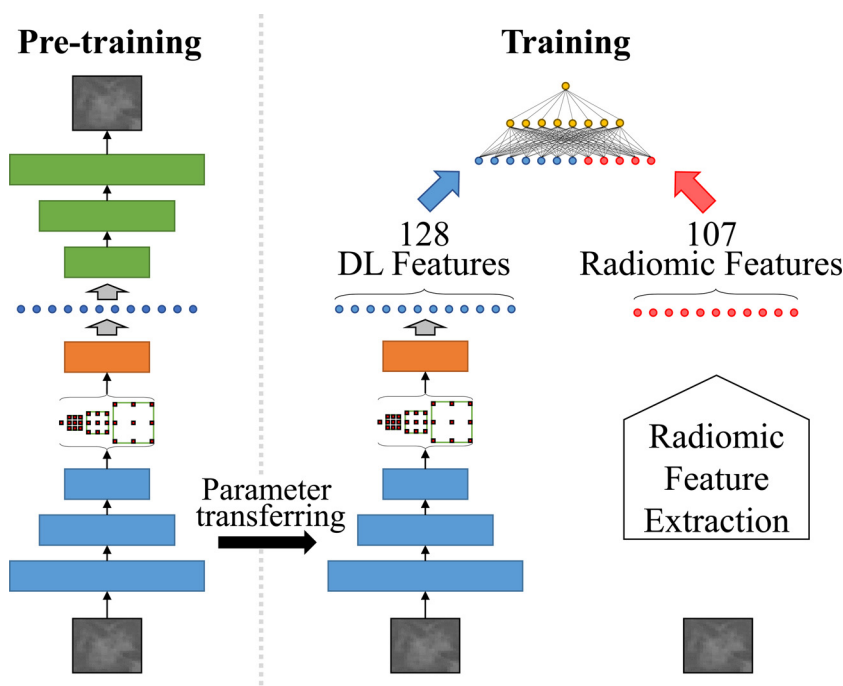
Specific design of image features in the deep learning model was not introduced. So that all features are automatically computed. The pure deep learning model's accuracy, sensitivity, and specificity are 0.744, 0.610, and 0.790, respectively. The accuracy, sensitivity, and specificity of the RDL model are 0.913, 0.734, and 0.964, respectively. Table 4 summarizes the results of all different methods with a classification cut-off of 0.5. Statistically, the proposed RDL outperforms other methods with a  $p$ -value  $< 0.01$  under the analysis of variation (ANOVA) using

**Table 3**

Examples of gradient-related features.

	With Micropapillary	Without Micropapillary	$p$ -Value
Laplace [64,128]	2636.40 (3085.77)	830.44 (1251.40)	0.0044
Sobel (gx) [64,128]	4059.10 (5378.17)	1040.10 (1204.22)	0.0054
Robert2 [64,128]	1999.87 (2858.87)	545.83 (1072.06)	0.0118
Robert1 [64,128]	3712.57 (5716.93)	854.52 (1044.38)	0.0120
Sobel (gy) [128,192]	1384.30 (2510.69)	238.87 (563.33)	0.0200

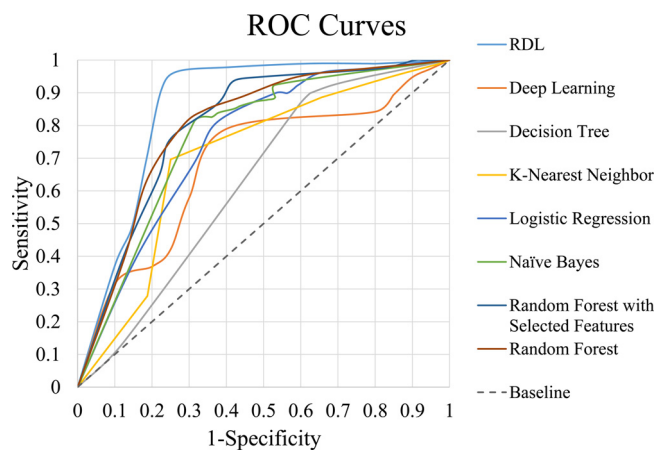
MATLAB software. In addition to the accuracy, sensitivity, and specificity, we have evaluated different methods using receiver operating characteristic (ROC) curves (Fig. 5). The proposed RDL achieves a high area under curve (AUC) value of 0.861, which also outperforms other methods as shown in Table 4. In addition, all the models were tested on 28 independent data. Table 5 showed the results of different models on this independent dataset, where the RDL framework significantly



**Fig. 4.** A classification framework that combines radiomics with deep learning: (1) Left part: the parameters of the auto-encoder are optimized during an offline pretraining process, and (2) Right part: 128 deep learning features and 107 radiomic features are fused to form a new feature vector that is used for the final prediction of the pathological subtypes of GGOs.

**Table 4**  
Evaluation of different methods: Accuracy, Sensitivity, Specificity, and AUC value.

Methods	Accuracy	Sensitivity	Specificity	AUC
Radiomics with gray-level features				
K-Nearest Neighbor	0.784	0.911	0.469	0.744
Radiomics with textural features				
Naïve Bayes	0.775	0.861	0.562	0.798
Logistic Regression	0.757	0.873	0.469	0.757
Decision Tree	0.730	0.861	0.406	0.673
Random Forest (RF)	0.802	0.899	0.562	0.819
RF with Feature Selection	0.847	0.937	0.625	0.818
Deep learning related models				
Deep Learning	0.744	0.610	0.790	0.689
Proposed RDL	0.913	0.734	0.964	0.861



**Fig. 5.** The ROC curves of different methods: the proposed RDL shows the largest area under curve.

**Table 5**  
Evaluation of different methods on an independent dataset: Accuracy, Sensitivity, and Specificity.

Methods	Accuracy	Sensitivity	Specificity
Radiomics with gray-level features			
K-Nearest Neighbor	0.786	0.786	0.548
Radiomics with textural features			
Naïve Bayes	0.821	0.854	0.798
Logistic Regression	0.801	0.813	0.769
Decision Tree	0.786	0.793	0.665
Random Forest (RF)	0.929	0.930	0.908
RF with Feature Selection	0.893	0.901	0.874
Deep learning related models			
Deep Learning	0.831	0.896	0.795
Proposed RDL	0.966	0.901	1.000

exceeded other methods with an accuracy of 0.966.

#### 4. Discussion

The increasing number of patients with ADC presenting as GGO is a critical problem in determining the grade of malignancy. Detailed preoperative or intraoperative examinations of histological subtype may provide important information for prediction of tumor progression and decisions regarding the surgical procedure [40]. Either an extensive or an insufficient dissection of lung tissues would bring about unnecessary damage and risk to the patients [41]. Our results provided a competitive accuracy for the prediction of high-grade subtype for GGOs [42], so we might assume that if the features from the DICOM data were extracted previously and the prediction was made, surgeons might choose the adequate extent for resection or lymphadenectomy as soon

as the specimen is confirmed as ADC from a frozen pathological result, thus changed surgical planning. We shall recommend this tool for the surgeons before resection of GGOs. Surgeons would know the result before resection, so they might choose lobectomy other than wedge resection if high-grade result were calculated, and at the same time, pathologists might be alerted and make more efforts to look for evidence for high-grade patterns like micropapillary during frozen pathological analysis.

Although the criteria for subtype differentiation were quite clear, heterogeneity and misdiagnosis still take place occasionally [43], as the accuracy of diagnosis of micropapillary pattern was poor [44,45] due to the deviation of tissue sampling. Therefore the RDL framework might be a supplemental tool for pathologists.

To predict the high-grade subtypes for GGOs, several methods have been designed and evaluated. By observing the CT image, micropapillary and solid patterns in GGOs were found to usually have different Hounsfield unit values from other areas. Such a difference can be represented by gray-level features, which are used to distinguish between high-grade GGOs and other types. In the experiment, we verified that there exists a statistical difference in the first- and second-order gray-level features of GGOs with different degrees of aggressiveness. However, the internal textures of GGOs are complex, so, unlike the classification of ordinary lung tissues or other types of lung tumors, it is difficult to obtain an accurate GGO classification using a small number of gray-level features.

Radiomics with textural features helps resolve such issues. For example, in this work, the gray level co-occurrence matrix (GLCM) is used to compute a subset of radiomic features. GLCM describes a set of the recurring image textures, which are the joint distributions of Hounsfield unit values (pixel/voxel intensity) with a certain positional relationship. For tumors (GGOs), such textures may correspond to special pathological structures. As shown in Table 4, radiomic features can significantly improve the accuracy of the classification model. However, the feature extraction process of radiomics is relatively fixed, which ignores the individual differences of patients. We expect to find a more flexible method to further improve the accuracy of the classification model.

Deep learning has achieved great success in the fields of image analysis and computer vision and has accomplished many complicated image classification tasks. However, during the experiments, it has been found that pure deep learning is not able to effectively improve the classification results. As shown in Table 4, the deep learning method produces a sensitivity of only 0.610 and a specificity of 0.790. Different from the lower order statistical features extracted by radiomics, the features extracted by pure deep learning with such limited data are only higher order semantic features. Establishing the relationship between these complex high order features and the classification results requires a large amount of training data. However, clinical image dataset is often difficult to meet such requirements.

In addition to increasing the training dataset, an alternative way is to impose a constraint on deep learning. Based on the previous discussion, radiomics plays the role of such constraint. Combining radiomics and deep learning is thus able to effectively classify GGO on the small image dataset in this work. The lower order features extracted by radiomics can be used to describe the general morphological, positional, and textural information of GGOs whereas deep learning can extract higher-order features to form a more personalized classification model for each patient. Experiments have shown that the combination of the two methods has indeed achieved the highest accuracy.

This paper has proposed a method to combine radiomics and deep learning (RDL) to classify GGOs. Although we show the more accurate results of the proposed method, several limitations are worth further investigation:

1. Only 31 positive samples of high-grade GGOs are included in this work. GGO images with micropapillary or solid patterns are difficult to collect, so previous works [46] did not use large amounts of data as

well. Although an independent dataset is used to test the generalizability and practical effectiveness of the RDL framework, a dataset with richer learning targets could greatly improve the performance of a model, especially in the field of deep learning. So we will try to collect more representative data for the analyses of radiomics and deep learning in the subsequent works.

2. This work involves retrospective cross-sectional data of ADCs, thus only the classification of existing GGO data is feasible. In future works, we will strengthen the research of longitudinal data, training the models to pursue prospective studies such as to predict the speed of GGO changes or to perform invasive examinations.

3. This model hasn't been validated to test GGO before surgery yet, while that should be the best scenario for this predicting tool, so we need to do more work to promote the use of this model for further study. We still need to focus on how this tool could help improving the accuracy of diagnoses from the aspect of pathologists.

In conclusion, high-grade lung ADCs based on histological pattern spectrum in GGO lesions from CT scans might be predicted by a framework combining radiomics with deep learning. This method reveals significant advantages over the radiomic methods with different input features, or deep learning alone. The exploration of tumor images according to these results may contribute to better treatment planning and personalized medicine.

## Funds

This work was supported by National Key R&D Program of China (No. 2018YFC0910700); Beijing Municipal Science & Technology Commission (No. Z161100000516063); Beijing Human Resources and Social Security Bureau (Beijing Millions of Talents Project, 2018A05); Special Fund of Beijing Municipal Administration of Hospitals Clinical Medicine Development (No. XMLX201841); The National Natural Science Foundation of China (NSFC) under Grants 81801778, 11831002; Beijing Municipal Administration of Hospitals Youth Programme (No. QML20171103); Peking University Medicine Seed Fund for Interdisciplinary Research (BMU2018MX008).

## CRedit authorship contribution statement

**Xing Wang:** Conceptualization, Methodology, Software. **Li Zhang:** Data curation, Writing - original draft. **Xin Yang:** Methodology. **Lei Tang:** Visualization, Investigation. **Jie Zhao:** Methodology. **Gaoxiang Chen:** Software, Validation. **Xiang Li:** Writing - review & editing. **Shi Yan:** Writing - review & editing. **Shaolei Li:** Writing - review & editing. **Yue Yang:** Supervision. **Yue Kang:** Software. **Quanzheng Li:** Supervision. **Nan Wu:** Supervision.

## Declaration of Competing Interest

The authors have no conflicts of interest to declare.

## References

- [1] K. Shedden, et al., Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study, *Nat. Med.* 14 (August 8) (2008) 822–827, <https://doi.org/10.1038/nm.1790>.
- [2] C.F. Mountain, Staging classification of lung cancer. A critical evaluation, *Clin. Chest Med.* 23 (March 1) (2002) 103–121 [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/11901906>.
- [3] G.M. Strauss, D.J. Kwiatkowski, D.H. Harpole, T.J. Lynch, A.T. Skarin, D.J. Sugarbaker, Molecular and pathologic markers in stage I non-small-cell carcinoma of the lung, *J. Clin. Oncol.* 13 (May 5) (1995) 1265–1279, <https://doi.org/10.1200/JCO.1995.13.5.1265>.
- [4] W.D. Travis, 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 (February 2) (2011) 244–285, <https://doi.org/10.1097/JTO.0b013e318206a221>.
- [5] M. Okada, Subtyping lung adenocarcinoma according to the novel 2011 IASLC/ATS/ERS classification: correlation with patient prognosis, *Thorac. Surg. Clin.* 23 (May 2) (2013) 179–186, <https://doi.org/10.1016/j.thorsurg.2013.01.001>.
- [6] H. Sakurai, et al., Differences in the prognosis of resected lung adenocarcinoma according to the histological subtype: a retrospective analysis of Japanese lung cancer registry data, (in eng), *Eur. J. Cardiothorac. Surg.* 45 (January 1) (2014) 100–107, <https://doi.org/10.1093/ejcts/ezt284>.
- [7] A. Warth, et al., The novel histologic International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society classification system of lung adenocarcinoma is a stage-independent predictor of survival, (in eng), *J. Clin. Oncol.* 30 (May 13) (2012) 1438–1446, <https://doi.org/10.1200/jco.2011.37.2185>.
- [8] N. Yanagawa, S. Shiono, M. Abiko, S.Y. Ogata, T. Sato, G. Tamura, New IASLC/ATS/ERS classification and invasive tumor size are predictive of disease recurrence in stage I lung adenocarcinoma, *J. Thorac. Oncol.* 8 (May 5) (2013) 612–618, <https://doi.org/10.1097/JTO.0b013e318287c3eb>.
- [9] A. Yoshizawa, et al., Impact of proposed IASLC/ATS/ERS classification of lung adenocarcinoma: prognostic subgroups and implications for further revision of staging based on analysis of 514 stage I cases, (in eng), *Mod. Pathol.* 24 (May 5) (2011) 653–664, <https://doi.org/10.1038/modpathol.2010.232>.
- [10] S. Gao, et al., Micropapillary and/or solid histologic subtype based on pre-treatment biopsy predicts local recurrence after thermal ablation of lung adenocarcinoma, (in eng), *Cardiovasc. Intervent. Radiol.* 41 (February 2) (2018) 253–259, <https://doi.org/10.1007/s00270-017-1760-8>.
- [11] N. Kobayashi, et al., Non-BAC component but not epidermal growth factor receptor gene mutation is associated with poor outcomes in small adenocarcinoma of the lung, (in eng), *J. Thorac. Oncol.* 3 (July 7) (2008) 704–710, <https://doi.org/10.1097/JTO.0b013e31817c6080>.
- [12] J.J. Hung, et al., Prognostic value of the new International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society lung adenocarcinoma classification on death and recurrence in completely resected stage I lung adenocarcinoma, *Ann. Surg.* 258 (December 6) (2013) 1079–1086, <https://doi.org/10.1097/SLA.0b013e31828920c0>.
- [13] J.J. Hung, et al., Predictive value of the international association for the study of lung cancer/American Thoracic Society/European Respiratory Society classification of lung adenocarcinoma in tumor recurrence and patient survival, *J. Clin. Oncol.* 32 (August 22) (2014) 2357–2364, <https://doi.org/10.1200/JCO.2013.50.1049>.
- [14] T. Miyoshi, et al., Early-stage lung adenocarcinomas with a micropapillary pattern, a distinct pathologic marker for a significantly poor prognosis, *Am. J. Surg. Pathol.* 27 (January 1) (2003) 101–109 [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/12502932>.
- [15] J.H. Pedersen, Z. Saghir, M.M. Wille, L.H. Thomsen, B.G. Skov, H. Ashraf, Ground-glass opacity lung nodules in the era of lung Cancer ct screening: radiology, pathology, and clinical management, *Oncology (Williston Park)* 30 (March 3) (2016) 266–274 [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/26984222>.
- [16] T. Mimae, et al., What are the radiologic findings predictive of indolent lung adenocarcinoma? (in eng), *Jpn. J. Clin. Oncol.* 45 (April 4) (2015) 367–372, <https://doi.org/10.1093/jcco/hyv005>.
- [17] Y.C. Yeh, J.I. Nitadori, K. Kadota, A. Yoshizawa, W.D. Travis, Using frozen section to identify histologic patterns in stage I lung adenocarcinoma  $\leq 3$  cm: accuracy and interobserver agreement, *Histopathology* 66 (7) (2014).
- [18] H.E.T. Bittar, P. Incharoen, A.D. Althouse, S. Dacic, Accuracy of the IASLC/ATS/ERS histological subtyping of stage I lung adenocarcinoma on intraoperative frozen sections, *Modern Pathology An Official Journal of the United States & Canadian Academy of Pathology Inc* 28 (8) (2015) 1058.
- [19] V. Kumar, et al., Radiomics: the process and the challenges, *Magn. Reson. Imaging* 30 (November 9) (2012) 1234–1248, <https://doi.org/10.1016/j.mri.2012.06.010>.
- [20] P. Lambin, et al., Radiomics: extracting more information from medical images using advanced feature analysis, (in eng), *Eur. J. Cancer* 48 (March 4) (1990) 441–446, <https://doi.org/10.1016/j.ejca.2011.11.036>.
- [21] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016) 770–778.
- [22] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, *Advances in Neural Information Processing Systems*, (2012), pp. 1097–1105.
- [23] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv 1409.1556* (2014).
- [24] B.Q. Huynh, H. Li, M.L. Giger, Digital mammographic tumor classification using transfer learning from deep convolutional neural networks, *J. Med. Imaging* 3 (3) (2016) 034501.
- [25] N. Antropova, B.Q. Huynh, M.L. Giger, A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets, *Med. Phys.* 44 (10) (2017) 5162–5171.
- [26] S. Liu, Y. Xie, A. Jirapatnakul, A.P. Reeves, Pulmonary nodule classification in lung cancer screening with three-dimensional convolutional neural networks, *J. Med. Imaging* 4 (4) (2017) 041308.
- [27] L. Fu, J. Ma, Y. Ren, Y.S. Han, J. Zhao, Automatic detection of lung nodules: false positive reduction using convolution neural networks and handcrafted features, *Medical Imaging 2017: Computer-Aided Diagnosis* 10134 International Society for Optics and Photonics, 2017, p. 101340A.
- [28] S. Chen, et al., Automatic scoring of multiple semantic attributes with multi-task feature leverage: a study on pulmonary nodules in CT images, *IEEE Trans. Med. Imaging* 36 (3) (2016) 802–814.
- [29] B.-C. Kim, Y.S. Sung, H.-I. Suk, Deep feature learning for pulmonary nodule classification in a lung CT, 2016 4th International Winter Conference on Brain-Computer Interface (BCI), IEEE, 2016, pp. 1–3.
- [30] J. Lao, et al., A deep learning-based radiomics model for prediction of survival in



- glioblastoma multiforme, *Sci. Rep.* 7 (1) (2017) 1–8.
- [31] R. Paul, et al., Deep feature transfer learning in combination with traditional features predicts survival among patients with lung adenocarcinoma, *Tomography* 2 (4) (2016) 388.
- [32] J.H. Austin, et al., Glossary of terms for CT of the lungs: recommendations of the nomenclature committee of the fleischner society, *Radiology* 200 (no. 2) (1996) 327–331.
- [33] R. Thawani, et al., Radiomics and radiogenomics in lung cancer: a review for the clinician, *Lung Cancer* 115 (2018) 34–41, <https://doi.org/10.1016/j.lungcan.2017.10.015> January.
- [34] J.J.M.V. Griethuysen, A. Fedorov, C. Parmar, A. Hosny, H.J.W.L. Aerts, Computational radiomics system to decode the radiographic phenotype, *Cancer Res.* 77 (21) (2017) e104–e107.
- [35] C.A. Kumar, M. Sooraj, S.J.Pcs. Ramakrishnan, A Comparative Performance Evaluation of Supervised Feature Selection Algorithms on Microarray Datasets vol. 115, (2017), pp. 209–217.
- [36] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, *IEEE Trans. Pattern Anal. Mach. Intell.* 40 (4) (2017) 834–848.
- [37] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, P.-A. Manzagol, Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion, *J. Mach. Learn. Res.* 11 (December) (2010) 3371–3408.
- [38] S.J. Pan, Q. Yang, A survey on transfer learning, *IEEE Trans. Knowl. Data Eng.* 22 (10) (2009) 1345–1359.
- [39] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, *arXiv preprint arXiv:1412.6980* (2014).
- [40] M. Mikubo, et al., Relevance of intraoperative pleural lavage cytology and histologic subtype in lung adenocarcinoma, *Ann. Thorac. Surg.* 106 (December 6) (2018) 1654–1660, <https://doi.org/10.1016/j.athoracsur.2018.07.035>.
- [41] P.A. Russell, Z. Wainer, G.M. Wright, M. Daniels, M. Conron, R.A. Williams, Does lung adenocarcinoma subtype predict patient survival?: a clinicopathologic study based on the new International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society international multidisciplinary lung adenocarcinoma classification, in eng, *J. Thorac. Oncol.* 6 (September 9) (2011) 1496–1504, <https://doi.org/10.1097/JTO.0b013e318221f701>.
- [42] S.M. Yang, et al., Extraction of radiomic values from lung adenocarcinoma with near-pure subtypes in the International Association for the Study of Lung Cancer/the American Thoracic Society/the European Respiratory Society (IASLC/ATS/ERS) classification, *Lung Cancer* 119 (2018) 56–63, <https://doi.org/10.1016/j.lungcan.2018.03.004> May.
- [43] E. Thunnissen, et al., Reproducibility of histopathological subtypes and invasion in pulmonary adenocarcinoma. An international interobserver study, *Mod. Pathol.* 25 (December 12) (2012) 1574–1583, <https://doi.org/10.1038/modpathol.2012.106>.
- [44] H.E. Trejo Bittar, P. Incharoen, A.D. Althouse, S. Dacic, Accuracy of the IASLC/ATS/ERS histological subtyping of stage I lung adenocarcinoma on intraoperative frozen sections, *Mod. Pathol.* 28 (August 8) (2015) 1058–1063, <https://doi.org/10.1038/modpathol.2015.71>.
- [45] A. Warth, et al., Interobserver variability in the application of the novel IASLC/ATS/ERS classification for pulmonary adenocarcinomas, *Eur. Respir. J.* 40 (November 5) (2012) 1221–1227, <https://doi.org/10.1183/09031936.00219211>.
- [46] S.H. Song, et al., Imaging phenotyping using radiomics to predict micropapillary pattern within lung adenocarcinoma, in eng, *J. Thorac. Oncol.* 12 (April 4) (2017) 624–632, <https://doi.org/10.1016/j.jtho.2016.11.2230>.