



# A Dataset of Pulmonary Lesions With Multiple-Level Attributes and Fine Contours

Ping Li<sup>1†</sup>, Xiangwen Kong<sup>2†</sup>, Johann Li<sup>2†</sup>, Guangming Zhu<sup>2\*</sup>, Xiaoyuan Lu<sup>1</sup>, Peiyi Shen<sup>1</sup>, Syed Afaq Ali Shah<sup>3</sup>, Mohammed Bennamoun<sup>4</sup> and Tao Hua<sup>5</sup>

<sup>1</sup> Shanghai BNC, Shanghai, China, <sup>2</sup> Embedded Technology & Vision Processing Research Center, Xidian University, Xi'an, China, <sup>3</sup> College of Science, Health, Engineering and Education, Murdoch University, Perth, WA, Australia, <sup>4</sup> School of Computer Science and Software Engineering, The University of Western Australia, Perth, WA, Australia, <sup>5</sup> Pet Center, Huashan Hospital, Fudan University, Shanghai, China

## OPEN ACCESS

### Edited by:

Kezhi Li,  
University College London,  
United Kingdom

### Reviewed by:

Tao Chen,  
Virginia Tech, United States  
Jian Guo,  
RIKEN Center for Computational  
Science, Japan  
Zhimin Liu,  
Janssen Pharmaceuticals, Inc.,  
United States

### \*Correspondence:

Guangming Zhu  
gmzhu@xidian.edu.cn

<sup>†</sup>These authors have contributed  
equally to this work and share first  
authorship

### Specialty section:

This article was submitted to  
Health Informatics,  
a section of the journal  
Frontiers in Digital Health

**Received:** 23 September 2020

**Accepted:** 09 December 2020

**Published:** 17 February 2021

### Citation:

Li P, Kong X, Li J, Zhu G, Lu X, Shen P,  
Shah SAA, Bennamoun M and Hua T  
(2021) A Dataset of Pulmonary  
Lesions With Multiple-Level Attributes  
and Fine Contours.  
*Front. Digit. Health* 2:609349.  
doi: 10.3389/fgdh.2020.609349

Lung cancer is a life-threatening disease and its diagnosis is of great significance. Data scarcity and unavailability of datasets is a major bottleneck in lung cancer research. In this paper, we introduce a dataset of pulmonary lesions for designing the computer-aided diagnosis (CAD) systems. The dataset has fine contour annotations and nine attribute annotations. We define the structure of the dataset in detail, and then discuss the relationship of the attributes and pathology, and the correlation between the nine attributes with the chi-square test. To demonstrate the contribution of our dataset to computer-aided system design, we define four tasks that can be developed using our dataset. Then, we use our dataset to model multi-attribute classification tasks. We discuss the performance in 2D, 2.5D, and 3D input modes of the classification model. To improve performance, we introduce two attention mechanisms and verify the principles of the attention mechanisms through visualization. Experimental results show the relationship between different models and different levels of attributes.

**Keywords:** deep learning, radiology, pulmonary dataset, classification, attention

## 1. INTRODUCTION

Lung cancer is caused by tumors which leads to the fastest increase in morbidity and mortality. It has a significant negative impact on the health of subjects. Therefore, the early diagnosis of lung lesions is of great significance for the treatment of lung cancer.

The early form of lung cancer is categorized as pulmonary nodules, which are clinically examined using computed tomography (CT). The characteristics of pulmonary nodules in CT images are diverse, which results in a large workload for radiologists to diagnosis the disease and leads to the subjective assessment of features. Therefore, accurate and quantitative analysis of the appearance characteristics of lung nodules is very essential for doctors to determine whether the nodules will grow into malignant tumors.

In recent years, with the development of deep learning technology (1), lung nodule diagnosis has made unprecedented progress in detection (2–7), segmentation (8–11), classification (2, 6, 12–15), and registration (16, 17) tasks. In order to improve the performance of the model, there is a great need of large datasets and accurate annotation of pulmonary lesions.

There are many publicly available datasets of pulmonary nodules. However, there are some shortcomings in the existing datasets, and the diversity of lesions cannot be balanced in these

datasets. For example, LIDC/IDRI (18) has rich attributes, however, it only marks nodules, and the prediction of other pulmonary diseases cannot be performed.

In this paper, we propose a dataset of lung lesions that could help the development of a pulmonary computer-aided diagnosis system. Our dataset is multi-centered, data-diversified, and informative. The proposed dataset is rich in lesion types and covers most of the signs of lung lesions. The lesions of the dataset are labeled with contours and attribute annotations by experienced radiologists using a professional tool. The attribute annotations are composed of nine attributes that are most useful for pathological assessment. In order to make the selected attributes hierarchical, we have selected multi-level attributes:

- **Low-level attributes:** Margin, spiculation, etc, which can be judged basically by the local features of the lesion;
- **Middle-level attributes:** Pleural indentation, vessel convergence, etc, which need to be judged by the relationship with the surrounding tissue around the lesion or cavity and calcification, which need to be judged by the relationship between local features and global features of the lesion;
- **High-level attributes:** The type and the location of the lesion, which requires to be judged by the abstract features of the entire lesion.

In order to describe the proposed dataset clearly, we first count the characteristics of our dataset, define the data storage format and data annotation rules for our dataset. We then propose the contours annotation format. We also focus on the correlation between the attributes of the lesions. In order to study the relationship between multiple attributes, we calculated the probability of a total of 27 categories of 9 different attributes using the chi-square test and conditional probability, and infer the correlation with the attributes by probability.

In order to illustrate the practical significance of our dataset, we discuss several applications that could be studied using our dataset, and then select the attribute classification for further study. First, we model the attribute classification and then explored the performance of the 2D, 2.5D, and 3D input modes on the accuracy of the model. Through experiments, we demonstrate that there is implicit competition between multiple attributes, we, therefore, use two attention mechanisms to filter different feature activations for different attributes. Our experiments show that the attention mechanisms have different effects on attribute classification.

## 2. RELATED WORK

In this section, we briefly discuss the existing datasets of lung nodules and the relevant classification methods.

### 2.1. Lung Nodule Datasets

#### 2.1.1. LUNA16 Dataset

The LUNA16 (4) dataset was designed for the Open Pulmonary Nod Challenge, which screened 888 CT volumes from a large dataset LIDC/IDRI as challenge data. Their slice thickness is within 2.5 mm and the nodule size is greater than 3 mm, which was annotated by more than 3 experimental doctors using tow-phase annotation. The detection annotations of a nodule

in LUNA16 use the center coordinates and diameter of the inscribed circle of the nodule. In contrast, we use the gravity center coordinates as the center coordinates of the nodule and the longer geometric moment as the diameter to generate the world coordinates. For small round nodules, the two datasets are not much different, but the need is to detect large lesions with irregular shapes and our proposed approach achieves better results for large lesion detection.

#### 2.1.2. LIDC/IDRI Dataset

The LIDC/IDRI (18) dataset labels each nodule with a contour and nine attributes. Besides the benign and malignant nodules, the other eight attributes are all the appearance attributes of the nodules. In contrast, in our dataset, two of the attributes are the basic attributes of the lesion, five are appearance attributes, and two have relationships with the tissue surrounding the lesion in context. These attributes are richer and can better represent a lesion.

#### 2.1.3. LISS Database

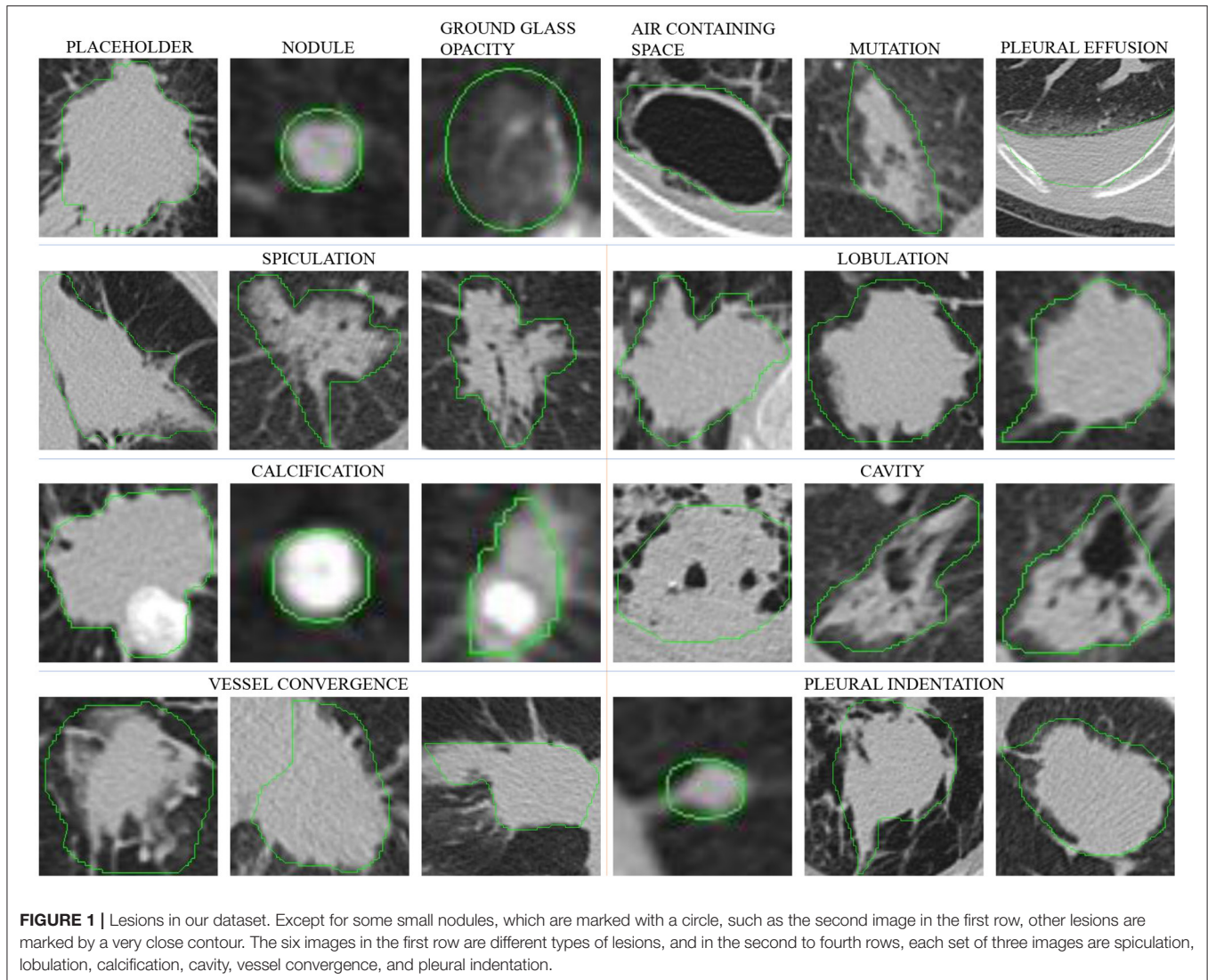
The LISS (19) database has 271 CT volumes, including 677 abnormal regions. These abnormal regions are divided into nine categories, which are called common CT imaging signs of lung disease (CISLs). In other words, there is only one CISLs label for each abnormal region. Although it can better help medical scholars learn a certain type of disease (12), it is not very good for CAD system development, because it cannot capture the relationship between disease signs.

#### 2.1.4. ILD Database

The ILD (20) database has 108 image series with more than 1946 ROIs. This dataset is a multimedia collection of cases of interstitial lung disease (ILDs). These ROIs are divided into 13 categories, which are lung tissue patterns from histological diagnoses of ILDs. The lesions in the ILD dataset are large, and the annotations are all high-level attributes. The dataset does not focus on a certain nodule, but on the pathology presented by a piece of tissue.

## 2.2. Lung Nodule Classification

The classification of lung nodules based on deep learning can be divided into two types of methods: one is to judge the benign and malignant lung nodules. Some methods directly predict the benign and malignant nodules by CT images, and other methods use different attributes of the nodules as the auxiliary basis to judge the benign and malignant nodules, such as (21–23). The other type of method has classified the disease, such as DeepLung (2) or LISCs classification (12). Dey et al. (21) have built a network that produces multiple outputs from multi-scale features to judge the benign and malignant nodules. Nibali et al. (22) has made a three-column configuration to fuse the features generated from three axes. Song et al. (14, 23) proposed methods that split the whole image into patches and predict the lesions. In contrast, Gao et al. (13) have used the whole image for classification. With the development of computationally efficient computers, the 3D models such as (24) has achieved an impressive performance in nodule classification. He (12) proposed a method to generate images for data augmentation,



which achieved a good improvement in performance. Zhu et al. (2) detected the position of the nodules first, then cropped the nodules before feeding it into a classification model to predict one of nine attributes.

Multi-attribute classification is a problem to classify multiple targets using one model. There are currently two approaches to solve this problem. The first is to regard it as a classification task with a fixed number of categories, and solve attribute correlation in one model by using multiple branches to decompose the relationship between multiple targets onto each branch. The second is to treat it as a multi-label classification task, with the positive attribute as the label of the lesion, then each lesion has a floating number of labels, and the labels are decoupled using different methods. In this paper, we use the first method to classify different attributes in a model using a fixed number of branches, and use two attention mechanisms to help decouple the correlation among the attributes.

### 3. LUNG LESION DATASET

In this section, we provide a description of our dataset. CT data were collected from four hospitals. The body parts examined are mainly the chest and abdomen. Among them, the chest CT was mostly thin (less than 3 mm), and the abdomen CT was mostly thick (greater than or equal to 5 mm). **Figure 1** shows examples of lesions in our dataset. As shown in **Figure 1**, except for some small nodules, which are marked with circles, such as the second image in the first row, other lesions are marked by a very close contour.

**Table 1** shows the parameter comparison of our dataset with several other public datasets. Same with LUNA16, our dataset annotates lesion with contour, which is shown in **Figure 1**. Compared with box and polygon, contour annotation has more generalization ability to different tasks, such as location, detection, and segmentation. At the same time, though the

**TABLE 1** | The statistical result of comparing our dataset parameters with other datasets.

Dataset	Annotation	Lesion attributes	Multiple categories	Scans	Lesion amount	Lesion size (mm)	Slice thickness (mm)	Pixel spacing (mm)
LUNA16	contour	9	✓	888	1,186	3.25–32.27	0.45–2.50	0.461–0.976
LISS (2D)	Box	9	×	252	511	–	5.0	0.42–1.00
LISS (2D)	Box	9	×	19	166	–	1, 1.25	0.60–0.87
ILD	Polygon	13	×	108	1,946	–	1.00–2.00	0.40–1.00
Ours	Contour	9	✓	694	5,113	0.83–191.32	1.00–2.00	0.176–0.977

number of scans in our dataset is not the largest, the number of lesion annotations and the range of lesion size in our dataset are. These annotations support more robust models. Moreover, the thickness of the slices of our datasets is relatively uniform, especially compared to LUNA 16. It reduces unnecessary processing of the data and makes it easier to use.

### 3.1. File Storage and Annotation Format

The raw data obtained from the hospital contains some sensitive information of subjects, and the data collected from different hospitals are stored in different ways, making the data difficult to use directly for analysis. Therefore, we first desensitize the data by removing subjects' sensitive information and retain only the necessary information, such as weight. Then, we store the CT volumes and annotation files as described below.

We define the directory structure to store files as follows:

```
ct_type / hospital / year / month / day / subject_id / series_id .
```

The directory with `series_id SE01` stores the CT data with DICOM format, and the directory with `series_id SE01_01_0n` stores the contour annotation file `aid_loc.anno`, where  $n$  is the identification number of the doctor who annotated the scans; `aid` is the number of the annotation in the CT for correspondence with the attribute information; `loc` is the slice number in the CT volume, and the description in the DICOM file is `SliceLocation (0020, 1041)`. An `anno` file represents an annotation. Each `anno` file has a different `aid`, but two `anno` files can have the same `loc`, indicating that the two annotations are in the same slice. It uses a dictionary to store the annotation information we need to use in the CAD tasks. The keywords of the `anno` format are `SeriesID`, `NoduleSerialNumber`, `InstanceNumber`, `Origin`, `Dimension`, `Spacing`, `Coords`, `XMin`, `XMax`, `YMin`, `YMax`. Among them, `SeriesID` is a unique number of a DICOM volume which described as `SeriesInstanceUID (0020, 000E)`, `NoduleSerialNumber` and `InstanceNumber` are `aid` and `loc`, respectively as mentioned above, `Origin`, `Dimension`, `Spacing` are the information from DICOM volume, `Coords` is the contour coordinate of this annotation, and its value is relative to the size of this slice. (`XMin`, `YMin`), (`XMax`, `YMax`) are the coordinates of the lower left and upper right corners of the bounding box of this annotation.

The CT volumes in our dataset contain lesions, while those without lesions have been removed by manually screening of RIS reports. For repeated subject numbers, such as two volumes of one subject, we map one of them to a new subject number and retain the correspondence to restore the original number.

## 3.2. Two-Phase Annotation Process

We use a two-phase annotation process to label the lesions. We label the contours of the lesions in the first phase, then label the attributes of the lesions in the second phase.

### 3.2.1. Contour Annotation Criterion

The contours are marked by experienced radiologists. In order to save the doctor's time and to increase the density of the lesion, we first manually screen the RIS report, retain the CT volume with the lesion in the description, and remove the volume without the lesion from the dataset. In order to standardize the process of marking the lesions, we have prescribed a rule for marking lesions with the doctor as follows:

- Mark all visible lesions;
- If the lesion is too small to draw the contour, circle the lesion

with a circle tool;

- If the lesion is larger than one slice, mark the lesion every three consecutive slices;
- Draw a contour as close as possible to the edge of a lesion.

After the marking process, we perform a secondary screening to remove the annotations which are too discontinuous to be processed as contours. Then, we convert the annotations into `anno` format and mark lesion numbers. In this way, the contour annotations and the attribute annotations correspond with respective file names.

### 3.2.2. Attribute Annotation Criterion

After discussed with the doctor, we selected nine attributes that are commonly used in clinical diagnosis as attribute annotations for the dataset. A detailed description of these attributes will be provided in section 5.2. Each lesion is independently labeled by a doctor, and we record the doctor's number for each lesion that can be used to identify the doctor if an error is discovered in the annotation.

In order to simplify the labeling of attributes, we implement an attribute labeling tool to collect and manage labels. We associate the slice of the contour with the lesion number so that it is convenient to label the attributes with the corresponding slice. When the attribute information is marked, the corresponding subject number and label number are recorded to correspond to the contour number. It should be noted that the contour annotation and the attribute annotation are not one-to-one



matched. Some problematic contour annotations are filtered out in the previous step, and no attribute annotation is performed. Finally, we only select lesions with both contour and attribute annotations into the dataset. The number of attributes is reported in **Table 1**. As can be noted, the categories of some attributes are very unbalanced. This brings great challenges to the performance of our attribute classification algorithm.

## 4. ATTRIBUTES AND PATHOLOGY

We initially selected 15 attributes that are commonly used in clinical diagnosis, and then selected 9 attributes for our dataset based on their importance. The number of categories of these attributes is not balanced and the distributions are not independent. Here we briefly describe the importance of these attributes in clinical diagnosis and then discuss the correlation between attributes from the statistics point of view.

### 4.1. Attributes Description

Among the 9 attributes we selected, besides the basic attribute, lesion type, and lesion location, there are vessel convergence and pleural indentation which represent the relationship between the lesion and the surrounding tissue. On the other hand, margin, calcification, lobulation, spiculation, cavity represent the apparent features of the lesion. The description of the significance of these nine attributes is as follows.

#### 4.1.1. Lesion Type

The first row of **Figure 1** shows six different lesion types. For the lesion type, we choose *placeholder*, *nodule*, *ground glass opacity*, *air containing space*, *mutation*, and *pleural effusion*. The difference between placeholders and nodules is that the lesions with a diameter of less than 30 mm are nodules, and those larger than 30 mm are placeholders. Except for the difference in size, the other attributes of the two lesion types are roughly similar. The air containing space is different from the cavity in pathology. The air containing space (**Figure 1**, the fifth image in the first row) is a pathological enlargement of the physiological cavity in the lung, while the cavities (**Figure 1**, the last three images in the third row) often appears in nodules or placeholders. In the air containing space lesions, the wall of the lesion is thinner and more uniform, mostly occurring in the subpleural area, and the size varies greatly. This means that the location of the air containing space is fixed and there are no apparent attributes such as spiculation and lobulation.

#### 4.1.2. Lesion Location

The location of the nodule is represented by five categories of lobes, including the *right upper lobe*, the *right middle lobe*, the *right lower lobe*, the *left upper lobe*, and the *left lower lobe*. Statistics show that the occurrence of lesions has little relationship with the location. The lesion location is only a basic attribute of the lesion, and it cannot be used as a basis for judging its pathological nature. Some lesions are large and span multiple lung lobes, so we mark them as 0, and do not include it in the five categories above.

#### 4.1.3. Margin

The margin attribute describes whether the outer boundary of a nodule is clear. We defined two main categories for this attribute: *clear* and *unclear* margin. Though the margin of a benign mass is often smooth, while that of a malignant mass is often unclear, inflammation may also cause an unclear margin of placeholder. Therefore, it cannot be used as the sole basis for judging benign and malignant lesion, and needs to be judged in combination with other attributes.

#### 4.1.4. Calcification

The calcification attribute describes lesions whose density is significantly higher than other soft tissues in the mediastinal window, usually with CT values above 100 Hu. The first three images in the third row of **Figure 1** show lesions of calcification. The white region in the images represents calcification. Calcification is a pathologically metamorphic lesion, which is more common in the healing stage of ductal tuberculosis lesions in the lung tissue or lymph nodes; calcification can also occur in tumor tissues or cyst walls. Usually, the greater the proportion of calcification in the lesion, the greater the likelihood of its being benign. Based on this, we classify the calcification attributes into three categories: *no*, *partial*, and *total* calcification.

#### 4.1.5. Lobulation

The lobulation attribute indicates that the nodule or mass grows at different speeds in various directions or is blocked by the surrounding structure. The contours may have a plurality of arcuate protrusions, and the curved phases are concave cuts to form a lobulated shape. The last three images in the second row of **Figure 1** show the lesions of lobulation. We can clearly see the convex part of the masses. We simply define two categories for this attribute: *with* and *without* lobulation.

#### 4.1.6. Spiculation

The spiculation attribute is characterized by a radial, unbranched, straight, and strong thin line shadow extending from the edge of the nodule to the periphery, and the proximal end of the shadow is slightly thicker. The first three images in the second row of **Figure 1** show lesions of spiculation. As shown in **Figure 1**, the burrs of the lesion are often not circled in the scope of annotation. The spiculation is not connected to the pleura, and distinct from the pleural depression. We classify the spiculation attributes into *no*, *short* and *long* spiculation; 5 mm burrs are called short spiculation, and larger than 5 mm burrs are called long spiculation. The pathological basis of the burr is the fiber band in which the tumor cells infiltrate into the adjacent bronchial sheath and local lymphatic vessels, or the tumor promotes connective tissue formation. Benign nodular inflammatory pseudotumor, tuberculoma can also be seen burrs, but longer, softer, more often formed by hyperplastic fibrous connective tissue. The possibility of lung cancer should be considered when there is a burr in solitary lung nodules.

#### 4.1.7. Cavity

The cancerous cavities are mostly located in the anterior segment of the upper lobe and the basal segment of the lower lobe. Most

of the cavities larger than 3 cm in diameter are tumors. Most cancerous cavities present an irregular or lobulated outer edge and irregular inner edge. Those with a wall thinner than 4 mm are mostly benign lesions, and those thicker than 15 mm are mostly malignant lesions. The last three images in the third row of **Figure 1** show the lesions of a cavity. We simply defined two categories for this attribute: *with* and *without* cavity.

#### 4.1.8. Vessel Convergence

The vessel convergence attribute appears on the slices as one or more vessels around the pulmonary nodule that touch with, cut or pass through the placeholder at its edge. The appearance of vessel convergence is related to the size of the placeholder or nodule. The lesions less than 1 cm in diameter have fewer vessel convergence signs. The first three images in the last row of **Figure 1** shows the lesions of vessel convergence. Images of the cavities and vessel convergence are similar, because the blood vessels look like cavities when they are transected. A multi-vessel-directed lesion presents vessel convergence, which leads to a higher chance of malignancy. In particular, the phenomenon that one blood vessel leads to a nodule or tumor is not only seen in malignant nodules, but also in benign lesions such as tuberculosis, inflammatory pseudotumor, or hamartoma. We simply defined two categories for this attribute: *with* and *without* vessel convergence.

#### 4.1.9. Pleural Indentation

The typical pleural indentation shows a small triangular shadow or a small trumpet shadow on the visceral surface of the visceral pleura. The bottom of the triangle is on the inside of the chest wall, the tip points on the nodule, and the nodule and the triangle shadow can be connected by a linear shadow. The last three images in the last row of **Figure 1** shows the lesions of pleural indentation. Peripheral lesions of the pleural indentation are often accompanied by other imaging signs. The pathological basis and imaging manifestations of pleural indentation in benign and malignant lesions are different. We simply define two categories for this attribute: *with* and *without* pleural indentation.

## 4.2. Correlation Between Attributes

In order to evaluate the correlation between attributes, we used the chi-square test. We assume that if the two attributes are independent of each other, their data distribution should not affect each other, which means that the proportional relationship between the categories of one attribute is the same under each category of the other attribute. If the chi-square test value calculated by the two attributes is greater than the statistical significance, there is a correlation between the two attributes. The approximate calculation equation for the chi-square test statistic is as follows:

$$\chi^2 = \sum \frac{(f_0 - f_e)^2}{f_e} \quad (1)$$

where  $f_0$  is the actual number of observations and  $f_e$  is the expected number of times. The larger the value of  $f_e$ , the Equation (1) approximately obeys the chi-square distribution. To simplify

the calculation of the chi-square test, we used a variant of Equation (1):

$$\chi^2 = \sum \frac{\left(f_{xy} - \frac{f_x f_y}{N}\right)^2}{\frac{f_x f_y}{N}} = N \left( \sum_{x=1}^R \sum_{y=1}^C \frac{f_{xy}^2}{f_x f_y} - 1 \right) \quad (2)$$

where  $f_x$  and  $f_y$  represent the number of samples of the categories of two different attributes  $x$  and  $y$ , respectively,  $R$  and  $C$  are the number of categories of  $f_x$  and  $f_y$ , and the total number of attributes is  $N$ . The degree of freedom  $df$  of the independence test is calculated as follows:

$$df = R \times C - R - C - 1 = (R - 1)(C - 1) \quad (3)$$

We use the data shown in **Table 2** and select a significance level of 0.05 for calculation. **Figure 2A** shows the result of the chi-square test. As the results show, there is a strong correlation between the three attributes of margin, spiculation, and lobulation. Meanwhile, there is a strong correlation between vessel convergence and spiculation, margin, lobulation and lesion type, pleural indentation, and margin.

To further explore the specific relationship between the various categories of attributes, we calculated the conditional probability between a total of 27 categories for all attributes. The equation for calculating the conditional probability is as follows:

$$P(X|Y) = \frac{P(XY)}{P(Y)} \quad (4)$$

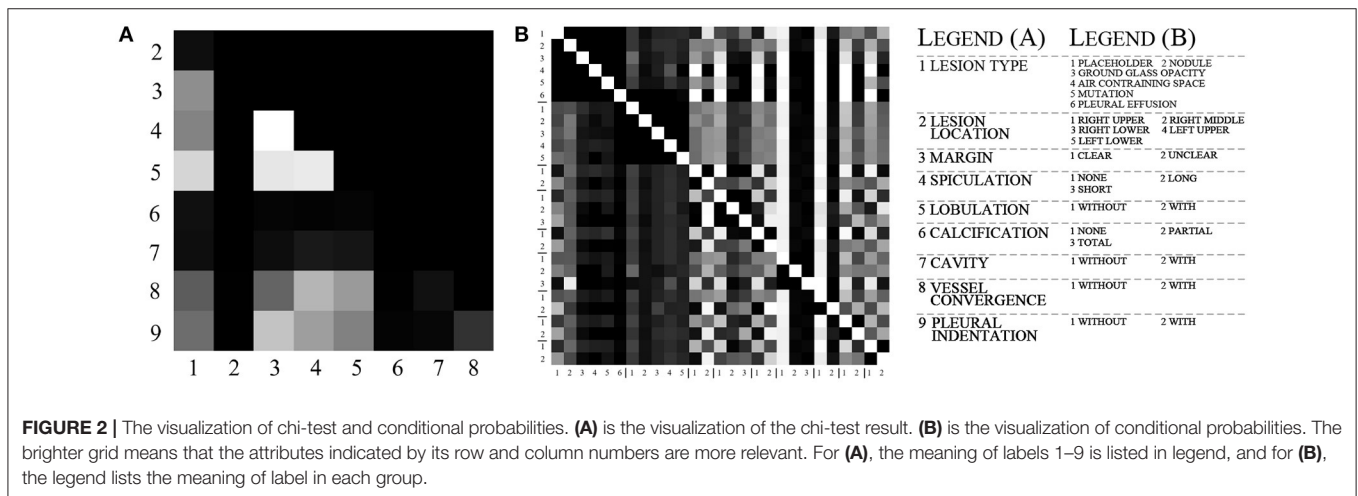
where  $P(X)$  and  $P(Y)$  represent the probabilities of two categories  $X$  and  $Y$ ,  $P(X|Y)$  represents the probability of  $X$  to occur when  $Y$  is present, and  $P(XY)$  represents the probability of co-occurrence for  $X$  and  $Y$ . The value of  $P(X|X)$  is 1, which is represented by white color in **Figure 2B**. We calculated the conditional probability between each of the two categories. As shown in **Figure 2B**, the white color represents a probability of 1 and the black color represents a probability of 0, while the lighter gray color represents higher conditional probability values.

According to the statistical results, there is a strong correlation between different lesion types and other attributes. For the placeholder, their margins are almost unclear, the degree of lobulation is more obvious, the degree of spiculation and the degree of pleural indentation are the highest among other lesion types. The nodules, ground glass, and mutation categories have a small number of spiculation and lobulation, and more features of vessel convergence and pleural indentation. For cavity and pleural effusion, they almost have no other attributes and their margins are all clear.

The margin attribute is highly correlated with lobulation, vessel convergence, and pleural indentation. When vessel convergence and pleural indentation are present, they are often accompanied by lobulation, and the margin is not very clear. The calcification attribute is concentrated in the nodules, and the cavity is also related to the margin and lobulation.

**TABLE 2** | The distribution of each attribute category used for experiments.

Attribute	Categories	Lesions	Attribute	Categories	Lesions	Attribute	Categories	Lesions
Lesion type	Placeholder	675	Lesion location	Right upper	496	Calcification	None	1,902
	Nodule	728		Right middle	151		Partial	62
	Ground glass opacity	220		Right lower	286		Total	50
	Air containing space	153		Left upper	374	Cavity	Without	1,924
	Mutation	208		Left lower	271		With	90
	Pleural effusion	30		Margin	Clear	887	Vessel Convergence	Without
Spiculation	None	1,198	Unclear	1,127	With	553		
	Long	307	Lobulation	Without	1,015	Without	1,222	
	Short	509		With	999	With	792	

**FIGURE 2** | The visualization of chi-test and conditional probabilities. (A) is the visualization of the chi-test result. (B) is the visualization of conditional probabilities. The brighter grid means that the attributes indicated by its row and column numbers are more relevant. For (A), the meaning of labels 1–9 is listed in legend, and for (B), the legend lists the meaning of label in each group.

## 5. TASKS OF DATASET

Our dataset is rich in data and diverse in annotations, which means that our dataset can be used for several tasks and aid in the development of CAD systems. We recommend using our dataset for the following tasks:

- (1) **Detection:** Some of the lesions in our dataset are smaller than 30 mm, which are nearly circular and suitable for lung nodules detection. This can be helpful for the initial diagnosis of lung cancer.
- (2) **Segmentation:** The lesions larger than 30 mm are all marked with precise contours. These lesions are more complex in shape and are suitable for the lung lesion segmentation task. This can be helpful for volume measurement and further treatment.
- (3) **Classification:** Multiple attributes of the lesion are suitable for multi-task lung disease prediction. This can be helpful to judge benign and malignant tumors.
- (4) **Reconstruction:** At present, medical datasets are small, and their size is not enough for deep learning. Our dataset has various types of data, and we can use real data to train generative adversarial networks to generate synthetic data.

In this paper, we focus on exploring the correlation between attributes. We, therefore, perform multi-attribute

classification and report our experimental results in section 6.

### 5.1. 2D, 2.5D, 3D Modes for Classification

In order to study the importance of the input mode for the model, we use different data dimensions for the same data and the model for classification experiments.

We use three input modes including 2D, 2.5D, and 3D. Assuming that the size of a CT volume is  $H \times W \times C$ , which corresponds to the three axes of X-Y-Z, the diameter of a lesion is  $d$ , the three input modes are expressed as follows:

#### 5.1.1. 2D Mode

The lesion is cut out from the grayscale slice in which it is located with a length  $d$  of side, and fed to a 2D network for prediction. The input size is  $d \times d \times 1$ . The 2D input mode can retain the lesion at the spatial structure in the X-Y direction, but the context information in the Z direction cannot be captured.

#### 5.1.2. 2.5D Mode

The grayscale image of the lesion and the five images above and below are cut out by the bounding box, and fed to the 2D network for prediction. The number of input channels is 5, and the input size is  $d \times d \times 5$ . Compared to the 2D input mode, the 2.5D

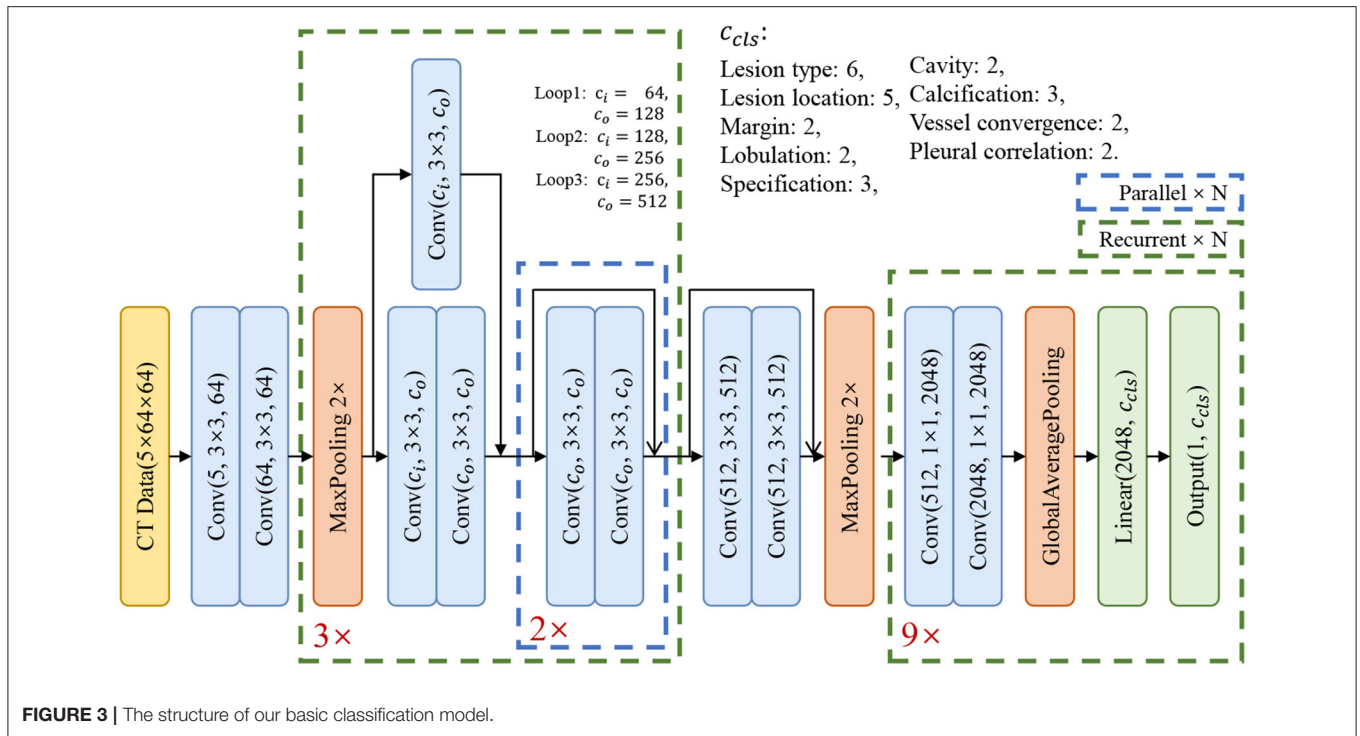


FIGURE 3 | The structure of our basic classification model.

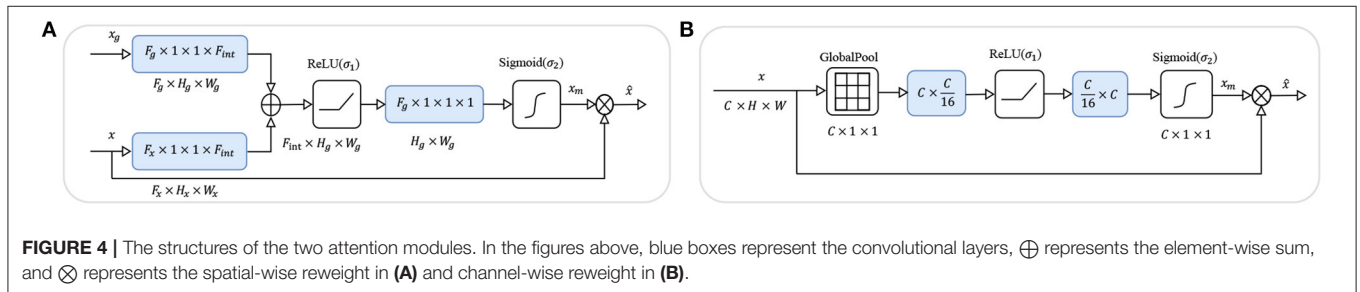


FIGURE 4 | The structures of the two attention modules. In the figures above, blue boxes represent the convolutional layers,  $\oplus$  represents the element-wise sum, and  $\otimes$  represents the spatial-wise reweight in (A) and channel-wise reweight in (B).

input mode is supplemented by a fixed number of slices in the Z direction.

### 5.1.3. 3D Mode

In the X-Y-Z direction where the lesion is located, the bounding box ( $d \times d \times d$ ) is cropped and fed to the 3D network for prediction. 3D network can capture the correlation on the Z-axis of the whole lesion by convolution. Compared with 2D, the information of 2.5D is more detailed, but the amount of 3D network parameters is more than that of 2D network, which can cause the deep learning model to overfit as the size of training data is small.

The architecture of our basic model is shown in Figure 3. In order to extract the relationship of nine attributes, we use a ResNet-based network (25) to extract the characteristics of the nodule and then use nine classification branches to predict nine attributes independently. We will explain the details and the results in section 6.1.

## 5.2. Two Attention Mechanisms

Through the experiments, we found that there is an implicit competition between multiple attributes during training. In the training phase, when the loss value is stable, the accuracy of some attributes increases while the accuracy of other attributes decreases. To solve this problem, we add an attention module in front of each attribute classifier to focus the activation on the features which are useful for classification. In this way, different input features for attributes are extracted, which could mitigate the conflict between attributes. Inspired by (26–28), we employed soft-attention and self-attention, commonly used mechanisms that compute a weight matrix used to filter noise and to focus on important features. These two attention mechanisms are described below in our model, and Figure 4 shows the structure of the two attention modules.

### 5.2.1. Soft-Attention Module

As shown in Figure 4A, we add a soft-attention module (26) before feeding the features into the classifier to filter out



shallower features with deeper features. While preserving the spatial structure, the attention module extracts a mask from the features to suppress noise which is not related to the attribute to improve accuracy.

Assuming that feature map  $x \in \mathbb{R}^{N \times C_x \times H \times W}$  from the basic model is the input feature for the attention model, and feature map  $x_g \in \mathbb{R}^{N \times C_g \times H \times W}$  is from a deeper layer as the gate, we firstly use  $1 \times 1$  convolutional layer to get the same number of channels  $C_g$  for both the features, then sum the features  $x$  and  $x_g$  together and add a non-linear transform ReLU which can be formulated as  $\sigma_1(x) = \max(0, x)$ . So far, the feature  $x$  is mixed with richer semantic information  $x_g$ , and we use a  $1 \times 1$  convolutional layer to fuse the channel information and retain the spatial information, and get a mask  $x_m$  with a value of  $[0, 1]$  through the sigmoid function which can be

formulated as  $\sigma_2(x) = (1 + e^{-x})^{-1}$ . Finally, we use the mask  $x_m$  to spatial-wise reweight the feature map  $x$  and get the output feature  $\hat{x}$ . After filtering by the soft-attention module, the features  $\hat{x}$  are re-weighted by high-dimensional semantic information in the spatial dimension, which is more conducive to multi-attribute classification.

## 5.2.2. Self-Attention Module

As shown in **Figure 4B**, we add a self-attention module (27, 28) before the features and fed to the classifier to squeeze the spatial structure of a feature map into one vector with spatial information. Then, we gather and filter the information to enhance the activation related to that attribute, and add the information to the original feature map to enhance the feature.

**TABLE 3** | Performance of the basic model on the 3D, 2.5D, and 2D modes.

Attributes	Categories	Accuracy			Sensitivity			Specificity		
		3D	2.5D	2D	3D	2.5D	2D	3D	2.5D	2D
Lesion type	Placeholder	0.8636	0.8182	<b>0.8864</b>	0.7451	<b>0.8780</b>	0.7959	0.9500	0.9344	<b>0.9561</b>
	Nodule	0.7460	0.6984	<b>0.7619</b>	0.8545	0.8462	<b>0.9057</b>	0.8621	0.8288	<b>0.8636</b>
	Ground glass opacity	<b>0.8800</b>	0.8000	0.8000	<b>0.8462</b>	0.8333	0.8000	<b>0.9793</b>	0.9640	0.9638
	Air containing space	0.9375	<b>1.0000</b>	0.9091	<b>0.9375</b>	0.4400	0.7143	0.9935	<b>1.0000</b>	0.9933
	Mutation	0.8235	<b>0.9286</b>	0.8571	0.8235	<b>0.8667</b>	0.8571	0.9805	<b>0.9932</b>	0.9866
	Pleural effusion	1.0000	1.0000	1.0000	<b>1.0000</b>	<b>1.0000</b>	0.7500	1.0000	1.0000	1.0000
Margin	Clear	0.8437	0.8523	<b>0.8636</b>	<b>0.8617</b>	0.8427	0.8261	0.8052	0.8243	<b>0.8310</b>
	Unclear	<b>0.8267</b>	0.8133	0.7867	0.8052	0.8243	<b>0.8310</b>	<b>0.8617</b>	0.8427	0.8261
Spiculation	None	<b>0.8672</b>	0.8083	0.8083	0.9407	<b>0.9604</b>	<b>0.9604</b>	<b>0.6792</b>	0.6290	0.6290
	Long	0.3333	0.6000	<b>0.7333</b>	<b>0.4167</b>	0.2571	0.2558	0.9371	0.9531	<b>0.9667</b>
	Short	<b>0.7143</b>	0.2857	0.4286	0.4878	0.2963	<b>0.6316</b>	<b>0.9385</b>	0.8529	0.8889
Lobulation	Without	0.8972	0.8990	<b>0.9091</b>	0.9143	0.9271	<b>0.9375</b>	0.8333	0.8507	<b>0.8657</b>
	With	0.8594	0.8906	<b>0.9062</b>	0.8333	0.8507	<b>0.8657</b>	0.9143	0.9271	<b>0.9375</b>
Calcification	None	<b>0.9937</b>	0.8684	0.7763	0.9464	<b>0.9565</b>	0.9516	<b>0.6667</b>	0.2000	0.1282
	Partial	0.0000	<b>0.2500</b>	<b>0.2500</b>	0.0000	<b>0.1176</b>	0.0741	0.9529	<b>0.9589</b>	0.9559
	Total	0.6667	<b>1.0000</b>	<b>1.0000</b>	<b>1.0000</b>	0.3750	0.2500	0.9941	<b>1.0000</b>	<b>1.0000</b>
Cavity	Without	<b>0.9819</b>	0.9557	0.9367	0.9702	0.9742	<b>0.9801</b>	0.0000	0.1250	<b>0.1667</b>
	With	0.0000	0.2000	<b>0.4000</b>	0.0000	0.1250	<b>0.1667</b>	0.9702	0.9742	<b>0.9801</b>
Vessel convergence	Without	<b>0.9618</b>	0.8618	0.8780	0.8936	0.8548	<b>0.9076</b>	<b>0.8333</b>	0.5641	0.6591
	With	0.6250	0.5500	<b>0.7250</b>	<b>0.8333</b>	0.5641	0.6591	0.8936	0.8548	<b>0.9076</b>
Pleural indentation	Without	<b>0.8500</b>	0.8125	0.7946	0.8430	0.8922	<b>0.9082</b>	0.6400	<b>0.6557</b>	0.6462
	With	0.6275	0.7843	<b>0.8235</b>	0.6400	<b>0.6557</b>	0.6462	0.8430	0.8922	<b>0.9082</b>
Lesion location	Right upper	<b>1.0000</b>	0.7083	0.7083	1.0000	1.0000	1.0000	<b>1.0000</b>	0.8793	0.8793
	Right middle	<b>0.8571</b>	0.7500	0.7500	<b>1.0000</b>	0.3000	0.3000	0.9934	<b>0.9929</b>	<b>0.9929</b>
	Right lower	<b>1.0000</b>	0.8750	0.8333	<b>0.9630</b>	0.6562	0.7143	<b>1.0000</b>	0.9746	0.9672
	Left upper	<b>1.0000</b>	0.7143	0.8571	0.9118	<b>0.9524</b>	0.9231	<b>1.0000</b>	0.9380	0.9677
	Left lower	0.9348	<b>0.9783</b>	0.9565	<b>1.0000</b>	0.8491	0.8462	0.9739	<b>0.9897</b>	0.9796
	Average	0.7513	0.7511	<b>0.7816</b>	<b>0.7671</b>	0.7006	0.7184	<b>0.8305</b>	0.7995	0.8116

*Bold value means better performance, compared between 2D, 2.5D, 3D.*

Assuming that feature map  $x \in \mathbb{R}^{N \times C_x \times H \times W}$  is generated from the basic model as the input feature of the attention model, we use a channel squeeze and spatial excitation branch to transform  $x$  to extract the spatial information and reweight the origin  $x$  with the transform of itself. We use a global pool which can squeeze  $x$  to a vector  $z \in \mathbb{R}^{N \times C_x \times 1 \times 1}$ . Then use two fully connected layers to transform the vector  $z$  to  $\hat{z} = W_1(\sigma_1(W_2 \cdot z))$  with  $W_1 \in \mathbb{R}^{C \times C/16}$  and  $W_2 \in \mathbb{R}^{C/16 \times C}$  and the activation  $\sigma_1$ . We also use the non-linear function  $\sigma_2$  to transform the values to  $[0, 1]$  to get the channel mask  $x_m$ . Finally, we use the  $x_m$  to channel-wise reweight the feature map  $x$  and get the output feature  $\hat{x}$ . After filtering by the self-attention module, the features  $\hat{x}$  are re-weighted by the information after squeeze and excitation in the spatial dimension, which is more conducive to multi-attribute classification.

## 6. EXPERIMENTAL RESULTS

In this section, we first verify that the proposed model can learn the correlation between attributes, and then empirically select the best input mode, and verify the attention mechanism on this input mode.

We used part of the data with a thickness of 1.0–2.0 mm in our experiments, which has 355 CT volumes and 2014 lesions labeled with 9 attributes in our dataset. The dataset has been split into 8:2 as the training set and validation set, with 1,847 lesions in the training set and 163 lesions in the validation set. During training, we randomly select 30% of the data for data augmentation i.e., random flip and rotation. As **Table 2** shows, the number of categories in the dataset is unbalanced, which could affect the convergence of the model. We use weighted cross

**TABLE 4** | Performance of the basic, soft-attention, and self-attention models on the 2D mode.

Attributes	Categories	Accuracy			Sensitivity			Specificity		
		Basic model	Soft-att	Self-att	Basic model	Soft-att	Self-att	Basic model	Soft-att	Self-att
attention Lesion type	Placeholder	<b>0.8864</b>	0.8182	0.8636	<b>0.7959</b>	0.7347	0.7755	<b>0.9561</b>	0.9298	0.9474
	Nodule	<b>0.7619</b>	0.6825	0.7302	0.9057	0.8776	<b>0.9787</b>	<b>0.8636</b>	0.8246	0.8534
	Ground glass opacity	0.8000	<b>0.9200</b>	0.8800	<b>0.8000</b>	0.7419	0.7857	0.9638	<b>0.9848</b>	0.9778
	Air containing space	0.9091	<b>1.0000</b>	0.9091	0.7143	<b>0.7857</b>	0.7143	0.9933	<b>1.0000</b>	0.9933
	Mutation	0.8571	0.9286	<b>1.0000</b>	0.8571	<b>1.0000</b>	0.7368	0.9866	0.9933	<b>1.0000</b>
	Pleural effusion	<b>1.0000</b>	<b>1.0000</b>	0.8333	0.7500	<b>0.8571</b>	0.8333	<b>1.0000</b>	<b>1.0000</b>	0.9936
Margin	Clear	<b>0.8636</b>	0.8523	<b>0.8636</b>	0.8261	<b>0.8621</b>	0.8352	0.8310	0.8289	<b>0.8333</b>
	Unclear	0.7867	<b>0.8400</b>	0.8000	0.8310	0.8289	<b>0.8333</b>	0.8261	<b>0.8621</b>	0.8352
Spiculation	None	<b>0.8083</b>	0.7750	0.7917	0.9604	<b>0.9894</b>	0.9500	<b>0.6290</b>	0.6087	0.6032
	Long	<b>0.7333</b>	0.5333	0.4667	0.2558	<b>0.3478</b>	0.2500	<b>0.9667</b>	0.9500	0.9407
	Short	0.4286	<b>0.7500</b>	0.5357	<b>0.6316</b>	0.4565	0.4286	0.8889	<b>0.9402</b>	0.8984
Lobulation	Without	0.9091	0.9091	<b>0.9192</b>	0.9375	0.9474	<b>0.9479</b>	0.8657	0.8676	<b>0.8806</b>
	With	0.9062	<b>0.9219</b>	<b>0.9219</b>	0.8657	0.8676	<b>0.8806</b>	0.9375	0.9474	<b>0.9479</b>
Calcification	None	0.7763	0.8158	<b>0.8224</b>	0.9516	<b>0.9538</b>	0.9398	0.1282	<b>0.1515</b>	0.1000
	Partial	<b>0.2500</b>	<b>0.2500</b>	0.0000	0.0741	<b>0.1111</b>	0.0000	0.9559	<b>0.9586</b>	0.9437
	Total	1.0000	1.0000	1.0000	0.2500	0.2000	<b>0.3333</b>	1.0000	1.0000	1.0000
Cavity	Without	0.9367	0.8734	<b>0.9557</b>	0.9801	<b>0.9857</b>	0.9805	0.1667	0.1304	<b>0.2222</b>
	With	0.4000	<b>0.6000</b>	0.4000	0.1667	0.1304	<b>0.2222</b>	0.9801	<b>0.9857</b>	0.9805
Vessel convergence	Without	0.8780	0.7967	<b>0.8211</b>	0.9076	<b>0.9515</b>	0.9182	<b>0.6591</b>	0.5833	0.5849
	With	0.7250	<b>0.8750</b>	0.7750	<b>0.6591</b>	0.5833	0.5849	0.9076	<b>0.9515</b>	0.9182
Pleural indentation	Without	<b>0.7946</b>	0.7857	0.7589	0.9082	0.9167	<b>0.9551</b>	<b>0.6462</b>	0.6418	0.6351
	With	0.8235	0.8431	<b>0.9216</b>	<b>0.6462</b>	0.6418	0.6351	0.9082	0.9167	<b>0.9551</b>
lesion Location	Right upper	0.7083	0.7083	0.7083	<b>1.0000</b>	<b>1.0000</b>	0.9714	<b>0.8793</b>	<b>0.8793</b>	0.8783
	Right middle	0.7500	0.7500	0.7500	0.3000	<b>0.3333</b>	0.3000	0.9929	0.9929	0.9929
	Right lower	0.8333	0.8333	<b>0.8750</b>	0.7143	0.6897	<b>0.7241</b>	0.9672	0.9669	<b>0.9752</b>
	Left upper	0.8571	0.8214	<b>0.9286</b>	0.9231	0.9200	<b>0.9630</b>	0.9677	0.9600	<b>0.9837</b>
	Left lower	0.9565	0.9565	0.9565	0.8462	0.8302	<b>0.8980</b>	0.9796	0.9794	<b>0.9802</b>
	Average	0.7816	<b>0.8032</b>	0.7763	<b>0.7184</b>	0.7183	0.7155	0.8116	0.8117	<b>0.8128</b>

*Bold value means better performance, compared between different models.*

entropy loss to reduce the impact of data imbalance during the training phase.

In the experiments, each model has four blocks. The first one is a convolutional block and the other three are residual blocks. At the end of the model, there are nine classifier blocks for the classification of nine attributes, respectively. We use the reweighted logistic loss to balance the numbers of categories. During the training phase, we set the learning rate to 0.01 with warm restart (29) and use SGD to optimize the model. The momentum was set to 0.09, the weight decay was set to  $10^{-4}$  and the batch size was set to 64. Since the model converges quickly, we have trained 200 epochs for each model and choose the model with the smallest validation loss as the best model.

The imbalanced data causes that no valid features can be learned, and results in low sensitivity of the model to this attribute. As shown in **Tables 3, 4**, categories with too few samples, such as *partial calcification* and *with cavity*, were not recognized. A given category prediction may have the following four cases: TP, True Positive; FP, False Positive; TN, True Negative; FN, False Negative.

To evaluate the imbalanced categories of each attribute, we use three metrics to score the results. **Accuracy** (ACC) is the basic metric to evaluate the result, which can be calculated as:

$$ACC = \frac{TP + TN}{P + N} \tag{5}$$

**Sensitivity** (SE), also called the true positive rate, means the probability that a sick person is diagnosed as positive, which can be calculated as:

$$SE = \frac{TP}{TP + FN} \tag{6}$$

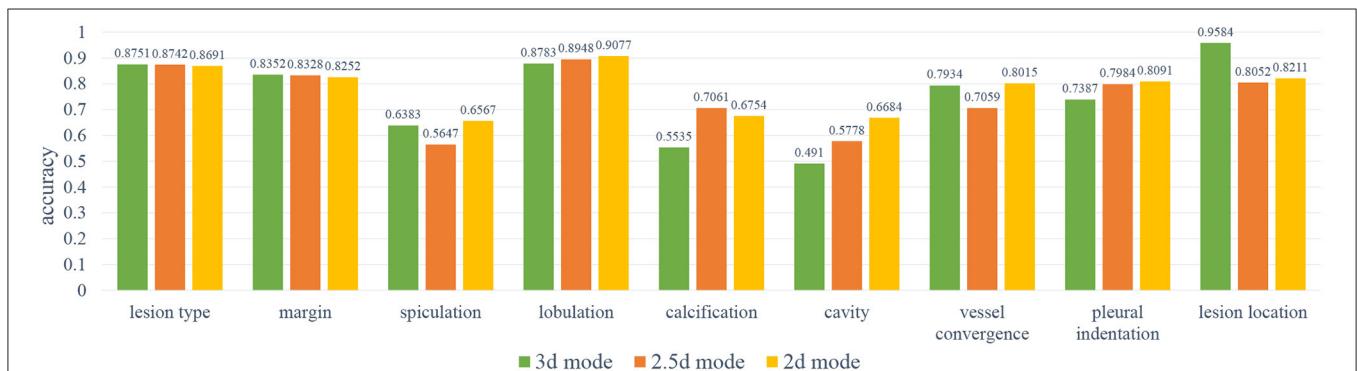
The larger the SE value, the more sensitive our model is in diagnosing this category.

**Specificity** (SP), also called the true negative rate, means the probability that a person who is actually not sick is diagnosed as negative, which can be calculated as:

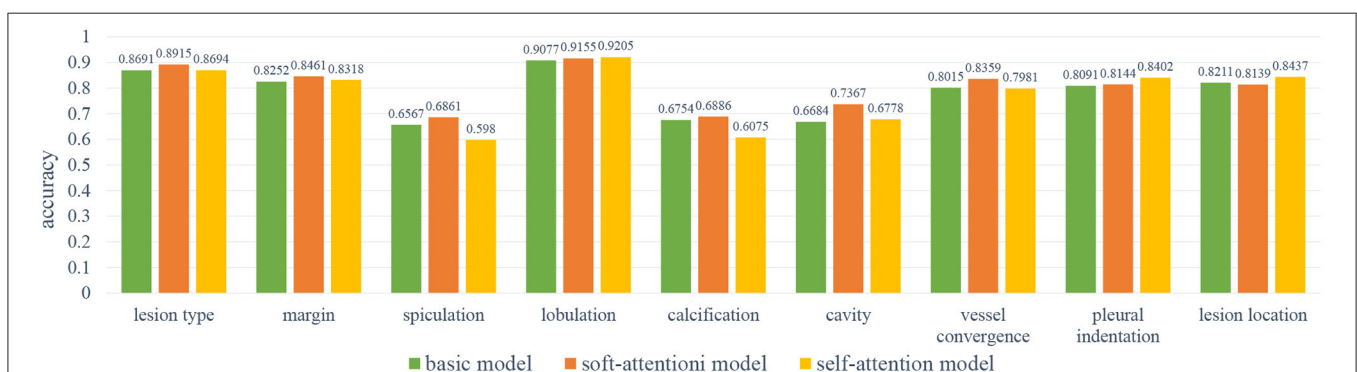
$$SP = \frac{TN}{FP + TN} \tag{7}$$

The larger the value of SP, the more accurate our model is for the diagnosis of this category.

We average out accuracies of all categories for each attribute, and average the scores of all attributes as the final score to represent the performance of the model.



**FIGURE 5 |** The results of the 2D, 2.5D, 3D input modes. As can be noted, the 3D mode has better results on spiculation, lobulation, cavity, vessel convergence, and pleural indentation.



**FIGURE 6 |** The results of the base model and two attention models. As can be noted, the self-attention module has better results on lobulation, pleural indentation, and lesion location attributes; the soft-attention module has better results on lesion type, margin, spiculation, calcification, cavity, and vessel convergence attributes.

## 6.1. Results for Input Modes

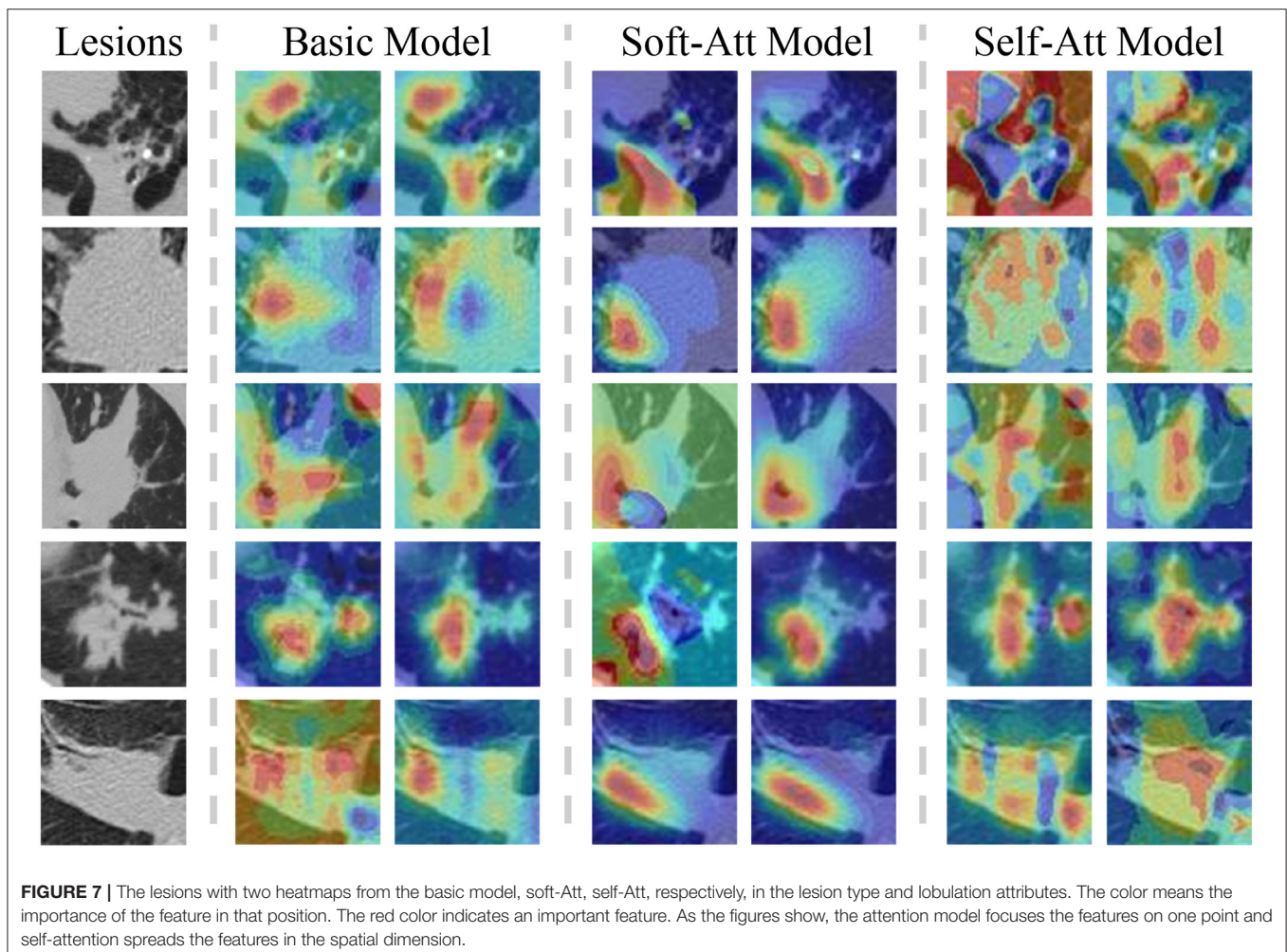
In order to select the most suitable input mode for the attribute classification of lung lesions, we train the 2D, 2.5D, and 3D model with the same structure described in **Figure 3**. To ensure the fairness of the three models, we do not adjust the hyper-parameters for different models. Each model was trained with 200 epochs and a batch size of 64. To evaluate the performance of the models, we chose the average accuracy of the model with the lowest validation loss as the metric. The average accuracy scores of the 3D, 2.5D, 2D model are 0.7513, 0.7511, and 0.7816; the average sensitivity are 0.7671, 0.7006, and 0.7184; and the average specificity are 0.8305, 0.7995, and 0.8116, respectively. As **Figure 5** shows, the three models have almost the same scores in lesion type and margin, and the model with 2D mode has better scores in spiculation, lobulation, vessel convergence, and pleural indentation. **Table 3** shows the accuracy, sensitivity, and specificity of each category for each attribute. From the experimental results, we note that the higher-level attributes, such as lesion type and lesion location, are more sensitive to the 3D mode and the lower-level attributes, such as spiculation and lobulation, are more sensitive to the 2D mode.

During training, we noticed that the 3D model has more parameters than the 2D models, which led to longer training time and slower convergence. Meanwhile, the 2D model has better average accuracy than the 3D model. So, we chose the 2D mode as the basic model for the following experiments.

## 6.2. Results for Attention Mechanisms

In order to improve the performance of the basic model, we have used two attention mechanisms to enhance the feature before feeding it to the classifiers. We called the model with the soft-attention module *Soft-Att*, and the model with the self-attention module *Self-Att*. Since the number of parameters of the two attention modules is not large, we use the same hyper-parameters as the basic model to train the two models. Similar to the previous section, we used a batch size of 64 and 200 epochs for training and taking the accuracy of the model with the lowest validation loss as the metric. The average accuracies scores of the basic model, *Soft-Att* and *Self-Att* are 0.7816, 0.8032, and 0.7763; the average sensitivities are 0.7184, 0.7183, and 0.7155; and the average specificities are 0.8116, 0.8117, and 0.8128, respectively.

As **Figure 6** shows, the soft-attention module has better results on margin, vessel convergence, lesion type, and spiculation





attributes, and the self-attention module has better results on lobulation, pleural indentation, and lesion location attributes. Due to the near-zero sensitivity of calcification and cavity attributes, we do not take their accuracy into comparison. As reported in **Table 4**, the two models with attention modules have better performance than the basic model.

The heatmaps in **Figure 7** visualize the attention mechanisms. Compared with the basic model, the red value of soft-attention is concentrated at one point. This is because soft-attention uses higher-layer semantic information to filter the low-layer features, which makes the features spatially smoother and more focused. This is a good feature for high-level attributes because it is concentrated at the point that best reflects the attribute, but it does not fully reflect the local information relationship. Compared with the basic model, the red value of self-attention is more scattered in the spatial dimension. This is because self-attention extracts channel information by compressing spatial information using its own features, and it is more comprehensive in spatial information due to multi-channel fusion. This is a good feature for low-level attributes because its local information relationships are more spatially refined, but because of the noise in the spatial dimension, it may not be appropriate for high-level attributes.

## 7. CONCLUSION

This paper presents a dataset of lung lesions with fine contour annotation and attribute and explores the correlation between the attributes of the dataset. To demonstrate the contribution of this dataset to the development of CAD systems, we explore two issues of medical data modeling using attribute classification tasks.

One of the issues is the effect of the 2D, 2.5D, 3D input mode on the classification model. The 2D mode works well for low-level attributes that do not require local information relationships between lesions and surrounding tissues, while the 3D mode works better for high-level attributes that require higher contextual relationships. The 2.5D mode is a trade-off between the lightweight of the 2D model and the context information of the 3D model.

The second is the impact of the two attention mechanisms on the model. Soft-attention can better handle the noise in the

spatial dimension and concentrate on the features at one point, which is beneficial for the classification of high-level attributes. Self-attention can better integrate the spatial information in the channel dimension, and complement the local relationship in the spatial dimension, which is beneficial for the classification of low-level attributes.

In the future, we mainly want to explore and address the following three issues:

1. For the three categories of cavity, partial calcification, and long spiculation, the sensitivity is almost zero due to the high degree of the category imbalance. We will explore novel methods to improve the accuracy of these three categories.
2. We will use the correlation between attributes to establish a loss function suitable for multi-attribute classification from the statistical learning strategy.
3. There is not a single metric that can well measure the performance of a multi-attribute model. We will build evaluation metrics for multi-task modeling.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

## FUNDING

This work was supported by the Shanghai Science and Technology Committee (No. 18411952100).

## ACKNOWLEDGMENTS

The authors would like to acknowledge all of the contributors to A Dataset of Pulmonary Lesions with Multiple-Level Attributes and Fine Contours.

## REFERENCES

1. Altaf F, Islam SMS, Akhtar N, Janjua NK. Going deep in medical image analysis: concepts, methods, challenges, and future directions. *IEEE Access*. (2019) 7:99540–72. doi: 10.1109/ACCESS.2019.2929365
2. Zhu W, Liu C, Fan W, Xie X. DeepLung: Deep 3D dual path nets for automated pulmonary nodule detection and classification. In: *Proceedings - 2018 IEEE Winter Conference on Applications of Computer Vision, WACV 2018*. Lake Tahoe (2018). p. 673–81. doi: 10.1109/WACV.2018.00079
3. Zhu W, Vang YS, Huang Y, Xie X. DeepEM: deep 3D ConvNets with EM for weakly supervised pulmonary nodule detection. In: *Lecture Notes in Computer Science*. Granada (2018). p. 812–20. doi: 10.1007/978-3-030-00934-2\_90
4. Setio AAA, Traverso A, de Bel T, Berens MSN, van den Bogaard C, Cerello P, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge. *Med Image Anal*. (2017) 42:1–13. doi: 10.1016/j.media.2017.06.015
5. Li Z, Wang C, Han M, Xue Y, Wei W, Li LJ, et al. Thoracic disease identification and localization with limited supervision. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT (2018). p. 8290–9. doi: 10.1109/CVPR.2018.00865
6. Masood A, Sheng B, Li P, Hou X, Wei X, Qin J, et al. Computer-assisted decision support system in pulmonary cancer detection and stage classification on CT images. *J Biomed Inform*. (2018) 79:117–28. doi: 10.1016/j.jbi.2018.01.005
7. Gonzalez G, Ash SY, Vegas-Sánchez-Ferrero G, Onieva JO, Rahaghi FN, Ross JC, et al. Disease staging and prognosis in smokers using deep learning in chest

- computed tomography. *Am J Respirat Crit Care Med.* (2018) 197:193–203. doi: 10.1164/rccm.201705-0860OC
8. Duan J, Schlemper J, Bai W, Dawes TJW, Bello G, Doumou G, et al. Deep nested level sets: fully automated segmentation of cardiac MR images in patients with pulmonary hypertension. In: *Lecture Notes in Computer Science*. Granada (2018). p. 595–603. doi: 10.1007/978-3-030-00937-3\_68
  9. LaLonde R, Bagci U. Capsules for object segmentation. *arXiv [Preprint]*. *arXiv:1804.04241*. (2018).
  10. Kim J. Lung nodule segmentation with convolutional neural network trained by simple diameter information. In: *Medical Imaging with Deep Learning (MIDL 2018)*. Amsterdam (2018). p. 3–5.
  11. Burlutskiy N, Gu F, Wilen LK, Backman M, Micke P. A deep learning framework for automatic diagnosis in lung cancer. *arXiv [Preprint]*. *arXiv:1807.10466*. (2018).
  12. He G. Lung CT Imaging sign classification through deep learning on small data. *arXiv [Preprint]*. *arXiv:1903.00183*. (2019).
  13. Gao M, Bagci U, Lu L, Wu A, Buty M, Shin HC, et al. Holistic classification of CT attenuation patterns for interstitial lung diseases via deep convolutional neural networks. *Comput Methods Biomech Biomed Eng Imaging Visual.* (2018) 6:1–6. doi: 10.1080/21681163.2015.1124249
  14. Song Y, Cai W, Zhou Y, Feng DD. Feature-based image patch approximation for lung tissue classification. *IEEE Trans Med Imaging.* (2013) 32:797–808. doi: 10.1109/TMI.2013.2241448
  15. Wang X, Chen H, Gan C, Lin H, Dou Q, Tsougenis E, et al. Weakly supervised deep learning for whole slide lung cancer image analysis. *IEEE Trans Cybern.* (2020). 50:3950–62. doi: 10.1109/TCYB.2019.2935141
  16. Lafarge MW, Moeskops P, Veta M, Pluim JPW, Eppenhof KAJ. Deformable image registration using convolutional neural networks. In: Angelini ED, Landman BA, editors. *Medical Imaging 2018: Image Processing*. SPIE. Houston, TX (2018). p. 27. doi: 10.1117/12.2292443
  17. Castillo E, Castillo R, Martinez J, Shenoy M, Guerrero T. Four-dimensional deformable image registration using trajectory modeling. *Phys Med Biol.* (2010) 55:305–27. doi: 10.1088/0031-9155/55/1/018
  18. Armato SG, McLennan G, Bidaut L, McNitt-Gray MF, Meyer CR, Reeves AP, et al. The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): a completed reference database of lung nodules on CT Scans. *Med Phys.* (2011) 38:915–31.
  19. Han G, Liu X, Han F, Santika INT, Zhao Y, Zhao X, et al. The LISS - A public database of common imaging signs of lung diseases for computer-aided detection and diagnosis research and medical education. *IEEE Trans Biomed Eng.* (2015) 62:648–56. doi: 10.1109/TBME.2014.2363131
  20. Depeursinge A, Vargas A, Platon A, Geissbuhler A, Poletti PA, Müller H. Building a reference multimedia database for interstitial lung diseases. *Comput Med Imaging Graph.* (2012) 36:227–38. doi: 10.1016/j.compmedimag.2011.07.003
  21. Dey R, Lu Z, Hong Y. Diagnostic classification of lung nodules using 3D neural networks. In: *Proceedings - International Symposium on Biomedical Imaging*. (2018). p. 774–8. doi: 10.1109/ISBI.2018.8363687
  22. Nibali A, He Z, Wollersheim D. Pulmonary nodule classification with deep residual networks. *Int J Comput Assist Radiol Surg.* (2017) 12:1799–808. doi: 10.1007/s11548-017-1605-6
  23. Song Y, Cai W, Huang H, Zhou Y, Feng DD, Wang Y, et al. Large margin local estimate with applications to medical image classification. *IEEE Trans Med Imaging.* (2015) 34:1362–77. doi: 10.1109/TMI.2015.2393954
  24. Biffi C, Oktay O, Tarroni G, Bai W, De Marvao A, Doumou G, et al. Learning interpretable anatomical features through deep generative models: application to cardiac remodeling. In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, editors. *Lecture Notes in Computer Science*. Granada: Springer (2018). p. 464–71. doi: 10.1007/978-3-030-00934-2\_52
  25. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV (2016). p. 770–8. doi: 10.1109/CVPR.2016.90
  26. Oktay O, Schlemper J, Le Folgoc L, Lee M, Heinrich M, Misawa K, et al. Attention U-Net: learning where to look for the pancreas. *arXiv [Preprint]*. *arXiv:1804.03999*. (2018).
  27. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT (2018). p. 7132–41. doi: 10.1109/CVPR.2018.00745
  28. Roy AG, Navab N, Wachinger C. Concurrent spatial and channel squeeze & excitation in fully convolutional networks. In: *Lecture Notes in Computer Science*. Granada (2018). p. 421–9. doi: 10.1007/978-3-030-00928-1\_48
  29. Loshchilov I, Hutter F. SGDR: stochastic gradient descent with warm restarts. In: *5th International Conference on Learning Representations, ICLR*. Toulon (2017).
- Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Copyright © 2021 Li, Kong, Li, Zhu, Lu, Shen, Shah, Bennamoun and Hua. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.