Relationship between pulmonary nodule malignancy and surrounding pleurae, airways and vessels: a quantitative study using the public LIDC-IDRI dataset

Yulei Qin\textsuperscript{a,b,c}, Yun Gu\textsuperscript{a,b}, Hanxiao Zhang\textsuperscript{b}, Jie Yang\textsuperscript{a,b,*}, Lihui Wang\textsuperscript{d}, Zhexin Wang\textsuperscript{e}, Feng Yao\textsuperscript{e,b}, Yue-Min Zhu\textsuperscript{c}

\textsuperscript{a}Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai 200240, China
\textsuperscript{b}Institute of Medical Robotics, Shanghai Jiao Tong University, Shanghai 200240, China
\textsuperscript{c}CREATIS, INSA Lyon, CNRS UMR 5220, INSERM U1206, Université de Lyon, Villeurbanne 69621, France
\textsuperscript{d}Key Laboratory of Intelligent Medical Image Analysis and Precise Diagnosis of Guizhou Province, School of Computer Science and Technology, Guizhou University, Guiyang 550025, China
\textsuperscript{e}Department of Thoracic Surgery, Shanghai Chest Hospital, Shanghai Jiao Tong University, Shanghai 200025, China

Abstract

Objectives: To investigate whether the pleurae, airways and vessels surrounding a nodule on non-contrast computed tomography (CT) can discriminate benign and malignant pulmonary nodules.

Materials and Methods: The LIDC-IDRI dataset, one of the largest publicly available CT database, was exploited for study. A total of 1556 nodules from 694 patients were involved in statistical analysis, where nodules with average scorings \(<3\) and \(>3\) were respectively denoted as benign and malignant. Besides, 339 nodules from 113 patients with diagnosis ground-truth were independently evaluated. Computer algorithms were developed to segment pulmonary structures and quantify the distances to pleural surface, airways and...

Abbreviations: CT: Computed tomography; LIDC: Lung image database consortium; IDRI: Image database resource initiative; XML: Extensible markup language; 2-D: two-dimensional; 3-D: three-dimensional.

*Corresponding author

Email addresses: qinyulei@sjtu.edu.cn (Yulei Qin), geron762@sjtu.edu.cn (Yun Gu), hanxiao.zhang@sjtu.edu.cn (Hanxiao Zhang), jieyang@sjtu.edu.cn (Jie Yang), wlh1984@gmail.com (Lihui Wang), wangzhexin001@hotmail.com (Zhexin Wang), yaofeng67966780126.com (Feng Yao), zhu@creatis.insa-lyon.fr (Yue-Min Zhu)
vessels, as well as the counting number and normalized volume of airways and vessels near a nodule. Odds ratio (OR) and Chi-square ($\chi^2$) testing were performed to assess the correlation between features of surrounding structures and nodule malignancy. A non-parametric receiver operating characteristic (ROC) analysis was conducted in logistic regression to evaluate discrimination ability of each structure.

**Results:** For the benign and malignant groups, the average distances from nodules to pleural surface, airways and vessels are respectively (6.56, 5.19), (37.08, 26.43) and (1.42, 1.07) mm. The correlation between nodules and the counting number of airways and vessels that contact or project towards nodules are respectively (OR=22.96, $\chi^2=105.04$) and (OR=7.06, $\chi^2=290.11$). The correlation between nodules and the volume of airways and vessels are (OR=9.19, $\chi^2=159.02$) and (OR=2.29, $\chi^2=55.89$). The areas-under-curves (AUCs) for pleurae, airways and vessels are respectively 0.5202, 0.6943 and 0.6529.

**Conclusion:** Our results show that malignant nodules are often surrounded by more pulmonary structures compared with benign ones, suggesting that features of these structures could be viewed as lung cancer biomarkers.

**Keywords:** Pulmonary nodule, Airway, Vessel, Pleura, Chest CT, Computer-assisted analysis

1. **Introduction**

The latest 2020 Global Cancer Statistics demonstrate that lung cancer remains the leading cause of cancer death worldwide [1]. The outcomes of lung cancer are highly dependent on the stage. Although the 5-year survival rate is 6% for patients with metastatic disease, if early diagnosis and treatment are made, the survival rate could be greatly increased to 60% [2, 3]. The National Lung Screening Trial shows that annual screening with computed tomography (CT) brings about a 20% reduction in lung cancer mortality [4].

Pulmonary nodule appears as a white spot inside lung on chest CT [5]. Its likelihood of malignancy indicates lung cancer. To reduce the burden of
radiologists in reading and assessing nodules on a slice-by-slice basis, many researchers have proposed computer algorithms for nodule malignancy classification [6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23]. Most of these methods relied on handcrafted features such as nodule intensity, sphericity and texture descriptors. Based on extracted features, classifiers like support vector machine and logistic regression were employed for malignancy estimation. Recently, deep learning methods have prevailed in the nodule classification task, where features are learned automatically from CT images. Such superior performance comes with a side effect of relatively low interpretability.

In the development of automatic classification methods, false positive predictions cause harmful consequences and unnecessary treatments (e.g., follow-up CT scans and invasive biopsies) [24, 25]. To improve performance, one potential solution is to incorporate clinically relevant context information as much as possible into nodule malignancy estimation. Most current methods only refer to the size, growth rate and morphologic characteristics of nodules since these attributes have been proved effective in diagnosis [26]. However, the context of a nodule, namely the relationship between a nodule and its surrounding structures, might provide supplementary diagnostic basis and should not be neglected. The exploration and understanding of such relationship is of high priority for both medical and computer vision researchers.

1.1. Related work

Relationship between a nodule and pleurae. Kim et al. [27] evaluated whether the attachment of a nodule to pleura related to visceral pleural invasion (VPI). The increase of solid portion in part-solid nodules suggested high risk of VPI. Heidinger et al. [28] showed that solid nodules that contacted pleural surface were associated with higher likelihood of VPI than part-solid nodules. Zhu et al. [29] performed a retrospective review and found that all non-calcified solid nodules attached to the costal pleura were benign if they had smooth margins, oval, semi-circular or triangular shapes, with diameters less than 10 mm.
**Relationship between a nodule and airways.** Gaeta et al. [30] investigated the value of bronchus sign in predicting the success of transbronchial biopsy and brushing for patients with peripheral lung lesions. They later studied how the type of tumor-bronchus relationship will determine the yield of transbronchial biopsy [31]. Qiang et al. [32] used multi-slice CT to study five different tumor-bronchus relationships. The categorization of such relationship reflected the pathological changes of nodules to some extent. Cui et al. [33] defined four types of nodule-bronchus relationship by bronchial morphology and confirmed their roles in determining the degree of differentiation of solid pulmonary nodules.

**Relationship between a nodule and vessels.** Mori et al. [6] studied tumor angiogenesis and reported that the involvement of veins converging to a nodule was strongly suggestive of malignancy. Kawata et al. [34] used geometry-based vector fields for quantification instead of explicitly segmenting vascular structures around nodules. Wang et al. [35] demonstrated that the morphology subtypes of bronchi, pulmonary arteries and veins correlated with size, location, pathology and stage of peripheral lung cancer. Wang et al. [36] validated that the vessels surrounding a nodule could help discriminate between benign and malignant nodules. The position and orientation of vessels relative to a nodule may imply lung cancer stage and pathology [37, 38, 39].

### 1.2. Challenges

There exist limitations in previous studies on the relationship between the malignancy of a nodule and its surrounding airways, vessels and pleurae. First, the size and diversity of CT database are limited in early work. Due to the difficulty of data collection and privacy issues, the number of patient cases is usually less than 100 and multi-center study was not available. Second, the extraction of pulmonary structures is a challenge without the help of image segmentation algorithms. It is tedious and time-consuming for radiologists to manually delineate and measure airways, vessels and pleurae from CT, which in turn imposes restrictions on the size of database. Third, the relationship
between a nodule and other structures is often categorized into subtypes by descriptive definition. Very few studies computed features to quantify such relationship. Without quantitative measurement, it is difficult to discover the value of surrounding structures in assessing nodule malignancy. Last but not least, the correlation analyses on categorical or dichotomized variables may not be adequate to reveal the discrimination ability of airways, vessels and pleurae. Classifiers are needed to evaluate the accuracy of utilizing such relationship for benign-malignant classification.

1.3. Contributions

In this study, we investigated if a pulmonary nodule’s surrounding structures like pleurae, airways and vessels (arteries and veins) could discriminate between benign and malignant nodules. Deep learning methods were developed to segment lung field (pleural surface), tubular airways and vessels (arteries and veins) from 3-D CT scans. Given the position and diameter of each nodule, a local volume-of-interest (VOI) was extracted to quantify features of pulmonary structures inside. Then, statistical correlation analyses and logistic regression experiments were performed to evaluate whether the presence of certain surrounding structures will be more frequent with malignant nodules than with benign ones. The database involved is one of the largest publicly available LIDC-IDRI lung nodule dataset [40]. According to our study on the database demographics and annotations, extensive experiments were designed and conducted. It is believed that investigation of such relationship on a large-scale database, by means of quantitative analysis, is indispensable and enlightening for improving the accuracy of nodule diagnosis.

2. Materials and Methods

2.1. Data sources

The LIDC-IDRI database [40] is one of the largest publicly available lung nodule dataset, consisting of 1018 CT scans collected retrospectively from seven
medical centers. These images were acquired by a wide range of scanner manufacturers and reconstruction parameters [41], with slice thickness ranging from 0.6 to 5.0 mm. The size of axial slices of all CT scans is 512×512.

2.2. Annotations and diagnosis results

A two-phase annotation process was adopted. In the first phase, each radiologist independently marked suspicious lesions on CT as non-nodule, nodule (diameter < 3 mm) and nodule (diameter ≥ 3 mm). For nodules (diameter ≥ 3 mm), their boundaries were outlined and nine attributes (subtlety, internal structure, calcification, sphericity, margin, lobulation, spiculation, texture and malignancy) were subjectively estimated as scorings. For nodules (diameter < 3 mm), only the coordinates of their centroids were provided. In the second phase, each radiologist reviewed annotations with reference to anonymized annotations from other radiologists. Modifications were allowed but no consensus among observers was enforced. All annotations were encoded in XML files.

We also selected a subset of 113 patient cases with diagnosis-definite ground-truth labels from LIDC-IDRI. The patient-level diagnostic results include: 1) benign; 2) malignant, primary lung cancer; 3) malignant metastatic. Diagnosis methods involve: 1) review of radiological images to show 2 years of stable nodule; 2) biopsy; 3) surgical resection; 4) progression or response.

2.3. Data processing and experiment design

Statistical analysis and logistic regression experiments were respectively designed with different ways of dataset partition and processing.

Statistical analysis. Following the procedures in previous studies [42, 43, 44, 45, 21, 22], all nodules (diameter ≥ 3 mm) with at least one annotation were extracted from XML files and isometric resampling of 1.0 × 1.0 × 1.0 mm³ was performed on CT scans to tackle a large variety of spatial resolution. The malignancy of each nodule was scored from 1 (highly unlikely malignant) to 5 (highly suspicious) and scorings from different radiologists were averaged. Nodules with mean malignancy score < 3, = 3 and > 3 are respectively benign, uncertain and
malignant. In total, there are 2505 nodules (1044 benign, 512 malignant and 949 uncertain). Note that the number of nodules extracted from the LIDC-IDRI database may vary from study to study (e.g., 1356 [42], 891 [43], 1145 [44], 2618 [45], 2272 [46], 2557 [21], 2669 [40]). The reasons behind are threefold: 1) the database was released in 2011 and has been updated in 2012, 2015, 2018 and 2020; 2) different software packages were developed for analyzing XML files; 3) various strategies of label fusion were adopted without establishing any single criterion. All these interpretations of data formatting were possible and justified [47, 48]. To reduce the impact of nodules with uncertain malignancy, we excluded these uncertain nodules in experiments. All the remaining 1556 nodules from 694 patients were involved in statistical analysis.

Although the malignancy scorings are subjective, they do statistically reflect certain characteristics of nodules with malignant tendency. Radiologists comprehensively considered multiple nodule attributes to assess malignancy. It is reasonable to refer to such scorings as proxy ground-truth labels. Besides, to quantify inter-observer variability, we followed Ref. [49, 50] to calculate the mean rating difference of each nodule between all pairs of radiologists. For example, given scorings \{1,2,4\} from three observers, the paired differences are \{|1-2|, |1-4|, |2-4|\} and the mean value is 2. The distribution of nodules with proxy malignancy label is summarized in Table 1.

**Logistic regression.** To verify if the surrounding structures can predict the malignancy of a nodule, a logistic regression classifier [51] was developed for receiver-operating-characteristic (ROC) analysis. We used 113 patient cases with diagnosis results as the testing set. Patients with diagnosis category=1 are benign and those with category=2 or 3 are malignant. Note that diagnosis was performed at patient-level as a multiple-instance learning (MIL) task [40]. For a patient with multiple nodule instances, he/she is malignant if at least one instance is malignant. Only if all instances are benign, the patient is benign. During training, we used 914 benign nodules and 429 malignant nodules. The classifier learned to predict the probability of each nodule being malignant.
Table 1: Subject demographics in statistical analysis experiments

| Nodule                          | Benign scoring < 3 | Malignant scoring > 3 | Uncertain scoring = 3 | All     |
|---------------------------------|--------------------|-----------------------|-----------------------|---------|
| No.                             | 1044               | 512                   | 949                   | 2505    |
| Equivalent diameter (Eqd.) *    |                    |                       |                       |         |
| Avg. eqd. †                     | 6.97 ± 2.93        | 14.52 ± 6.47          | 7.50 ± 3.20           | 8.64 ± 5.00 |
| No. eqd. ≤ 10 mm                | 965                | 143                   | 808                   | 1916    |
| No. 10 < eqd. ≤ 20 mm           | 71                 | 265                   | 130                   | 466     |
| No. eqd. > 20 mm                | 8                  | 104                   | 11                    | 123     |
| Texture †                       |                    |                       |                       |         |
| Solid                           | 766                | 324                   | 568                   | 1658    |
| Part-solid                      | 278                | 188                   | 381                   | 847     |
| Scorings                        |                    |                       |                       |         |
| No. radiologists = 1            | 316                | 66                    | 284                   | 666     |
| No. radiologists > 1            | 728                | 446                   | 665                   | 1839    |
| Avg. difference †               | 0.78 ± 0.63        | 1.05 ± 0.54           | 0.77 ± 0.59           | 0.84 ± 0.61 |
| Background lung diseases        |                    |                       |                       |         |
| Emphysema                       | 93                 | 40                    | 79                    | 212     |
| Fibrosis                        | 32                 | 19                    | 27                    | 78      |
| Pulmonary congestion            | 103                | 83                    | 92                    | 278     |

* Given the volume of a nodule, equivalent diameter (Eq. diameter, Eqd.) is calculated by assuming the nodule as an ideal sphere. Compared with nodule’s maximum diameter, the equivalent diameter is a normalized metric and reflects more closely its actual size.

† Statistical significance (p-value < 0.05) was observed between different nodule groups.
Table 2: Subject demographics in logistic regression experiments

| Nodule                  | Benign patient-level | Malignant patient-level | All   |
|-------------------------|----------------------|-------------------------|-------|
| No.                     | 81                   | 258                     | 339   |
| Equivalent diameter (Eqd.) |                       |                         |       |
| Avg. eqd.†              | 7.93±3.83            | 10.57±5.90              | 9.94±5.59 |
| No. eqd.≤10 mm          | 67                   | 160                     | 227   |
| No. 10<eqd.≤20 mm       | 12                   | 74                      | 86    |
| No. eqd.>20 mm          | 2                    | 24                      | 26    |
| Texture                 |                       |                         |       |
| Solid                   | 69                   | 208                     | 277   |
| Part-solid              | 12                   | 50                      | 62    |
| Scorings                |                       |                         |       |
| No. radiologists =1     | 23                   | 48                      | 71    |
| No. radiologists >1     | 58                   | 210                     | 268   |
| Avg. difference†        | 0.75±0.60            | 0.90±0.58               | 0.87±0.59 |
| Background lung diseases|                       |                         |       |
| Emphysema               | 8                    | 22                      | 30    |
| Fibrosis                | 2                    | 12                      | 14    |
| Pulmonary congestion    | 6                    | 46                      | 52    |

† Statistical significance (p-value<0.05) was observed between different nodule groups.
During testing, for each patient, the maximum of malignancy probabilities of all visible nodules was deemed as patient-level malignancy probability. The training set (1343 nodules) does not overlap with testing set (339 nodules). The distribution of nodules with patient-level ground-truth is summarized in Table 2.

2.4. Quantification of surrounding structures

![Image: Extraction of lung fields, airways and vessels.](image)

The lung fields, airways and vessels (arteries and veins) were segmented on CT scans using deep learning methods [52, 53]. The centerlines of airway and vessel branches were extracted via skeletonization [54]. The coordinates and masks of nodules were obtained from LIDC-IDRI annotations. For each nodule,
a spherical sub-volume located at the nodule centroid was cropped as VOI. Its radius is the nodule diameter and has to be at least 6 mm. Fig. 1 illustrates the steps of extracting pulmonary structures. To describe the proximity between a nodule and its surrounding structures, we measured the distance from the nodule surface to its closest target structure.

To quantify the surrounding airways and vessels, we had two choices: 1) all structures inside VOI are evaluated; 2) only structures that are connected to or project towards a nodule are evaluated. For the second choice, we referred to Ref. 36 for the definition of filtering rules in distance and orientation:

i. the centerline of airway or vessel is directly attached to a nodule;

ii. the distance between the centerline and the nodule surface is $\leq 3$ mm;

iii. the distance between the centerline and the nodule surface is $\leq 5$ mm and the centerline is projecting towards the nodule centroid.

Here, “projecting towards” means the angle, between centerline trajectory and the segment between a centerline end and the nodule centroid, is $\leq 15$ degrees. All conditions above were summarized from clinical practice, but they might neglect some pertinent pulmonary structures. Therefore, both the two choices were adopted for quantification and comparison. Fig. 2 illustrates the differences between choice 1 and choice 2 in identifying vessels.

The quantification of airways and vessels also includes counting their number and volume inside VOI. Different from Ref. 36, we divided the volume of structures by the volume of non-nodule region in VOI. Such normalized volume avoids the influence of nodule size. By nature, the VOI of a large nodule contains more airways and vessels. If the absolute volume is used, the volume increase of structures may be simply due to the enlarged VOI. Only through normalization can we properly validate the correlation between volume of surrounding structures and nodule malignancy.

2.5. Data analysis

Data analysis was performed using Python with Scipy library 55.
Figure 2: Differences between choice 1 and 2 in identifying vessels inside VOI. (a) In choice 1, all branches were counted in quantification. In choice 2, only V1, V2 and V3 were quantified since they respectively matched the condition ii, iii and i. (b) Vessels and skeletons before and after filtering. Vessel features include the distance, counting number and normalized volume.
Statistical analysis. The student’s t-tests were conducted for comparisons of quantified structures between benign and malignant nodule groups. Correlation between features of pleurae, airways and vessels was tested with Pearson correlation coefficient (PCC), where two-tailed p-value was estimated. To interpret the relationship between pulmonary structures and nodules, features of surrounding structures were dichotomized with respect to nodule categories. Odds ratio (OR) and Chi-square tests ($\chi^2$) were performed on the resulting contingency tables.

Logistic regression. For each structure, a classifier was trained using quantified features. The model was evaluated on nodules from the testing set. ROC analysis was performed on the patient-level malignancy prediction results.

3. Results

Both Table 1 and Table 2 showed that there existed significant differences between benign and malignant nodules in equivalent diameter and scoring disagreement among radiologists. The average diameters of benign and malignant nodules were respectively 6.97 mm and 14.52 mm. Malignant nodules tended to grow larger and lack expert consensus. Additionally, there were not significant differences between benign and malignant groups in diagnosed lung diseases such as emphysema and fibrosis. Pulmonary congestion was relatively frequently observed in patients with malignant nodules.

Table 3 reflected that the average distance from nodule to pleural surface, airways, vessels, arteries and veins was significantly smaller in the malignant group. The counting number of all surrounding structures was different between benign and malignant groups. Similar results were observed in the normalized volume except for arteries and veins (choice 2). More structures were present around malignant nodules.

The distance from nodule to pleurae, airways and vessels was negatively associated with nodule diameter, but the correlation was weak (see Table 4). For vessels, arteries and veins around benign nodules, the correlation between
Table 3: Comparison of features of pleurae, airways and vessels surrounding the nodule

| Features          | Benign     | Malignant  | All        |
|-------------------|------------|------------|------------|
| **Distance to nodule (mm)** |            |            |            |
| pleural surface†  | 6.56±6.93  | 5.19±6.09  | 6.11±6.69  |
| airways†          | 37.08±19.29| 26.43±21.28| 33.57±20.59|
| vessels†          | 1.42±1.39  | 1.07±0.48  | 1.31±1.18  |
| arteries†         | 2.53±2.14  | 1.32±0.88  | 2.13±1.91  |
| veins†            | 2.25±2.42  | 1.32±0.94  | 1.94±2.10  |
| **Counting number** |            |            |            |
| airways (choice 1)† | 0.05±0.27  | 0.57±1.08  | 0.22±0.70  |
| airways (choice 2)† | 0.05±0.25  | 0.49±0.90  | 0.19±0.59  |
| vessels (choice 1)† | 6.63±9.79  | 34.94±39.09| 15.94±27.28|
| vessels (choice 2)† | 2.42±1.95  | 5.74±4.68  | 3.52±3.49  |
| arteries (choice 1)† | 5.02±10.17 | 31.95±38.25| 13.88±26.66|
| arteries (choice 2)† | 2.47±4.41  | 5.90±9.00  | 3.60±6.50  |
| veins (choice 1)† | 4.07±7.88  | 26.36±31.98| 11.40±22.09|
| veins (choice 2)† | 1.93±3.27  | 4.64±4.89  | 2.82±4.08  |
| **Normalized volume (%)** |            |            |            |
| airways (choice 1)† | 0.09±0.72  | 0.29±0.88  | 0.16±0.78  |
| airways (choice 2)† | 0.09±0.72  | 0.27±0.84  | 0.15±0.77  |
| vessels (choice 1)† | 3.78±3.37  | 5.35±3.10  | 4.29±3.37  |
| vessels (choice 2)† | 2.75±3.15  | 3.62±2.93  | 3.03±3.11  |
| arteries (choice 1)† | 1.97±2.51  | 2.88±1.84  | 2.27±2.35  |
| arteries (choice 2) | 1.56±2.34  | 1.65±1.59  | 1.59±2.12  |
| veins (choice 1)† | 1.77±2.01  | 2.39±1.72  | 1.98±1.94  |
| veins (choice 2)† | 1.33±1.87  | 1.34±1.55  | 1.34±1.77  |

† Statistical significance (p-value<0.05) was observed between different nodule groups.
Table 4: Correlation among features of surrounding pulmonary structures and nodules

| Features                                      | PCC (r)     |
|-----------------------------------------------|-------------|
|                                               | Benign      | Malignant | All      |
| Pleurae                                       |             |           |          |
| Distance & Eq. diameter                       | -0.08\(\dagger\) | -0.29\(\dagger\) | -0.15\(\dagger\) |
| Airways                                       |             |           |          |
| Distance & Eq. diameter                       | -0.14\(\dagger\) | -0.23\(\dagger\) | -0.25\(\dagger\) |
| Counting number & Eq. diameter (choice 1)    | 0.34\(\dagger\) | 0.41\(\dagger\) | 0.50\(\dagger\) |
| Counting number & Eq. diameter (choice 2)    | 0.14\(\dagger\) | 0.42\(\dagger\) | 0.51\(\dagger\) |
| Normalized volume & Eq. diameter (choice 1)  | 0.14\(\dagger\) | 0.14\(\dagger\) | 0.16\(\dagger\) |
| Normalized volume & Eq. diameter (choice 2)  | 0.16\(\dagger\) | 0.14\(\dagger\) | 0.16\(\dagger\) |
| Counting number & normalized volume (choice 1)| 0.52\(\dagger\) | 0.46\(\dagger\) | 0.49\(\dagger\) |
| Counting number & normalized volume (choice 2)| 0.53\(\dagger\) | 0.43\(\dagger\) | 0.39\(\dagger\) |
| Vessels                                       |             |           |          |
| Distance & Eq. diameter                       | -0.10\(\dagger\) | -0.15\(\dagger\) | -0.14\(\dagger\) |
| Counting number & Eq. diameter (choice 1)    | 0.71\(\dagger\) | 0.63\(\dagger\) | 0.73\(\dagger\) |
| Counting number & Eq. diameter (choice 2)    | 0.29\(\dagger\) | 0.51\(\dagger\) | 0.56\(\dagger\) |
| Normalized volume & Eq. diameter (choice 1)  | 0.09\(\dagger\) | -0.08     | 0.13\(\dagger\) |
| Normalized volume & Eq. diameter (choice 2)  | 0.05        | -0.03     | 0.07\(\dagger\) |
| Counting number & normalized volume (choice 1)| 0.15\(\dagger\) | 0.04      | 0.15\(\dagger\) |
| Counting number & normalized volume (choice 2)| 0.14\(\dagger\) | 0.06      | 0.13\(\dagger\) |
| Arteries                                      |             |           |          |
| Distance & Eq. diameter                       | -0.21\(\dagger\) | -0.31\(\dagger\) | -0.28\(\dagger\) |
| Counting number & Eq. diameter (choice 1)    | 0.71\(\dagger\) | 0.64\(\dagger\) | 0.73\(\dagger\) |
| Counting number & Eq. diameter (choice 2)    | 0.30\(\dagger\) | 0.23\(\dagger\) | 0.34\(\dagger\) |
| Normalized volume & Eq. diameter (choice 1)  | 0.08\(\dagger\) | -0.06     | 0.12\(\dagger\) |
| Normalized volume & Eq. diameter (choice 2)  | 0.01        | -0.17\(\dagger\) | -0.02 |
| Counting number & normalized volume (choice 1)| 0.13\(\dagger\) | 0.06      | 0.14\(\dagger\) |
| Counting number & normalized volume (choice 2)| 0.09\(\dagger\) | 0.09\(\dagger\) | 0.08\(\dagger\) |
| Veins                                         |             |           |          |
| Distance & Eq. diameter                       | -0.11\(\dagger\) | -0.27\(\dagger\) | -0.19\(\dagger\) |
| Counting number & Eq. diameter (choice 1)    | 0.74\(\dagger\) | 0.64\(\dagger\) | 0.73\(\dagger\) |
| Counting number & Eq. diameter (choice 2)    | 0.53\(\dagger\) | 0.35\(\dagger\) | 0.48\(\dagger\) |
| Normalized volume & Eq. diameter (choice 1)  | 0.04        | -0.09\(\dagger\) | 0.06\(\dagger\) |
| Normalized volume & Eq. diameter (choice 2)  | -0.02       | -0.21\(\dagger\) | -0.09\(\dagger\) |
| Counting number & normalized volume (choice 1)| 0.14\(\dagger\) | 0.04      | 0.12\(\dagger\) |
| Counting number & normalized volume (choice 2)| 0.12\(\dagger\) | 0.03      | 0.09\(\dagger\) |

\(\dagger\) Low probability (p-value <0.05) was observed if the two feature variables were uncorrelated.
Table 5: Relationship between dichotomized distances to structures and nodules

| Distance to nodule (mm) | Malignancy | Eq. diameter | Texture |
|------------------------|------------|--------------|---------|
|                        | Benign     | Malignant    | <10 mm  | ≥10 mm  | Solid | Part-solid |
| Pleurae                | ≤1         | 407          | 253     | 400     | 260   | 475        | 185    |
|                        | >1         | 637          | 259     | 708     | 188   | 615        | 281    |
| Airways                | ≤1         | 12           | 50      | 14      | 48    | 47         | 15     |
|                        | >1         | 1032         | 462     | 1094    | 400   | 1043       | 451    |
| Vessels                | ≤1         | 1250         | 14      | 158     | 4     | 124        | 38     |
|                        | >1         | 1407         | 402     | 1094    | 400   | 1043       | 451    |
| Arteries               | ≤1         | 896          | 498     | 950     | 444   | 966        | 428    |
|                        | >1         | 1407         | 498     | 950     | 444   | 966        | 428    |
| Veins                  | ≤1         | 540          | 433     | 572     | 401   | 633        | 340    |
|                        | >1         | 504          | 79      | 536     | 47    | 457        | 126    |
|                        | ≤1         | 363          | 80      | 397     | 46    | 310        | 133    |
|                        | >1         | 681          | 432     | 711     | 402   | 780        | 333    |

counting number and normalized volume was weak. Such correlation was weaker for malignant nodules. The counting number of all surrounding structures was strongly correlated with diameter.

Table 8 provided the correlation testing results of Table 5. The attachment to pleurae, airways, vessels and veins was hardly relevant to nodule texture. The attachment to arteries was more often observed in part-solid nodules. Nodules contacting pleural surface and tubular structures were less likely to be benign and small.

Table 9 provided the correlation testing results of Table 6. The counting number of all structures except vessels (choice 2) was irrelevant to nodule texture. The more tubule branches exist inside VOI, the more likely the nodule is malignant.

Table 10 provided the correlation testing results of Table 7. The normalized volume of all structures except artery (choice 1) was irrelevant to nodule texture. Except for arteries and veins (choice 2), the volume of all surrounding structures was positively related to nodule malignancy and diameter. The larger percentage of VOI these structures occupy, the more likely the nodule is malignant.
| Counting number | Malignancy | Eq. diameter | Texture |          |          |
|-----------------|------------|--------------|---------|----------|----------|
| Airways         | Benign     | Malignant    | <10 mm  | >10 mm   | Solid    |
| ≤1              | 1038       | 436          | 1104    | 370      | 1035     |
|                 | ≥10 mm     |              |         |          | 439      |
| Airways         | ≤1         | 1038         | 452     |          | 1048     |
| (choice 1)      | >1         | 6            | 76      | 4        | 55       |
| Airways         | ≤1         | 1038         | 452     |          | 1048     |
| (choice 2)      | >1         | 6            | 60      | 3        | 42       |
| Airways         | ≤10        | 907          | 108     |          | 733      |
| (choice 1)      | >10        | 137          | 404     |          | 357      |
| Airways         | ≤3         | 850          | 196     |          | 748      |
| (choice 2)      | >3         | 194          | 316     |          | 342      |
| Vessels         | ≤10        | 964          | 144     |          | 788      |
| (choice 1)      | >10        | 80           | 368     |          | 302      |
| Arteries        | ≤10        | 991          | 174     |          | 827      |
| (choice 1)      | >10        | 53           | 338     |          | 263      |
| Arteries        | ≤3         | 916          | 280     |          | 849      |
| (choice 2)      | >3         | 128          | 232     |          | 241      |
| Veins           | ≤10        | 991          | 174     |          | 827      |
| (choice 1)      | >10        | 53           | 338     |          | 263      |
| Veins           | ≤3         | 916          | 280     |          | 849      |
| (choice 2)      | >3         | 128          | 232     |          | 241      |
| Normalized volume (%) | Malignancy | Eq. diameter | Texture |
|-----------------------|------------|--------------|---------|
|                       | Benign     | Malignant    | <10 mm  | ≥10 mm |
| Airways (choice 1)    | ≤0.1       | 1010         | 391     | 1074   | 327     | 978     | 423     |
| Airways (choice 2)    | >0.1       | 34           | 121     | 34     | 121     | 112     | 43      |
| Airways (choice 1)    | ≤0.1       | 1010         | 391     | 1074   | 327     | 978     | 423     |
| Airways (choice 2)    | >0.1       | 34           | 121     | 34     | 121     | 112     | 43      |
| Vessels (choice 1)    | ≤2         | 365          | 44      | 379    | 30      | 296     | 113     |
| Vessels (choice 2)    | >2         | 679          | 468     | 729    | 418     | 794     | 353     |
| Arteries (choice 1)   | ≤2         | 704          | 181     | 721    | 164     | 639     | 246     |
| Arteries (choice 2)   | >2         | 340          | 331     | 387    | 284     | 451     | 220     |
| Veins (choice 1)      | ≤2         | 689          | 254     | 715    | 228     | 654     | 289     |
| Veins (choice 2)      | >2         | 355          | 258     | 393    | 220     | 436     | 177     |
| Veins (choice 1)      | ≤2         | 810          | 408     | 846    | 372     | 854     | 364     |
| Veins (choice 2)      | >2         | 234          | 104     | 262    | 76      | 236     | 102     |
Table 8: Correlation testing between dichotomized distance and nodule attribute

| Distance to nodule (mm) | Nodule attribute | OR  | $\chi^2$ |
|------------------------|------------------|-----|---------|
|                        | Malignancy       | 0.65†| 15.30† |
| Pleurae                | Eq. diameter     | 0.41†| 62.84† |
|                        | Texture          | 1.17 | 2.01   |
|                        | Malignancy       | 0.11†| 66.66† |
| Airways                | Eq. diameter     | 0.11†| 74.48† |
|                        | Texture          | 1.35 | 1.02   |
|                        | Malignancy       | 0.17†| 48.22† |
| Vessels                | Eq. diameter     | 0.05†| 61.11† |
|                        | Texture          | 0.69 | 3.63   |
|                        | Malignancy       | 0.20†| 158.19†|
| Arteries               | Eq. diameter     | 0.13†| 195.42†|
|                        | Texture          | 0.51†| 30.88† |
|                        | Malignancy       | 0.35†| 61.83† |
| Veins                  | Eq. diameter     | 0.20†| 102.36†|
|                        | Texture          | 1.00 | 0.00   |

† Statistical significance (p-value<0.05) was observed between different nodule groups.
Table 9: Correlation testing between dichotomized counting number and nodule attribute

| Counting Nodule | Choice 1          | Choice 2          | \( \chi^2 \)  | \( \chi^2 \)  |
|----------------|-------------------|-------------------|----------------|----------------|
| Airways        | Malignancy        | 30.16†            | 140.11†        | 22.96†         | 105.04†        |
|                | Eq. diameter      | 58.18†            | 185.76†        | 60.27†         | 149.39†        |
|                | Texture           | 1.16              | 0.37           | 1.35           | 1.35           |
| Vessels        | Malignancy        | 24.77†            | 655.47†        | 7.06†          | 290.11†        |
|                | Eq. diameter      | 55.39†            | 817.72†        | 10.54†         | 388.09†        |
|                | Texture           | 1.34†             | 6.52†          | 1.23           | 3.24           |
| Arteries       | Malignancy        | 30.79†            | 690.87†        | 4.94†          | 196.73†        |
|                | Eq. diameter      | 52.83†            | 844.45†        | 5.12†          | 197.83†        |
|                | Texture           | 1.19              | 2.09           | 1.07           | 0.32           |
| Veins          | Malignancy        | 36.32†            | 678.06†        | 5.93†          | 211.03†        |
|                | Eq. diameter      | 72.14†            | 884.64†        | 8.08†          | 285.87†        |
|                | Texture           | 1.19†             | 1.93†          | 1.21           | 2.16           |

† Statistical significance (p-value<0.05) was observed between different nodule groups.
| Normalized volume (%) | Nodule attribute | Choice 1 | Choice 2 |
|-----------------------|-----------------|---------|---------|
|                       | Malignancy      | OR      | OR      |
| Airways               | 9.19†           | 159.02† | 9.19†   | 159.02† |
|                       | Eq. diameter    | 11.69†  | 203.85† | 11.69†   | 203.85† |
|                       | Texture         | 0.89    | 0.40    | 0.89    | 0.40    |
|                       | Malignancy      | 5.72†   | 123.27† | 2.29†   | 55.89†  |
| Vessels               | 7.24†           | 124.60† | 2.20†   | 46.65†  |
|                       | Eq. diameter    | 1.16    | 1.42    | 0.98    | 0.03    |
|                       | Texture         | 3.79†   | 144.15† | 1.39†   | 7.61†   |
| Arteries              | 7.24†           | 124.60† | 1.00    | 0.00    |
|                       | Eq. diameter    | 1.27†   | 4.53†   | 1.10    | 0.63    |
|                       | Texture         | 1.97†   | 38.64†  | 0.88    | 0.89    |
| Veins                 | 1.76†           | 24.85†  | 0.66†   | 8.38†   |
|                       | Eq. diameter    | 0.92    | 0.56    | 1.01    | 0.01    |

† Statistical significance (p-value<0.05) was observed between different nodule groups.

malignant and large.

Performance and ROC analysis of logistic regression were reported in Table 11 and Fig. 3. The pleural feature only involved the distance to nodule. For other structures, the distance, counting number and normalized volume were combined as features. The artery features (choice 2) achieved the highest accuracy, recall and $F_1$-score while the vein features (choice 1) gained the highest areas-under-curve (AUC). The highest precision was obtained by airway features of both choices 1 and 2.

4. Discussion

Our study suggests that the distance, counting number and normalized volume of airways, vessels, arteries and veins can be viewed as important biomark-
Table 11: Performance of logistic regression for patient-level malignancy prediction

| Structure features      | Accuracy | Precision | Recall | F1-score | AUC   |
|-------------------------|----------|-----------|--------|----------|-------|
| Pleurae                 | 0.7080   | 0.7708    | 0.8706 | 0.8177   | 0.5202|
| Airways (choice 1)      | 0.5841   | **0.9524**| 0.4706 | 0.6299   | 0.6968|
| Airways (choice 2)      | 0.5841   | **0.9524**| 0.4706 | 0.6299   | 0.6943|
| Vessels (choice 1)      | 0.6814   | 0.8451    | 0.7059 | 0.7692   | 0.7336|
| Vessels (choice 2)      | 0.6903   | 0.8125    | 0.7647 | 0.7879   | 0.6529|
| Arteries (choice 1)     | 0.6991   | 0.8228    | 0.7647 | 0.7927   | 0.6975|
| Arteries (choice 2)     | **0.7345**| 0.7895    | **0.8824**| **0.8333**| 0.6424|
| Veins (choice 1)        | 0.6991   | 0.8400    | 0.7412 | 0.7875   | **0.7378**|
| Veins (choice 2)        | 0.7257   | 0.8553    | 0.7647 | 0.8075   | 0.7076|

Figure 3: ROC analysis of logistic regression. Features of surrounding pulmonary structures were used to discriminate benign and malignant nodules and predict patient-level malignancy.
ers for discriminating benign nodules from malignant ones. To the best of our knowledge, this is the first quantitative investigation using a large-scale publicly available chest CT dataset to discover the relation between pulmonary structures and nodule malignancy. It may provide new potential diagnosis basis for indeterminate nodules, and thereafter improve diagnostic accuracy and reduce unnecessary intervention. Our previously proposed segmentation methods make it possible to quantify structures from massive CT samples efficiently and accurately, which is a prerequisite to draw general conclusions.

**Pleurae.** The average distance from pleurae to malignant nodules is smaller than that to benign nodules. Meanwhile, the correlation between distance and nodule size is weak, revealing that: 1) the distance is not necessarily negatively related to nodule size because the position distribution of nodules is diverse; 2) the malignancy is indeed associated with the distance. After dichotomization, the malignant inclination of nodules that directly contact pleurae is clearer. The competitive accuracy versus low AUC suggests that the pleural feature is not robust enough on patient-level malignancy prediction and the binarization threshold needs to be tuned carefully.

**Airways.** Most nodules in LIDC-IDRI dataset are not adjacent to airways. Both the counting number and normalized volume of airways are close to 0. Consequently, correlation between the two features is relatively high. Nearly 90% of the nodules that have more than one airway branch around are malignant. The classifier may simply learn to predict a nodule as malignant provided that airways exist nearby. As a result, the airway feature achieved the highest precision. But for the remaining nodules without airways, the classifier retrieved few malignant ones with a rather low recall.

**Vessels.** Around 97% of the malignant nodules are juxta-vascular and their average distance to vessels is 1.07 mm. Malignant nodules have more vessel branches than benign ones. The correlation between volume and nodule size is very weak, implying that the increase of vessel volume is not merely an outcome
of the expanded VOI. The proportion of malignant nodules contacting vessels or having more than 10 branches in the vicinity is much higher than the opposite. Vessel features exhibited modest power to recognize malignant nodules.

Arteries and Veins. Similar to the findings of vessel features, the distance to nodule, the counting number and normalized volume of arteries and veins are respectively smaller and larger in the malignant group. But the correlation between counting number and nodule size is strong. Nodules surrounded by more arteries and veins tend to be malignant and large. Moreover, either artery or vein features performed well in patient-level malignancy discrimination.

Verification of previous findings. We compared findings in previous studies with ours. In this study, there were 103 benign, 8 malignant, and 99 uncertain non-calcified solid nodules attached to the costal pleura that have smooth margins with diameters less than 10 mm. Such distribution is a bit different from Ref. [29] where all these nodules were benign. Nevertheless, the proportion of benign nodules is still much larger. Consistent with Ref. [30], vessels surrounding a nodule did play a role in benign-malignant distinction. Another interesting finding is that there exists no significant difference between the presence of arteries and veins around malignant nodules, which is in accord with Ref. [39] but against Ref. [6]. Since only 26 patients were investigated by Mori et al. [6], their conclusion might be biased due to limited sample size.

Choice 1 or 2. The difference between choice 1 and 2 lies in the filtering of target structures, but such difference did not lead to contradictory findings. Features of vessels, arteries and veins in choice 2 outperformed those in choice 1 in accuracy, recall and \( F_1 \)-score. The vessels (choice 2) are more closely related to nodule malignancy, corroborating the value of filtering weakly correlated structures.

Limitations. First, some low-resolution CT scans in LIDC-IDRI restrict the extent to which the details of structures are segmented. Errors accumulate along the pipeline of segmentation and quantification. More high-quality scans should be collected. Second, the number of pathologically-proved nodules is
insufficient. Subjective malignancy scorings cannot substitute for pathological ground-truth. On the other hand, percutaneous biopsy is usually carried out on one highly-suspicious malignant nodule (solid, diameter over 15 mm) for patients with multiple nodules \[^{56, 57}\].

Considering that benign-definite and malignant-definite patients under investigation respectively had 2.33 and 2.86 nodules on average, it is laborious or even impractical to obtain labels of all nodules in each patient. Therefore, it takes time to collect more nodule samples with ground-truth. In the future, the effectiveness of structure features as lung cancer biomarkers could be further evaluated. Third, more elaborate features could be considered in quantification, which may improve the discrimination ability of pulmonary structures on nodule malignancy.

5. Conclusion

We investigated the relationship between pleurae, airways and vessels surrounding a nodule and nodule malignancy on a large public chest CT dataset. Correlation analysis on quantified structures demonstrated that the distance from nodule to pleurae, airways and vessels, together with counting number and normalized volume of airways and vessels, can be viewed as potential lung cancer biomarkers. Features of either arteries or veins benefit nodule diagnosis, which could be useful in further studies on lung cancer.

Acknowledgments

This work was partly supported by National Natural Science Foundation of China (Nos. 62003208, 61661010), National Key R&D Program of China (No. 2019YFB1311503), Committee of Science and Technology, Shanghai, China (Nos. 19411963900, 19510711200), Shanghai Sailing Program (No. 20YF1420800), Shanghai Municipal of Science and Technology (Project No. 20JC1419500), International Research Project METISLAB, Program PHC-Cai Yuanpei 2018 (No. 41400TC).
The authors acknowledged the National Cancer Institute and the National Institutes of Health for their critical roles in the creation of the LIDC-IDRI database. The authors also thanked Dr. Yutong Xie (Northwestern Polytechnical University) for the enlightening comments.

Conflicts of Interest

The authors have no conflicts to disclose.

References

References

[1] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA: a Cancer Journal for Clinicians 71 (3) (2021) 209–249.

[2] G. X. Wu, D. J. Raz, Lung cancer screening, Lung Cancer (2016) 1–23.

[3] R. L. Siegel, K. D. Miller, H. E. Fuchs, A. Jemal, Cancer statistics, 2021., CA: a Cancer Journal for Clinicians 71 (1) (2021) 7–33.

[4] N. L. S. T. R. Team, Reduced lung-cancer mortality with low-dose computed tomographic screening, New England Journal of Medicine 365 (5) (2011) 395–409.

[5] A. T. Society, et al., What is a lung nodule, American Journal of Respiratory and Critical Care Medicine 193 (2016) 11–12.

[6] K. Mori, Y. Saitou, K. Tominaga, K. Yokoi, N. Miyazawa, A. Okuyama, M. Sasagawa, Small nodular lesions in the lung periphery: new approach to diagnosis with CT, Radiology 177 (3) (1990) 843–849.
[7] J. W. Gurney, S. J. Swensen, Solitary pulmonary nodules: determining the likelihood of malignancy with neural network analysis., Radiology 196 (3) (1995) 823–829.

[8] M. F. McNitt-Gray, E. M. Hart, N. Wyckoff, J. W. Sayre, J. G. Goldin, D. R. Aberle, A pattern classification approach to characterizing solitary pulmonary nodules imaged on high resolution CT: preliminary results, Medical Physics 26 (6) (1999) 880–888.

[9] K. Nakamura, H. Yoshida, R. Engelmann, H. MacMahon, S. Katsuragawa, T. Ishida, K. Ashizawa, K. Doi, Computerized analysis of the likelihood of malignancy in solitary pulmonary nodules with use of artificial neural networks, Radiology 214 (3) (2000) 823–830.

[10] Y. Matsuki, K. Nakamura, H. Watanabe, T. Aoki, H. Nakata, S. Katsuragawa, K. Doi, Usefulness of an artificial neural network for differentiating benign from malignant pulmonary nodules on high-resolution CT: evaluation with receiver operating characteristic analysis, American Journal of Roentgenology 178 (3) (2002) 657–663.

[11] S. Matsuoka, Y. Kurihara, K. Yagihashi, H. Niimi, Y. Nakajima, Peripheral solitary pulmonary nodule: CT findings in patients with pulmonary emphysema, Radiology 235 (1) (2005) 266–273.

[12] K. Suzuki, F. Li, S. Sone, K. Doi, Computer-aided diagnostic scheme for distinction between benign and malignant nodules in thoracic low-dose CT by use of massive training artificial neural network, IEEE Transactions on Medical Imaging 24 (9) (2005) 1138–1150.

[13] T. Okada, S. Iwano, T. Ishigaki, T. Kitasaka, Y. Hirano, K. Mori, Y. Sue-naga, S. Naganawa, Computer-aided diagnosis of lung cancer: definition and detection of ground-glass opacity type of nodules by high-resolution computed tomography, Japanese Journal of Radiology 27 (2) (2009) 91–99.
[14] H. Chen, Y. Xu, Y. Ma, B. Ma, Neural network ensemble-based computer-aided diagnosis for differentiation of lung nodules on CT images: clinical evaluation, Academic Radiology 17 (5) (2010) 595–602.

[15] N. Emaminejad, W. Qian, Y. Guan, M. Tan, Y. Qiu, H. Liu, B. Zheng, Fusion of quantitative image and genomic biomarkers to improve prognosis assessment of early stage lung cancer patients, IEEE Transactions on Biomedical Engineering 63 (5) (2015) 1034–1043.

[16] W. Shen, M. Zhou, F. Yang, C. Yang, J. Tian, Multi-scale convolutional neural networks for lung nodule classification, in: International Conference on Information Processing in Medical Imaging, Springer, 2015, pp. 588–599.

[17] J. S. Kirby, S. G. Armato, K. Drukker, F. Li, L. Hadiiski, G. D. Tourassi, L. P. Clarke, R. M. Engelmann, M. L. Giger, G. Redmond, et al., Lungx challenge for computerized lung nodule classification, Journal of Medical Imaging 3 (4) (2016) 044506.

[18] A. A. Farag, A. Ali, S. Elshazly, A. A. Farag, Feature fusion for lung nodule classification, International Journal of Computer Assisted Radiology and Surgery 12 (10) (2017) 1809–1818.

[19] A. Nibali, Z. He, D. Wollersheim, Pulmonary nodule classification with deep residual networks, International Journal of Computer Assisted Radiology and Surgery 12 (10) (2017) 1799–1808.

[20] F. Shaukat, G. Raja, A. Gooya, A. F. Frangi, Fully automatic detection of lung nodules in CT images using a hybrid feature set, Medical Physics 44 (7) (2017) 3615–3629.

[21] Y. Xie, Y. Xia, J. Zhang, Y. Song, D. Feng, M. Fulham, W. Cai, Knowledge-based collaborative deep learning for benign-malignant lung nodule classification on chest CT, IEEE Transactions on Medical Imaging 38 (4) (2018) 991–1004.
[22] Y. Xie, J. Zhang, Y. Xia, Semi-supervised adversarial model for benign-malignant lung nodule classification on chest CT, Medical Image Analysis 57 (2019) 237–248.

[23] H. Zhang, Y. Gu, Y. Qin, F. Yao, G.-Z. Yang, Learning with sure data for nodule-level lung cancer prediction, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2020, pp. 570–578.

[24] E. R. Park, I. F. Gareen, A. Jain, J. S. Ostroff, F. Duan, J. D. Sicks, W. Rakowski, M. Diefenbach, N. A. Rigotti, Examining whether lung screening changes risk perceptions: National lung screening trial participants at 1-year follow-up, Cancer 119 (7) (2013) 1306–1313.

[25] P. M. Thalanayar, N. Altintas, J. L. Weissfeld, C. R. Fuhrman, D. O. Wilson, Indolent, potentially inconsequential lung cancers in the pittsburgh lung screening study, Annals of the American Thoracic Society 12 (8) (2015) 1193–1196.

[26] A. R. Larici, A. Farchione, P. Franchi, M. Ciliberto, G. Cicchetti, L. Calandriello, A. Del Ciello, L. Bonomo, Lung nodules: size still matters, European Respiratory Review 26 (146).

[27] H.-J. Kim, J. Y. Cho, Y. J. Lee, J. S. Park, Y.-J. Cho, H. I. Yoon, J.-H. Chung, S. Cho, K. Kim, K. W. Lee, et al., Clinical significance of pleural attachment and indentation of subsolid nodule lung cancer, Cancer Research and Treatment 51 (4) (2019) 1540.

[28] B. H. Heidinger, U. Schwarz-Nemec, K. R. Anderson, C. de Margerie-Mellon, A. C. Monteiro Filho, Y. Chen, M. E. Mayerhoefer, P. A. VanderLaan, A. A. Bankier, Visceral pleural invasion in pulmonary adenocarcinoma: differences in CT patterns between solid and subsolid cancers, Radiology: Cardiothoracic Imaging 1 (3) (2019) e190071.
[29] Y. Zhu, R. Yip, N. You, C. I. Henschke, D. F. Yankelevitz, Management of nodules attached to the costal pleura at low-dose CT screening for lung cancer, Radiology 297 (3) (2020) 710–718.

[30] M. Gaeta, I. Pandolfo, S. Volta, E. Russi, G. Bartiromo, G. Girone, F. La Spada, M. Barone, G. Casablanca, A. Minutoli, Bronchus sign on CT in peripheral carcinoma of the lung: value in predicting results of transbronchial biopsy., American Journal of Roentgenology 157 (6) (1991) 1181–1185.

[31] M. Gaeta, M. Barone, E. Russi, S. Volta, G. Casablanca, P. Romeo, F. La Spada, A. Minutoli, Carcinomatous solitary pulmonary nodules: evaluation of the tumor-bronchi relationship with thin-section CT, Radiology 187 (2) (1993) 535–539.

[32] J. Qiang, K. Zhou, G. Lu, Q. Wang, X. Ye, S. Xu, L. Tan, The relationship between solitary pulmonary nodules and bronchi: multi-slice CT—pathological correlation, Clinical Radiology 59 (12) (2004) 1121–1127.

[33] Y. Cui, D.-q. Ma, W.-h. Liu, Value of multiplanar reconstruction in MSCT in demonstrating the relationship between solitary pulmonary nodule and bronchus, Clinical Imaging 33 (1) (2009) 15–21.

[34] Y. Kawata, N. Niki, H. Ohmatsu, M. Kusumoto, R. Kakinuma, K. Mori, H. Nishiyama, K. Eguchi, M. Kaneko, N. Moriyama, Differential geometry based vector fields for characterizing surrounding structures of pulmonary nodules, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2000, pp. 348–357.

[35] Y. Wang, K.-r. Liang, X.-g. Liu, J.-a. Wang, J.-h. Kang, M.-z. Liang, Relationship between peripheral lung cancer and the surrounding bronchi, pulmonary arteries, pulmonary veins: a multidetector CT observation, Clinical imaging 35 (3) (2011) 184–192.
[36] X. Wang, J. K. Leader, R. Wang, D. Wilson, J. Herman, J.-M. Yuan, J. Pu, Vasculature surrounding a nodule: a novel lung cancer biomarker, Lung Cancer 114 (2017) 38–43.

[37] Z.-G. Yang, S. Sone, S. Takashima, F. Li, T. Honda, T. Yamanda, Small peripheral carcinomas of the lung: thin-section CT and pathologic correlation, European Radiology 9 (9) (1999) 1819–1825.

[38] V. Rigau, T. Molina, C. Chaffaud, G. Huchon, J. Audouin, S. Chevret, J. Brechot, Blood vessel invasion in resected non small cell lung carcinomas is predictive of metastatic occurrence, Lung Cancer 38 (2) (2002) 169–176.

[39] F. Gao, M. Li, X. Ge, X. Zheng, Q. Ren, Y. Chen, F. Lv, Y. Hua, Multi-detector spiral CT study of the relationships between pulmonary ground-glass nodules and blood vessels, European Radiology 23 (12) (2013) 3271–3277.

[40] S. G. Armato III, G. McLennan, L. Bidaut, M. F. McNitt-Gray, C. R. Meyer, A. P. Reeves, B. Zhao, D. R. Aberle, C. I. Henschke, E. A. Hoffman, et al., The lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on CT scans, Medical Physics 38 (2) (2011) 915–931.

[41] A. Gupta, T. Saar, O. Martens, Y. L. Moullec, Automatic detection of multisize pulmonary nodules in CT images: Large-scale validation of the false-positive reduction step, Medical Physics 45 (3) (2018) 1135–1149.

[42] F. Han, H. Wang, G. Zhang, H. Han, B. Song, L. Li, W. Moore, H. Lu, H. Zhao, Z. Liang, Texture feature analysis for computer-aided diagnosis on pulmonary nodules, Journal of Digital Imaging 28 (1) (2015) 99–115.

[43] A. K. Dhara, S. Mukhopadhyay, A. Dutta, M. Garg, N. Khandelwal, A combination of shape and texture features for classification of pulmonary nodules in lung CT images, Journal of Digital Imaging 29 (4) (2016) 466–475.
[44] S. Hussein, R. Gillies, K. Cao, Q. Song, U. Bagci, Tumornet: Lung nodule characterization using multi-view convolutional neural network with gaussian process, in: IEEE International Symposium on Biomedical Imaging, IEEE, 2017, pp. 1007–1010.

[45] W. Shen, M. Zhou, F. Yang, D. Yu, D. Dong, C. Yang, Y. Zang, J. Tian, Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification, Pattern Recognition 61 (2017) 663–673.

[46] W. Shen, M. Zhou, F. Yang, D. Dong, C. Yang, Y. Zang, J. Tian, Learning from experts: Developing transferable deep features for patient-level lung cancer prediction, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2016, pp. 124–131.

[47] A. Fedorov, M. Hancock, D. Clunie, M. Brochhausen, J. Bona, J. Kirby, J. Freymann, S. Pieper, H. Aerts, R. Kikinis, et al., Standardized representation of the LIDC annotations using DICOM, Tech. rep., PeerJ Preprints (2019).

[48] A. Fedorov, M. Hancock, D. Clunie, M. Brochhausen, J. Bona, J. Kirby, J. Freymann, S. Pieper, H. JWL Aerts, R. Kikinis, et al., Dicom re-encoding of volumetrically annotated lung imaging database consortium (lidc) nodules, Medical Physics 47 (11) (2020) 5953–5965.

[49] W. H. Horsthemke, D. S. Raicu, J. D. Furst, Evaluation challenges for bridging semantic gap: Shape disagreements on pulmonary nodules in the lung image database consortium, International Journal of Healthcare Information Systems and Informatics 4 (1) (2009) 17–33.

[50] H. Lin, C. Huang, W. Wang, J. Luo, X. Yang, Y. Liu, Measuring interobserver disagreement in rating diagnostic characteristics of pulmonary nodule using the lung imaging database consortium and image database resource initiative, Academic Radiology 24 (4) (2017) 401–410.
[51] D. G. Kleinbaum, K. Dietz, M. Gail, M. Klein, M. Klein, Logistic regression, Springer, 2002.

[52] Y. Qin, H. Zheng, Y. Gu, X. Huang, J. Yang, L. Wang, Y.-M. Zhu, Learning bronchiole-sensitive airway segmentation cnns by feature recalibration and attention distillation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Cham, 2020, pp. 221–231.

[53] Y. Qin, H. Zheng, Y. Gu, X. Huang, J. Yang, L. Wang, F. Yao, Y.-M. Zhu, G.-Z. Yang, Learning tubule-sensitive cnns for pulmonary airway and artery-vein segmentation in ct, IEEE Transactions on Medical Imaging.

[54] T.-C. Lee, R. L. Kashyap, C.-N. Chu, Building skeleton models via 3-d medial surface axis thinning algorithms, CVGIP: Graphical Models and Image Processing 56 (6) (1994) 462–478.

[55] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, Í. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, SciPy 1.0 Contributors, SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, Nature Methods 17 (2020) 261–272.

[56] M. Anzidei, A. Porfiri, F. Andrani, M. Di Martino, L. Saba, C. Catalano, M. Bezzi, Imaging-guided chest biopsies: techniques and clinical results, Insights into Imaging 8 (4) (2017) 419–428.

[57] N. Kothary, L. Lock, D. Y. Sze, L. V. Hofmann, Computed tomography–guided percutaneous needle biopsy of pulmonary nodules: impact of nodule size on diagnostic accuracy, Clinical Lung Cancer 10 (5) (2009) 360–363.