



# Article Transformer-Based Recognition Model for Ground-Glass Nodules from the View of Global 3D Asymmetry Feature Representation

Jun Miao<sup>1,\*</sup>, Maoxuan Zhang<sup>1</sup>, Yiru Chang<sup>1</sup> and Yuanhua Qiao<sup>2</sup>

- <sup>1</sup> School of Computer Science, Beijing Information Science and Technology University, Beijing 100101, China; mxzhang123@bistu.edu.cn (M.Z.); cc13623840893@163.com (Y.C.)
- <sup>2</sup> College of Applied Sciences, Beijing University of Technology, Beijing 100124, China; qiaoyuanhua@bjut.edu.cn
- \* Correspondence: jmiao@bistu.edu.cn

Abstract: Ground-glass nodules (GGN) are the main manifestation of early lung cancer, and accurate and efficient identification of ground-glass pulmonary nodules is of great significance for the treatment of lung diseases. In response to the problem of traditional machine learning requiring manual feature extraction, and most deep learning models applied to 2D image classification, this paper proposes a Transformer-based recognition model for ground-glass nodules from the view of global 3D asymmetry feature representation. Firstly, a 3D convolutional neural network is used as the backbone to extract the features of the three-dimensional CT-image block of pulmonary nodules automatically; secondly, positional encoding information is added to the extracted feature map and input into the Transformer encoder layer for further extraction of global 3D asymmetry features, which can preserve more spatial information and obtain higher-order asymmetry feature representation; finally, the extracted asymmetry features are entered into a support vector machine or ELM-KNN model to further improve the recognition ability of the model. The experimental results show that the recognition accuracy of the proposed method reaches 95.89%, which is 4.79, 2.05, 4.11, and 2.74 percentage points higher than the common deep learning models of AlexNet, DenseNet121, GoogLeNet, and VGG19, respectively; compared with the latest models proposed in the field of pulmonary nodule classification, the accuracy has been improved by 2.05, 2.05, and 0.68 percentage points, respectively, which can effectively improve the recognition accuracy of ground-glass nodules.

Keywords: ground-glass nodules; 3D ResNet; transformer; support vector machine; ELM-KNN

# 1. Introduction

Lung cancer is one of the cancers with the highest global mortality rate, and one of the significant factors contributing to the high mortality rate of lung cancer is late discovery and treatment [1]. In the early stage of lung cancer, lung lesions often appear in the form of pulmonary nodules, among which ground-glass nodules (GGN) are one of the main manifestations of early lung cancer [2]. Interventional treatment can achieve a cure rate of over 80% for patients with early malignant lung diseases accompanied by ground-glass nodules [3]. For lung examination, low-dose spiral CT technology is usually used. Doctors can judge lung lesions by observing the image characteristics of lung nodules in computer tomography (CT) [4]. However, with the continuous increase of medical imaging data, doctors' diagnostic work is becoming more and more arduous [5], and doctors will be affected by various factors in the process of reading films, resulting in missed diagnoses and misdiagnoses [6]. The recognition of ground-glass pulmonary nodules with the help of computer-aided diagnosis (CAD) technology based on medical imaging can reduce doctors' work intensity and improve their ability to diagnose lesions [7].



Citation: Miao, J.; Zhang, M.; Chang, Y.; Qiao, Y. Transformer-Based Recognition Model for Ground-Glass Nodules from the View of Global 3D Asymmetry Feature Representation. *Symmetry* 2023, *15*, 2192. https:// doi.org/10.3390/sym15122192

Academic Editors: Sergei D. Odintsov and Zhixun Su

Received: 15 June 2023 Revised: 7 September 2023 Accepted: 16 September 2023 Published: 12 December 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).

At present, domestic and foreign scholars have proposed many effective lung nodule recognition algorithms. The traditional machine learning method is to manually design and extract nodule features, and then select appropriate classifiers for lung nodule classification [8]. However, the feature dimensions of medical images are often high, and there may be redundant features [9], which limits the use of traditional machine learning. Deep learning can automatically learn different medical image data and obtain key features from it [10], further improving the accuracy of disease diagnosis. More and more researchers are applying it to the field of medical imaging. Gao et al. [11] used a multi-input 2D Convolutional Neural Network (CNN) [12] for the prediction of benign and malignant pulmonary nodules. Gao et al. [13] proposed a lung nodule classification method based on fused prior knowledge. Liu et al. [14] introduced center clipping into DenseNet to recognize pulmonary nodules. Zhang et al. [15] fused the semantic features of lung nodules with the benign and malignant discrimination network of lung nodules into a multi-branch network, completing the task of diagnosing benign and malignant lung nodules. However, these methods are all based on two-dimensional image classification algorithms, ignoring the characteristics of pulmonary nodules in three-dimensional spaces, and have certain limitations [16]. Some works attempt to use the three-dimensional features of pulmonary nodules for classification. Wang et al. [17] used an enhanced three-dimensional VGG16 [18] network to classify pulmonary nodules. Raunak et al. [19] used multitask learning convolutional neural networks for the segmentation and classification of pulmonary nodules. Li Wei et al. [20] fused the features of multi-layer sections of pulmonary nodules to achieve the classification of malignant degrees of pulmonary nodules. Halder et al. [21] utilized a deep learning framework based on adaptive morphology for lung nodule classification. Yi et al. [22] applied multi-scale and multipath convolutional neural networks for lung nodule classification. These models are effective in 3D image classification, but due to the limitations of local receptive fields in convolutional operations in capturing long-distance pixel relationships [23], recognition models based on convolutional neural networks have shortcomings in extracting global features of lung nodule images.

The preferred model in the field of Natural Language Processing (NLP), Transformer [24], is capable of long-distance attention computation. Inspired by Transformer, Google directly applied the standard Transformer to image classification in 2020 and proposed the model Vision Transformer [25]. Afterwards, more and more researchers are applying Transformer to image classification. However, Transformer lacks inductive bias similar to CNN [26], and this prior knowledge needs to be learned through a large amount of data. The lack of labeled data is a common problem in the field of medical images [27]. Based on the above information, this paper attempts to combine a 3D Convolutional neural network and Transformer and proposes a recognition model of ground-glass lung nodules based on 3D ResNet [28] and Transformer for better 3D asymmetry feature representation. In response to the traditional machine learning methods that rely on manual experience to extract features, this article utilizes a deep learning model with powerful feature extraction capabilities to extract features. At the same time, considering that medical images are threedimensional block images, the 3D convolution method is used in this paper to enable the neural network to better extract medical image features, and selects the ResNet model with good robustness as the backbone for feature extraction; in response to the shortcomings of CNN in extracting global image features, this paper cites the Transformer encoder layer to enable the model to learn richer global image features and enhance its feature expression ability; finally, an Support Vector Machine (SVM) [29] classifier with strong generalization ability is selected to output the recognition results. Through 10-fold cross validation on a private lung CT dataset, and comparing the model proposed in this paper with other mainstream recognition models, the experimental results show that the fused 3D ResNet and Transformer lung nodule recognition models have better classification accuracy than mainstream ground-glass lung nodule recognition methods.

To summarize, our contributions are as follows. We propose a Transformer model for 3D pulmonary nodule image recognition. This model does not rely on manual feature

3 of 16

extraction and utilizes 3D convolution to extract better image features of three-dimensional medical image blocks. The Transformer encoder layer is introduced to learn richer global image features. The recognition accuracies of this model reached 95.89% and 95.72% on a private dataset and the open Luna16 [30], respectively.

The rest of the paper is organized as follows: Section 2 describes the relevant works to our paper. In Section 3, we first describe the overall structure of the proposed model and then detail the three parts of the model, the 3D ResNet, The Transformer Encoder, and the SVM. In Section 4, we present the extensive experimental results and analysis. Finally, in Section 5, we present a summary and the limitations and future research directions of the model proposed in this paper.

## 2. Related Work

The pulmonary nodule recognition algorithm is generally based on doctors providing the location of the lung nodule, and then training the CT image of the lung nodule with a classifier to distinguish between nodules and non-nodules. It mainly utilizes relevant machine learning algorithms to extract image features. At present, computer-aided diagnosis tasks based on deep learning have achieved outstanding performance in the diagnosis of thyroid cancer, breast cancer, and skin cancer, among others [31], and the classification method of ground-glass nodules is also developing in medical diagnosis [32]. For example, Hu et al. [33] used the deep Convolutional neural network and residual architecture to segment GGN and then used the Transfer learning method to build a deep neural network to classify benign and evil GGNs. Li et al. [34] used the deep Convolutional neural network to solve the identification of solid, half-time center, and ground glass. Ni et al. [35] used a GGN detector composed of 3D U-Net and 3D multiple Receptive fields to detect the position of GGN and then used the Attention-v1 depth 3D Convolutional neural network to identify the invasiveness of GGN. Different deep learning models have also been used in the classification of ground-glass nodules, such as AlexNet [36], VGG16 [18], GoogLeNet [37], ResNet [28], etc., which have been applied in medical image analysis. Improving the network structure of CNN [12] to continuously improve the accuracy of classification has also become the focus of current image classification research [38]. For example, Dai et al. [39] proposed a CNN architecture with mixed convolutional scales, which solves the problem of reduced accuracy in electroencephalography (EEG) motor imagery classification when training data is limited. Lei et al. [40] constructed an extended CNN model by replacing the convolutional kernel of traditional CNN with an extended convolutional kernel and solved the problem of detail loss in the extended CNN model by stacking extended convolutional kernels with different expansion rates. Harsh Sharma et al. [41] used Convolutional neural network architecture with different depths to extract features from chest radiograph images and classify the images to detect whether a person has pneumonia. Most deep learning models are applied to 2D image classification [42], while CT medical images of lung nodules are three-dimensional block images, and its three-dimensional image block features are better at representing spatial asymmetry information and texture information [43]. For example, Wu Baorong et al. [44] trained the three-dimensional and two-dimensional multi-scale images of nodules and weighted their classification results to achieve the classification of pulmonary nodules. Shen et al. [45] extracted distinguishing features from images of lung nodules at different scales and sent them to a Convolutional neural network to obtain the heterogeneity between benign and malignant nodules, thus improving the classification accuracy. Zhang et al. [46] used the DenseNet architecture as the backbone, as well as 3D filters and pooling kernels, achieving a decent performance. Yuan et al. [47] designed a gradient boosting method for 3D dual-path network features. The experimental results showed that the accuracy of lung nodule classification obtained using this method was improved by about 3% compared to 2D CNN, strongly proving the feasibility of 3D CNN.

Transformer [24], first proposed for Natural Language Processing, has achieved great success in computer vision tasks. It can encode long-term dependencies and learn efficient

asymmetry feature representation. Vision Transformer [25] introduced pure Transformer into visual tasks for the first time, dividing images into a series of image patches and taking them as input, achieving promising results and demonstrating the effectiveness of Transformer in spatial asymmetry feature representation and image classification. Afterward, Transformer was also applied to various tasks in medical imaging [48]. For example, You et al. [49] used Transformer for 2D medical image segmentation, proposed a generative adversarial model, Castformer, and designed a novel multi-scale Transformer module that uses adversarial training strategies to improve segmentation accuracy. Cao et al. [50] proposed a pure Transformer model similar to Unet for medical image segmentation, Swin-Unet, which jumps and connects labeled images to a U-shaped Encoder-Decoder for local and global semantic feature learning. Chen et al. [51] introduced the hybrid Transformer ConvNet model into volume medical Image registration and proposed the model Trans-Morph. Jeya Maria et al. [52] proposed a gated axial attention model that extends the existing architecture by introducing additional control mechanisms into the self-attention module. The method proposed in this paper combines 3D convolutional neural networks with Transformer to extract 3D features while also learning richer global image features.

## 3. Model Architecture

The model architecture of the proposed algorithm is shown in Figure 1, which contains three main components: a 3D ResNet as the backbone to extract deep features, a Transformer encoder layer, and a Support vector machine classification layer.



Figure 1. Model architecture.

## 3.1. 3D ResNet

Ground-glass lung nodules are stereoscopic image blocks, and their 3D image block features are better in characterizing spatial information and texture information. Different from 2D convolution, 3D convolution can better extract the features of the three dimensions of ground-glass pulmonary nodules and can fully extract the context information in pulmonary nodules, which effectively alleviates the dependence of traditional machine learning methods on manual extraction of features. Therefore, according to the location information of lung nodules marked by doctors, this paper cuts the corresponding size of the stereo block on the CT image and extracts the 3D image block features through the 3D ResNet network.

ResNet has strong robustness, which solves the problem of gradient vanishing when the network depth in the neural network reaches very deep. In addition, given the large demand for training data sets for the Transformer model, this paper uses ResNet as the basis to make the model generalize on general data sets. Therefore, this paper selects the ResNet network as the Backbone. In this paper, the residual module of the original 2D version of the ResNet model is improved, and its convolutional layer and pooling layer are transformed into a 3D version.

When constructing a 3D convolution layer, a set of 3D feature extractors is first established to extract the feature representation of the image, and multiple sets of 3D convolution kernels are used to convolve different input feature quantities to generate new feature volumes. The convolution function operated by the 3D convolutional layer is expressed by the following Equation [12]:

$$y = conv(x, w)$$

where x represents the original data operated by the convolution function, w represents the filter, y represents the output layer of the convolution, and the input x has  $B \times C \times D \times H \times W$  dimensions, where B is the batch size, C is the number of channels, D represents the length of the feature map, that is, the number of slices on the z-axis, and H and W represent the height and width, respectively.

The pooling layer is an important layer running on each feature channel, combining the surrounding feature values into one by appropriate operation. The 3D pooling layer is used to reduce the size of the feature map to reduce the amount of computation while retaining the original image information. The common pooling is Max pooling and average pooling, and the model in this paper uses the Max pooling method.

In this paper, a 3D ResNet with a depth of 18 is used as the feature extraction Backbone. Its detailed network structure is shown in Table 1, where padding is all set to 1, 16 in the Input Size represents the batch size of 16 image blocks of pulmonary nodules that are input at a time.

Table 1. Three-dimensional ResNet network structure.

| Network<br>Composition | Input Size                               | Output Size                              | Layer Structure   |
|------------------------|--|--|---|
| Input layer            | $16\times1\times48\times48\times48$      | $16\times32\times48\times48\times48$     | $[3 \times 3 \times 3$ conv+ BN +ReLU]  |
| Pooling layer          | $16\times32\times48\times48\times48$     | $16\times32\times24\times24\times24$     | $3 \times 3 \times 3$ in stride of 2,<br>Maximum pooling with padding of 1  |
| First layer            | $16\times32\times24\times24\times24$     | $16\times 32\times 24\times 24\times 24$ | $[3 \times 3 \times 3$ conv (in stride of 1) + BN + ReLU] $\times 2$  |
| Second layer           | $16x32\times24\times24\times24$          | $16\times 64\times 12\times 12\times 12$ | $[3 \times 3 \times 3 \text{conv} \text{ (in stride of 2) + BN + ReLU,}$<br>3 × 3 × 3 conv (in stride of 1) + BN + ReLU] × 2                                |
| Third layer            | $16\times 64\times 12\times 12\times 12$ | $16\times 128\times 6\times 6\times 6$   | $[3 \times 3 \times 3 \text{conv} \text{ (in stride of 2) + BN + ReLU,}$<br>3 × 3 × 3 conv (in stride of 1) + BN + ReLU] × 2                                |
| Fourth layer           | $16\times128\times6\times6\times6$       | $16\times 256\times 3\times 3\times 3$   | $ [3 \times 3 \times 3 \text{conv} \text{ (in stride of 2) + BN + ReLU} \\ 3 \times 3 \times 3 \text{conv} \text{ (in stride of 1) + BN + ReLU]} \times 2 $ |

# 3.2. Position Encoding

After extracting features through 3D ResNet, the image is then input into the Transformer Encoder. The input of the Transformer model is a sequence, so position encoding needs to be added to the feature map extracted by 3D ResNet. The purpose of adding position encoding is to preserve the position information of each pixel in the feature map. The position encoding used in this experiment is the sine position encoding method used in Transformer. The calculation formulas are [24]:

$$\begin{aligned} \text{PE}_{(\text{pos},2i)} &= \sin\left(\text{pos}/10000^{2i/d_{\text{model}}}\right) \\ \text{PE}_{(\text{pos},2i+1)} &= \cos\left(\text{pos}/10000^{2i/d_{\text{model}}}\right) \end{aligned}$$

Among them, pos is the position of the graph vector in the sequence, and i is the index of the channel. This paper calculates a position encoding in the three-dimensional directions x, y, and z, calculates the sine of odd positions, calculates the cosine of even

positions, and then converts pos\_x, pos\_y, and pos\_z is concatenated to obtain an array of BHWDC, which is then transformed into BCDHW through permute (0, 4, 3, 1, 2). Therefore, the output and input feature shapes of position encoding are the same.

## 3.3. Transformer Encoder

The universal Transformer module is composed of an encoder-decoder structure, where the encoder converts the input image into a vector representation and outputs it. In the recognition algorithm for ground-glass nodules proposed in this paper, we only use the encoder structure in the Transformer model as a feature extractor and use its self-attention mechanism to obtain global features to assist in the recognition task of ground-glass nodules.

The 3D ResNet processes the ground-glass nodule image blocks, and the size of each image block is converted to a  $256 \times 3 \times 3 \times 3$  feature map. Then, the feature map is transferred into the Transformer encoder for further feature extraction. The Transformer encoder requires a one-dimensional encoding sequence as input, so we segment the feature map in pixels, and the number of pixels obtained is the effective input sequence length of the Transformer encoder. In this paper, the sequence length is 27.

Next, we will introduce the Transformer encoder in detail. The Transformer encoder consists of a Multi-head self-attention and Multilayer Perceptron (MLP) [53]. Each sublayer is preceded by Layer Normalization [54], and each sublayer is followed by residual connection [28], as shown in Figure 2.



Figure 2. Transformer Encoder Structure Diagram.

The use of the Multi-head attention mechanism can combine information learned from different heads, which is beneficial for the model to achieve better results. The multi-head attention mechanism divides the Q, K, and V matrices into multiple parts in the Embedding dimension, each part being a head. Next, calculate each group of attention in parallel. The calculation of attention needs to introduce the concepts of Query, Key, and Value. Key is matched with Query, and Value represents the information extracted from the input. The process of matching Query and Key can be understood as calculating the correlation between the two. The larger the correlation, the larger the weight of the corresponding Value. The attention result is obtained by multiplying the correlation by Value. Finally, the calculation results of each header are concatenated to generate an output value.

Q, K, and V are derived from self-attention and input X<sub>embedding</sub>, generating a Linear map [24]:

 $\begin{array}{lll} Q = & X_{embedding} \ast W_Q \\ K = & X_{embedding} \ast W_K \\ V = & X_{embedding} \ast W_V \end{array}$ 

 $W_Q$ ,  $W_K$ , and  $W_V$  are the parameter matrices of linear mapping, and the dimension is 256 × 256. The dimensions of Q, K, and V is consistent with  $X_{embedding}$ , which is batch × 27 × 256. According to the larger dot product of two vectors, the more similar they are, the attention matrix of each head is calculated in parallel through QK<sup>T</sup>, and then the Value is weighted according to the attention matrix. Finally, the attention results of each head are concatenated to obtain the output of multi-head attention, the formula is as follows [24]:

$$\begin{split} O_i &= Attention(Q_i, K_i, V_i) = softmax(\frac{Q_i K_i^1}{\sqrt{d_k}}) V_i \\ MultiHeadAttention(X_{embedding}) &= Concat[O_1, \ldots, O_h] W_C \end{split}$$

where  $\sqrt{d_k}$  turns the attention matrix into a standard normal distribution, and softmax is normalized to make the sum of attention weights 1,  $O_i$  is the attention calculation result for each head, h is the number of heads, and  $W_O$  is the parameter matrix.

Next, we will introduce the Residual connection and Layer Normalization. The residual connection adds up the inputs and outputs of the previous layer. The Layer Normalization normalizes the hidden layer in the neural network into a standard normal distribution to accelerate convergence. Finally, the Feedforward network uses two layers of Linear map and then passes through the Relu Activation function to get the output.

## 3.4. The SVM Classifier

The ground-glass nodules undergo feature extraction through the upper two layers of the network structure, and the output features are classified using SVM to obtain recognition results. SVM has the characteristic of strong generalization ability for datasets, so this paper uses an SVM classifier. SVM maps data from low dimensional space to high dimensional space through kernel functions and classifies it in high dimensional space. Common kernel functions include linear kernel function, polynomial kernel function, and radial basis kernel function. Due to the advantages of parameter variables, low computational complexity, and high computational efficiency of the radial basis function, we select it as the kernel function of SVM in this paper. Its mathematical expression is as follows [55].

$$K(x_i, x_j) = exp\left[\frac{-\|X - X_i\|^2}{2\sigma^2}\right]$$

Among them,  $\sigma$  is the width of the radial basis kernel function, and C and  $\sigma$  are the important parameters of the SVM model.

In order to obtain the best parameter combination of C and  $\sigma$ , this paper uses a grid search algorithm optimized through K-fold cross-validation to ensure the optimal performance of SVM.

The basic idea of optimizing SVM model parameters through a grid search algorithm [56] is: list all parameter combinations and generate a grid; perform SVM modeling on all parameter combinations in sequence; select the parameter combination with the highest modeling accuracy.

The algorithm steps are as Algorithm 1.

The experimental results indicate that the optimal parameter combination is as follows: C' = 5,  $\sigma' = 10\sqrt{5}$ .

# 3.5. ELM-KNN Classifier

The ground-glass nodules are input into the ELM-KNN [57] model for classification after feature extraction by the Transformer. The Extreme Learning Machine (ELM) [58] is used to map input features from low-dimensional space to high-dimensional space, and the K-Nearest Neighbor (KNN) algorithm [59] is used to classify the mapped features. For the specific content of the ELM-KNN model, please refer to the literature [57]. Through cross-validation, parameters K of KNN in this study are selected as 10, weights as 'distance', and P as 4 for the experiment.

# 4. Experimental Design and Result Analysis

# 4.1. Dataset and Preprocessing

This paper adopts a private lung CT dataset for experiments in Sections 4.2–4.5. and a public lung CT dataset of LUNA16 [30] for experiments in Section 4.6, respectively.

The first dataset contains a total of 500 lung CT images and was labeled by five experienced radiologists. There are 369 ground-glass pulmonary nodules, 1559 non-ground-glass pulmonary nodules, and 488 uncertain types in this labeled dataset. In order to make the experimental data set balanced, this paper selected 369 non-ground-glass pulmonary nodules and 369 ground-glass pulmonary nodules, totaling 738 pulmonary nodule samples for experiments.

The second data set of LUNA16 contains a total of 888 lung CT images, which are illustrated in Figure 3. In order to balance the positive and negative samples, we randomly selected 1320 benign nodule samples and 1320 malignant nodule samples, giving a total of 2640 samples for experiments.



Figure 3. Examples of lung CT images.

The above two experimental datasets, are all divided into a training set and a testing set with the ratio of 80% vs. 20%, respectively.

For deep learning models, a larger amount of data may be required to obtain good results. The dataset used in this paper may not have enough data for the deep learning model. In order to enhance the learning ability of the network and reduce the occurrence of overfitting of the network model, this paper uses the following data augmentation processing methods:

- (1) 3D RandomRotation: For the original input data, rotate it 180 degrees by randomly fixing one of its edges.
- (2) 3D RandomShift: For the input 3D image, use random displacement to move it in any direction along the x, y, and z axes, with a range of 10% of the input image size.
- (3) Gaussian Noise: In order to avoid overfitting and enhance learning ability by adding an appropriate amount of noise, this experiment sets its mean (offset) to 0.2 and sigma (standard deviation) to 0.3.
- (4) 3D RandomScale: The ability to achieve internal and external image scaling. When scaling outward, the resulting image size will be larger than the original, but when scaling inward, it will cause its size to decrease. In this experiment, the scaling factor is set to be between 0.8 and 1.1.
- (5) 3D RandomFlip: For the original input image, flip it 180 degrees using a flip matrix.

This paper randomly executes one of the above data augmentation methods during each training round, and the quantity of expanded samples are ten times that of original samples.

#### 4.2. Model Parameter Settings and Experimental Environments

The training data for the model in this paper is the lung nodule three-dimensional image block, whose size is  $1 \times 48 \times 48 \times 48$ . The lung nodule recognition algorithm uses 3D ResNet and Transformer to extract features from the image. The depth selection of the 3D ResNet model is 18. The image size is  $256 \times 3 \times 3 \times 3$  after 3D convolution, and a 3D position code is added to the image to record the position information. Then, the image is flattened into 27 patches, and the size of patches is  $1 \times 1 \times 1$ . The Transformer encoder takes a sequence of  $27 \times 256$  vectors as input, where the number of heads parameter h of the multi-head attention mechanism is 8, as shown in Figure 4.



Figure 4. Input pulmonary nodule images into 3D ResNet and Transformer Encoder for feature extraction.

In this section, the experiments in this paper involve two environments:

- Vgg19 model: AMD EPYC 7543 32 CPU; two NVIDIA A40 graphics cards, each card contains 48GB graphics memory; Ubuntu 18.04 operating system; Pytorch 1.9.1 deep learning framework; Python 3.8 programming language.
- (2) Other models: Intel Core i7-7820X CPU; one NVIDIA GeForce RTX 2080 Ti graphics card with 11GB graphics memory; Windows operating system; Pytorch 1.9.1 deep learning framework; Python 3.7 programming language and integrated development tool PyCharm.

All the comparison models are consistent with the settings of the proposed model. The number of epochs of each data in the training sample is 100, the learning rate is 0.0001, and the batch size is 16.

# 4.3. Selection of Transformer Encoder Layers

For the number of Transformer encoder layers, it is not the case that the more layers there are, the better the classification effect. In order to select the optimal number of layers, this paper carried out experiments for the different number of layers (1, 2, 3, 4, 5) in the same experimental environment, and obtained the recognition accuracy of ground-glass nodules under the different numbers of layers. The results are shown in Figure 5. Experiments show that when the encoder layer comprises two layers, the accuracy is the highest, so the number of layers is two in this paper.



Figure 5. Selection of Transformer Encoder Layers.

#### 4.4. Classifier Selection

In order to obtain better recognition results of ground-glass nodules, this paper uses six classifier models: Fully Connected Layer, Decision Tree [60], Random Forest [61], AdaBoost [62], SVM, and ELM-KNN [57] in the selection of classifiers, and optimizes model parameters using a grid search algorithm. The experiment is carried out in the same experimental environment, and the results are shown in Table 2.

| Classifier             | Accuracy |
|------------------------|----------|
| Fully Connected Layers | 0.9315   |
| Decision Tree          | 0.9315   |
| Random forest          | 0.9383   |
| AdaBoost               | 0.9520   |
| SVM                    | 0.9589   |
| ELM-KNN                | 0.9589   |

Table 2. Model accuracy for different classifiers.

From Table 2, we find that the classification effect of ELM-KNN and SVM is 0.9589, which is 2.74 percentage points higher than that of using the Full Connection Layer. Compared with the classifier using the Decision Tree, the accuracy is improved by 2.74 percentage points. Compared with the classifier using Random Forest, the accuracy is improved by 2.06 percentage points. The accuracy of these two classifiers is 0.69 percentage points higher than that of the classifier using AdaBoost. We can see that using SVM or ELM-KNN as the classifier has the same good classification effect, so we chose these two classifiers to do a comparison experiment on the two databases.

## 4.5. Comparative Experiments

In order to verify the superiority of the proposed method, this paper conducts comparative experiments with the commonly used deep learning models ResNet18 [28], AlexNet, DenseNet121 [63], GoogLeNet, and VGG19 [18] in the same experimental environment, on the same dataset, and with the same data preprocessing. All the network models employ 3D convolution. In addition, this paper also uses ResNet34 [28] as a feature extraction network and only applies the Transformer model to complete the recognition task of ground-glass nodules. Finally, the models ResNet and GBM and Attention [64], MSMA-Net [65], and MVCSNet [66] proposed by the latest papers in the field of lung nodule classification are used to perform the recognition task of ground-glass nodules in the same experimental environment. The above experimental results are shown in Table 3.

Table 3. Performance comparison of different network models.

| Deep Network Model               | Accuracy | F-Measure |
|----------------------------------|----------|-----------|
| ResNet18                         | 0.9110   | 0.9128    |
| AlexNet                          | 0.9110   | 0.9103    |
| DenseNet121                      | 0.9384   | 0.9379    |
| GoogLeNet                        | 0.9178   | 0.9167    |
| VGG19                            | 0.9315   | 0.9315    |
| ResNet34 + Transformer           | 0.9452   | 0.9452    |
| Transformer                      | 0.9110   | 0.9127    |
| ResNet + GBM + Attention         | 0.9384   | 0.9404    |
| MSMA-Net                         | 0.9384   | 0.9388    |
| MVCSNet                          | 0.9521   | 0.9517    |
| ResNet18 + Transformer + SVM     | 0.9589   | 0.9600    |
| ResNet18 + Transformer + ELM-KNN | 0.9589   | 0.9594    |

As can be seen from Table 3, when the classifier is SVM, the accuracy and F-measure of the proposed method are 95.89% and 96.00%, respectively, which are 4.79 percentage points and 4.72 percentage points higher than that of ResNet18. From this data, it can be seen that adding the Transformer encoder layer to the model can improve the accuracy of recognition. When only the Transformer model is applied to extract features in this paper, the accuracy is 91.10%, which is lower than the method proposed in this paper. The reason is that when an experiment only uses Transformer, it needs a large amount of data to obtain a good learning effect. In addition, in the experiment, when this paper used a deeper ResNet34 as the feature extraction network, the accuracy and F-measure are reduced by 1.37 percentage

points and 1.48 percentage points. This result shows that when using a network with strong expressive power such as Transformer for modeling, the feature extraction network can use a lighter structure to balance effect and performance. Finally, compared with the common deep learning models of AlexNet, DenseNet121, GoogLeNet, and VGG19, the accuracy of this paper is improved by 4.79, 2.05, 4.11, and 2.74 percentage points, respectively, and the F-measures is improved by 4.97, 2.21, 4.33, and 2.85 percentage points, respectively. Compared with the models proposed in the latest paper in the field of nodule classification, the accuracy is improved by 2.05, 2.05, and 0.68 percentage points, respectively, and F-measure is improved by 1.96, 2.12, and 0.83 percentage points, respectively, which fully proves the superiority of the proposed model in ground-glass nodule recognition.

# 4.6. Experimental Results on the Luna16 Dataset

In order to verify the performance of the model in this paper on the public dataset, we used the dataset Luna16 for experiments. In this paper, the LUNA16 dataset is used for the classification of benign and malignant nodules. According to the nodule information provided by the LUNA16 dataset, we can find the malignant score given by the physician from the corresponding LIDC-IDRI [67] labeling file. In order to balance the positive and negative samples, we randomly selected 1320 benign samples and 1320 malignant samples, giving a total of 2640. In addition, the same data augmentation processing as the dataset in Section 4.1 is applied.

In terms of parameter setting, the Batchsize of model training is 16, the Epoch is 100, the learning rate is 0.0001, and the number of Transformer encoder layers is 5.

In this paper, we select the models ResNet and GBM and Attention [64], DPNCapsNet [68], MVF-CNN 25 [69] and MVCSNet [66] proposed in the latest papers on the Luna16 dataset for comparison, and the experimental results are shown in Table 4.

| Deep Network Model               | Accuracy | F-Measure |
|----------------------------------|----------|-----------|
| ResNet + GBM + Attention         | 0.9130   | N/A       |
| DPNCapsNet                       | 0.9156   | N/A       |
| MVF-CNN 25                       | 0.928    | N/A       |
| MVCSNet                          | 0.9239   | 0.9090    |
| ResNet18 + Transformer + SVM     | 0.9572   | 0.9567    |
| ResNet18 + Transformer + ELM-KNN | 0.9633   | 0.9633    |

Table 4. Performance comparison of different network models on LUNA16 dataset.

It can be seen from Table 4 when the classifier is SVM, the accuracy of the proposed method is higher than that of the above papers, and the accuracy is increased by 4.42, 4.16, 2.92, and 3.33 percentage points, respectively, which fully illustrates the effectiveness and generalization of the ground-glass nodule recognition model based on 3D ResNet and Transformer proposed in this paper on the LUNA16 public dataset.

# 5. Conclusions

In this paper, a ground-glass nodule recognition model based on Transformer from the view of global 3D asymmetry feature representation is proposed. Given the traditional machine learning methods relying on artificial experience to extract features, this paper uses the deep learning model with powerful feature extraction ability to represent asymmetry features. At the same time, considering that the medical image is a 3D block image, this paper uses the 3D convolution method to make the neural network better extract 3D medical image features. Aiming at the shortcomings of CNN in extracting global image features, this paper refers to the Transformer encoder, so that the model can learn richer global 3D asymmetry features and enhance the feature expression ability of the model. In the selection of the number of Transformer encoder layers, this paper conducts experiments on the number of Transformer encoder layers (1, 2, 3, 4, 5). The experimental results show that when the number of Transformer encoder layers is 2, the proposed model achieves the highest accuracy of 95.89%. Finally, the method proposed in this paper is compared with other deep learning models and the models of the latest papers in the field of lung nodules. The experimental results show that the transformer-based method proposed in this paper is superior to other recognition models, and the recognition accuracy reaches 95.89%, which verifies the effectiveness of the method in ground-glass nodule recognition.

Although the Transformer-based recognition method proposed in this paper shows good performance, there are some limitations which are also our future research directions:

- (1) The model proposed in this paper only uses sinusoidal position encoding in the position encoding section. In the future, more position encoding methods can be attempted to enable the model to obtain richer image information.
- (2) The number of samples used in this paper is limited. In subsequent studies, the number of samples can be increased to improve the accuracy and generalization performance of ground-glass pulmonary nodule recognition.
- (3) The main content of this paper is to study the classification of ground-glass pulmonary nodules. The samples used are already detected nodule areas. Subsequent research can include the detection of pulmonary nodules and visualize the test results of the detection section.

**Author Contributions:** Conceptualization, J.M. and Y.C.; methodology, J.M. and Y.C.; software, Y.C.; validation, M.Z.; formal analysis, Y.C.; resources, J.M. and Y.C.; visualization, M.Z.; supervision, J.M. and Y.Q.; project administration, J.M. and Y.Q.; funding acquisition, J.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported in part by the Beijing Natural Science Foundation (No. 4202025), Tianjin Anjian IoT Technology Enterprise Key Laboratory Research Project (No. VTJ-OT20230209-2) and Guizhou Provincial Sci-Tech project (Qiankehe foundation zk[2022] general 012).

**Institutional Review Board Statement:** This is a retrospective study which used a subset of existing data used in Ref. [70].

**Informed Consent Statement:** This is a retrospective study and that the private dataset used in this paper is a subset of the dataset used in Ref. [70]. The informed consent is not required for use of the original dataset as the risk of the subjects does not exceed the minimum.

**Data Availability Statement:** The private dataset used in this research is a subset of the original dataset used in Ref. [70], and this research is supported by the project where the original dataset was generated. Due to the nature of the original research, participants of this study did not agree for their data to be shared publicly, so supporting data is not available. The Luna16 dataset is publicly available in a public repository.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- 1. Qiu, L.; Ding, S.; Zhang, S.; Shen, M.; Ruan, Y. The effect of astaxanthin on human lung cancer A549 in nude mice. *Chin. J. New Drugs* 2019, *28*, 1477–1483. (In Chinese)
- Chen, T.; Li, S. Diagnostic value of different reconstruction methods of MSCT for pulmonary ground glass nodules within 3 cm. J. Med. Imaging 2019, 29, 1123–1127. (In Chinese)
- Zhang, H.; Zhou, W. CT signs and clinical diagnostic value of benign and malignant solitary focal ground glass density pulmonary nodules. J. Hebei Med. Univ. 2016, 37, 1458–1461. (In Chinese)
- Feng, M. Application of computer-aided CT image features in the diagnosis of early lung cancer with ground glass nodules. J. Imaging Res. Med. Appl. 2020, 4, 90–91. (In Chinese)
- Jiang, X.; Jiang, T.; Sun, J.; Song, J.; Jiang, W.; Ai, H.; Long, Z.; Su, J.; Chang, S.; Yu, T. Application of deep learning AI technology in Medical imaging aided analysis. *China Med. Devices* 2021, *36*, 164–171. (In Chinese)
- Xu, Y.; Guan, X. The Application of Computer Aided Diagnosis System in Teaching Pulmonary Nodules. *Chin. Foreign Med. Res.* 2021, 19, 185–188. (In Chinese) [CrossRef]
- 7. Chen, L.; Zhou, Y.; Xu, S. Breast mass detection based on multi spectral channel attention and two-way feature fusion. *J. South-Cent. Univ. Natl.* (*Nat. Sci. Ed.*) **2023**, *42*, 111–119. (In Chinese) [CrossRef]

- Zhao, K.; Qiu, H.; Li, X.; Xu, Z. Real time lung nodule detection algorithm combining attention and multipath fusion. *J. Comput. Appl.* 2023, 43, 1–11. Available online: http://kns.cnki.net/kcms/detail/51.1307.TP.20230731.1346.005.html (accessed on 6 September 2023). (In Chinese)
- 9. Yu, R.; Yu, H.; Wan, H. Classification of brain network features in schizophrenia based on resting statefunctional magnetic resonance imaging. *J. Biomed. Eng.* 2020, *37*, 661–669. (In Chinese)
- 10. Hu, J.; Wang, H. Malaria recognition algorithm based on unsupervised sample relationship embedding model. *J. Baoji Univ. Arts Sci. (Nat. Sci. Ed.)* **2022**, *42*, 30–39. (In Chinese) [CrossRef]
- 11. Gao, Y. Segmentation and Classification of Benign and Malignant Pulmonary Nodules Based on Deep Learning; Hebei Normal University: Shijiazhuang, China, 2020. (In Chinese) [CrossRef]
- 12. LeCun, Y.; Boser, B.; Denker, J.S.; Henderson, D.; Howard, R.E.; Hubbard, W.; Jackel, L.D. Backpropagation applied to handwritten zip code recognition. *Neural Comput.* **1989**, *1*, 541–551. [CrossRef]
- Gao, F.; Zhang, S. A deep learning classification method for pulmonary nodules based on fused prior knowledge. *China Med. Devices* 2021, *36*, 54–57+70. (In Chinese)
- Liu, Y.; Zhang, X.; Zhang, J.; Zhou, Z.; Feng, Y.; Chen, W. DenseNet centrop: A Convolutional Network for Classification of Pulmonary Nodules. J. Zhe Jiang Univ. (Sci. Ed.) 2020, 47, 20–26. (In Chinese)
- 15. Zhang, J.; Zhang, X. Multi branch convolutional neural network classification method for pulmonary nodules and its interpretability. *Comput. Sci.* 2020, 47, 135–140. (In Chinese)
- 16. Liu, Y.; Zhao, D.; Liu, J. 3D pulmonary nodule recognition based on K-L transform and support vector machine. *J. Northeast. Univ.* (*Nat. Sci.*) **2009**, *30*, 1249–1252. (In Chinese)
- 17. Wang, W.; Wang, Z.; Xu, Q.; Sun, H. Classification of pulmonary nodules based on three-dimensional convolutional neural networks. *J. Harbin Univ. Sci. Technol.* **2021**, *26*, 87–93. (In Chinese) [CrossRef]
- 18. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv 2014, arXiv:1409.1556.
- Raunak, D.; Lu, Z.; Hong, Y. Diagnostic classification of lung nodules using 3D neural networks. In Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging, Washington, DC, USA, 4–7 April 2018; Volume 38, pp. 774–778.
- Wei, L.; Zhao, X.; Duan, Y.; Liu, L.; Huang, Q. Classification of pulmonary nodules based on CNN multi-level second-order feature fusion. J. Front. Comput. Sci. Technol. 2020, 14, 1590–1601.
- Halder, A.; Chatterjee, S.; Dey, D. Adaptive morphology aided 2-pathway convolutional neural network for lung nodule classification. *Biomed. Signal Process. Control.* 2022, 72, 103347. [CrossRef]
- 22. Wang, Y.; Zhang, H.; Chae, K.J.; Choi, Y.; Jin, G.Y.; Ko, S.B. Novel convolutional neural network architecture for improved pulmonary nodule classification on computed tomography. *Multidimens. Syst. Signal Process.* **2020**, *31*, 1163–1183. [CrossRef]
- Zhao, C.; Qin, B.; Feng, S.; Zhu, W.; Sun, W.; Li, W.; Jia, X. Hyperspectral Image Classification with Multi-attention Transformer and Adaptive Superpixel Segmentation-based Active Learning. *IEEE Trans. Image Process.* 2023, 32, 3606–3621. [CrossRef]
- 24. Vaswani, A.; Shazeer, N. Attention Is All You Need. arXiv 2017, arXiv:1706.03762.
- 25. Dosovitskiy, A.; Beyer, L. An Image is Worth 16 × 16 Words: Transformers for Image Recognition at Scale. *arXiv* 2020, arXiv:2010.11929.
- Ji, W.; Liu, Q.; Huang, C.; Yang, R.; Huang, H.; Xu, G. YOLOX traffic sign detection based on Swin Transformer. *Radio Commun. Technol.* 2023, 49, 547–555. (In Chinese)
- 27. Wang, D. Research on lung cancer detection methods based on deep learning. Digit. Technol. Appl. 2020, 38, 85–89. (In Chinese)
- 28. He, K.; Zhang, X. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
- Hearst, M.A.; Dumais, S.T.; Osuna, E.; Platt, J.; Scholkopf, B. Support vector machines. *IEEE Intell. Syst. Their Appl.* 1998, 13, 18–28. [CrossRef]
- Setio, A.; Traverso, A. Validation, Comparison, and Combination of Algorithms for Automatic Detection of Pulmonary Nodules in Computed Tomography Images: The LUNA16 Challenge; Elsevier: Amsterdam, The Netherlands, 2017.
- 31. Qiao, H. Research Progress and Development Trend of Modern Medical imaging. China Health Ind. 2017, 14, 189–190. (In Chinese)
- 32. Cai, Y.; Zhou, X.; Zhang, Z.; Li, J.; Min, R.; Han, D.; Fan, M. The value of CT histogram analysis in distinguishing benign and malignant pure ground glass nodules of the lung. *Radiol. Pract.* **2020**, *35*, 949–952. (In Chinese)
- Hu, X.; Gong, J.; Zhou, W.; Li, H.; Wang, S.; Wei, M.; Peng, W.; Gu, Y. Computer-aided diagnosis of ground glass pulmonary nodule by fusing deep learning and radiomics features. *Phys. Med. Biol.* 2021, 66, 065015. [CrossRef]
- 34. Li, W.; Cao, P.; Zhao, D.; Wang, J. Pulmonary nodule classification with deep convolutional neural networks on computed tomography images. *Comput. Math. Methods Med.* **2016**, 2016, 6215085. [CrossRef]
- 35. Ni, Y.; Yang, Y.; Zheng, D.; Xie, Z.; Huang, H.; Wang, W. The invasiveness classification of ground-glass nodules using 3D attention network and HRCT. *J. Digit. Imaging* **2020**, *33*, 1144–1154. [CrossRef] [PubMed]
- Krizhevsky, A.; Sutskever, I.C.D. ImageNet Classification with Deep Convolutional Neural Networks. Adv. Neural Inf. Process. Syst. 2012, 25, 1097–1105. [CrossRef]
- Szegedy, C.; Liu, W. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.

- 38. Zhang, Z.; Gao, J.; Yan, L. Research on abnormal disturbance of power quality based on improved Convolutional neural network. *Microcomput. Appl.* **2023**, *39*, 135–139. (In Chinese)
- 39. Dai, G.; Zhou, J.; Huang, J.; Wang, N. HS-CNN: A CNN with hybrid convolution scale for EEG motor imagery classification. *J. Neural Eng.* **2020**, *17*, 016025. [CrossRef] [PubMed]
- 40. Lei, X.; Pan, H.; Huang, X. A dilated CNN model for image classification. IEEE Access 2019, 7, 124087–124095. [CrossRef]
- Sharma, H.; Jain, J.S.; Bansal, P.; Gupta, S. Feature Extraction and Classification of Chest X-Ray Images Using CNN to Detect Pneumonia. In Proceedings of the 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 29–31 January 2020; pp. 227–231. [CrossRef]
- 42. Wang, J. Research on V2X Early Warning Information Display Algorithm and System Implementation Based on Augmented Reality; Jilin University: Jilin, China, 2022. (In Chinese) [CrossRef]
- 43. Cao, P.; Sheng, Q.; Pan, Q.; Ning, G.; Wang, Z.; Fang, L. Segmentation method of Hippocampus combin-ng sequence learning and U-shaped network. *J. Comput.-Aided Des. Comput. Graph.* **2019**, *31*, 1382–1390. (In Chinese)
- 44. Wu, B.; Qiang, Y. Classification of pulmonary nodules based on multi-dimensional Convolutional neural network. *Comput. Eng. Appl.* **2019**, *55*, 171–177. (In Chinese)
- Shen, W.; Zhou, M. Multi-scale convolutional neural networks for lung nodule classification. In Proceedings of the International Conference on Information Processing in Medical Imaging, Isle of Skye, UK, 28 June–3 July 2015; Springer: Cham, Switzerland, 2015; pp. 588–599.
- 46. Zhang, G.; Lin, L. Lung Nodule Classification in CT Images Using 3D DenseNet. J. Phys. Conf. Ser. 2021, 1827, 012155. [CrossRef]
- 47. Yuan, Z.G.; Bi, L.X. Survey on Medical Image Computer Aided Detection and Diagnosis Systems. J. Softw. 2018, 29, 1471–1514.
- Li, J.; Li, Y. Research on Coronary Artery Segmentation Method Based on Semi supervised Collaborative Training. *Chin. J. Stereol. Image Anal.* 2023, 28, 77–85. (In Chinese) [CrossRef]
- Chenyu, Y.; Zhao, R.; Liu, F.; Chinchali, S.; Topcu, U.; Staib, L.; Duncan, J.S. Class-Aware Generative Adversarial Transformers for Medical Image Segmentation. *arXiv* 2022, arXiv:2201.10737.
- 50. Cao, H.; Wang, Y.; Chen, J.; Jiang, D.; Zhang, X.; Tian, Q.; Wang, M. Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation. *arXiv* 2021, arXiv:2105.05537. [CrossRef]
- 51. Chen, J.; Frey, E.C.; He, Y.; Segars, W.P.; Li, Y.; Du, Y. TransMorph: Transformer for unsupervised medical image registration. *arXiv* 2021, arXiv:2111.10480. [CrossRef] [PubMed]
- 52. Valanarasu, J.M.J.; Oza, P.; Hacihaliloglu, I.; Patel, V. Medical transformer: Gated axial-attention for medical image segmentation. In Proceedings of the Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, 27 September–1 October 2021; Part I 24. Springer International Publishing: Berlin/Heidelberg, Germany, 2021; pp. 36–46.
- 53. Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; Bengio, Y. Graph attention networks. arXiv 2017, arXiv:1710.10903.
- 54. Ba, J.L.; Kiros, J.R.; Hinton, G.E. Layer normalization. arXiv 2016, arXiv:1607.06450.
- 55. Powell, M.J.D. Radial basis function for multivariable approximations: A review. In Proceedings of the IMA Conference on Algorithms for the Approximation of Functions and Data, Shrivenham, UK, July 1985; pp. 143–167.
- 56. Dallimore, P.J. HAUSER-A Computer Code for the Calculation of Compound Nucleus Cross Sections Using the Hauser-Feshbach Theory; Australian National University: Canberra, Australia, 1970.
- 57. Li, P.; Liu, Z.; Anduv, B.; Zhu, X.; Jin, X.; Du, Z. Diagnosis for multiple faults of chiller using ELM-KNN model enhanced by multi-label learning and specific feature combinations. *Build. Environ.* **2022**, 214, 108904. [CrossRef]
- 58. Huang, G.B.; Zhou, H.; Ding, X.; Zhang, R. Extreme learning machine for regression and multiclass classification. *IEEE Trans. Syst. Man Cybern. Part B (Cybern.)* 2011, 42, 513–529. [CrossRef] [PubMed]
- 59. Cover, T.; Hart, P. Nearest neighbor pattern classification. IEEE Trans. Inf. Theor. 1967, 13, 21–27. [CrossRef]
- 60. Quinlan, J.R. Simplifying decision trees. Int. J. Man-Mach. Stud. 1987, 27, 221–234. [CrossRef]
- 61. Breiman, L. Random forests. *Mach. Learn.* 2001, 45, 5–32. [CrossRef]
- 62. Rätsch, G.; Onoda, T.; Müller, K.R. Regularizing adaboost. Adv. Neural Inf. Process. Syst. 1998, 11, 564–570.
- 63. Gao, H.; Zhuang, L.; van der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
- 64. Kuang, J.; Hong, M. Classification of pulmonary nodules based on attention mechanism. *Comput. Appl. Softw.* **2022**, 39, 163–167. (In Chinese)
- 65. Guo, F.; Huang, M. Classification of benign and malignant pulmonary nodules based on multi model fusion method. *J. Optoelectron. Laser* **2021**, *32*, 389–394. (In Chinese)
- 66. Zhu, Q.; Wang, Y.; Chu, X.; Yang, X.; Zhong, W. Multi-View Coupled Self-Attention Network for Pulmonary Nodules Classification. In Proceedings of the Asian Conference on Computer Vision, Macau SAR, China, 4–8 December 2022; pp. 995–1009.
- 67. Messay, T.; Hardie, R.C. Segmentation of pulmonary nodules in computed tomography using a regression neural network approach and its application to the Lung Image Database Consortium and Image Database Resource Initiative dataset. *Med. Image Anal.* **2015**, *22*, 48–62. [CrossRef]
- 68. Li, C. Classification Model of Lung Nodules Combining DPN Network and Capsule Network; Jilin University: Jilin, China, 2020. (In Chinese)

- 69. Tang, N.; Wei, Z. Comparative study on classification of pulmonary nodules based on multi-scale multi-mode images. *Comput. Eng. Appl.* **2020**, *56*, 165–175. (In Chinese)
- 70. Han, Y.; Qi, H.; Wang, L.; Chen, C.; Miao, J.; Xu, H.; Wang, Z.; Guo, Z.; Xu, Q.; Lin, Q.; et al. Pulmonary nodules detection assistant platform: An effective computer aided system for early pulmonary nodules detection in physical examination. *Comput. Methods Programs Biomed.* 2022, 217, 106680. [CrossRef]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.