applied sciences

MDPI

Article

Adaptive Aggregated Attention Network for Pulmonary Nodule Classification

Kai Xia, Jianning Chi, Yuan Gao, Yang Jiang and Chengdong Wu *

Faculty of Robot Science and Engineering, Northeastern University, No. 195, Chuangxin Road, Shenyang 110169, China; xiakai@stumail.neu.edu.cn (K.X.); chijianning@mail.neu.edu.cn (J.C.); gaoyuan@stumail.neu.edu.cn (Y.G.); jiangyang@mail.neu.edu.cn (Y.J.)
* Correspondence: wuchengdong@mail.neu.edu.cn; Tel.: +86-24-83688325

Abstract: Lung cancer has one of the highest cancer mortality rates in the world and threatens people's health. Timely and accurate diagnosis can greatly reduce the number of deaths. Therefore, an accurate diagnosis system is extremely important. The existing methods have achieved significant performances on lung cancer diagnosis, but they are insufficient in fine-grained representations. In this paper, we propose a novel attentive method to differentiate malignant and benign pulmonary nodules. Firstly, the residual attention network (RAN) and squeeze-and-excitation network (SEN) were utilized to extract spatial and contextual features. Secondly, a novel multi-scale attention network (MSAN) was proposed to capture multi-scale attention features automatically, and the MSAN integrated the advantages of the spatial attention mechanism and contextual attention mechanism, which are very important for capturing the salient features of nodules. Finally, the gradient boosting machine (GBM) algorithm was used to differentiate malignant and benign nodules. We conducted a series of experiments on the Lung Image Database Consortium image collection (LIDC-IDRI) database, achieving an accuracy of 91.9%, a sensitivity of 91.3%, a false positive rate of 8.0%, and an F1-score of 91.0%. The experimental results demonstrate that our proposed method outperforms the state-of-the-art methods with respect to accuracy, false positive rate, and F1-Score.

Keywords: lung cancer diagnosis; deep learning; 3D dual path network; attention network

Citation: Xia, K.; Chi, J.; Gao, Y.; Jiang, Y.; Wu, C. Adaptive Aggregated Attention Network for Pulmonary Nodule Classification. *Appl. Sci.* **2021**, *11*, 610. https://doi.org/10.3390/app11020610

Received: 2 December 2020
Accepted: 8 January 2021
Published: 10 January 2021

1. Introduction

Lung cancer is one of the most lethal cancers in the world, posing a threat to people's health [1]. Lung cancer mortality can be significantly reduced through early diagnosis and screening [2]. In the 1960s, computer-aided diagnosis (CAD) was proposed and used to diagnose lung cancer, which relieved the pressure on doctors and helped them to diagnose cases more accurately [3].

Traditionally, researchers manually extracted hand-crafted features and designed classifiers, but designing hand-crafted features was time-consuming and required professional medical knowledge. The effectiveness of feature extraction depends on doctors' expertise in lung cancer diagnosis and their understanding of traditional machine learning methods. Moreover, hand-crafted features were subjective and their generalization was poor. Therefore, it is very important to develop an algorithm for automatic feature extraction.

In recent years, end-to-end network has become very popular, and it can automatically extract deep features and learn the internal salient features through iterations. Architectural convolution neural network (CNN) has gained huge success in lung nodule classification. In the early stages, 2D CNN was mainly used for pulmonary nodule classification. Two-dimensional CNN could effectively extract features within an image but could not extract the features between the images which are important for lung nodule classification. In recent years, 3D CNN has led to better performances than 2D CNN [4]. Three-dimensional CNN can extract spatial features of nodules, which are very important in classifying nodules. Although the existing deep learning methods have made great progress in the

diagnosis of lung nodules, there are still deficiencies in feature extraction. Due to the small lung cancer dataset, it is difficult to extract rich and useful features. Researchers usually deepen the network depth and width to extract rich features, but networks are prone to overfitting. Moreover, not all of the extracted features are useful and some invalid features will reduce the performance of the model. Therefore, based on the small lung cancer dataset, it is very important to design a novel method which can extract rich and useful fine-grained features automatically, and this method should be able to avoid overfitting as much as possible, so that the robustness of the method can be guaranteed.

Attention networks are often effective for capturing fine-grained features. An attention mechanism plays an important role in image classification, and it can make networks capture useful attention features automatically. However, if the attention mechanism is not designed or applied properly, the attention mechanism will filter out important features and focus on some unimportant features instead, which will lead to performance degradation of the model. Although some attention methods have been successfully applied in the field of lung cancer diagnosis, there are still deficiencies in some aspects, such as the multi-scale problem between large nodules and small nodules, and the multi-resolution problem of different nodules.

To address these challenges, firstly, we used a 3D dual path network (DPN) as our backbone network, which aggregated the advantages of residual connection and dense connection. Based on the dense connection of a DPN, rich features were extracted and the residual connection avoided overfitting. The backbone network ensured the overall performance of our model. Secondly, we applied the squeeze-and-excitation module (SEN) [5] and residual-attention network (RAN) [6] to extract contextual features and spatial features, and the attention modules were useful to make the network focus on important features and to filter out redundant features. RAN is effective in solving multi-resolution problems because of its bottom-up and top-down structure. Especially, RAN was the first applied method in terms of pulmonary nodule classification. Moreover, we designed a multi-scale attention module to automatically obtain multi-scale nodule features, including sparse features and dense features, and the multi-scale attention network (MSAN) module can concatenate the spatial features and contextual features effectively. The attention mechanism can automatically concentrate on essential regions and can extract key features for pulmonary nodule classification. Finally, the gradient boosting machine (GBM) algorithm [7] was utilized to differentiate benign and malignant nodules.

To verify our model, we conducted experiments on the Lung Image Database Consortium image collection (LIDC-IDRI) dataset. The experimental results prove that our attention mechanisms are effective for improving nodule classification. At last, our proposed framework significantly outperformed previous state-of-the-art methods. Our main contributions are as follows: (1) We applied residual attention and squeeze-and-excitation module to automatically concentrate on regions which are essential for nodule classification and the residual structure ensures the integrity of the features, which can prevent overfitting to some extent and can improve the classification performance. (2) We propose a Multi-scale attention mechanism to automatically obtain multi-scale salient features of nodules which can better extract the features of different scales of nodules and can fuse sparse features and dense features effectively.

The rest of the paper is organized as follows. Section 2 discusses previous related works in medical image analysis and pulmonary nodules classification. Section 3 describes the data preprocessing and the proposed framework in detail. The experiments results and discussions are described in Section 4. Finally, we conclude our work in Section 5.

2. Related Works

Traditionally, researchers extracted hand-crafted features based on their professional recognition in medical imaging and pulmonary nodules and then designed a classifier to classify the extracted hand-crafted features. The hand-crafted features usually included texture, shape, density, and sphericity features, and the classifiers usually included Support

Vector Machine (SVM), Random Forest, and K-nearest classifier [8–10]. Carvalho et al. [11] utilized image preprocessing and pattern recognition methods to differentiate malignant and benign nodules. To classify nodules, they used phylogenetic diversity by means of particular indexes, which were intensive quadratic entropy, extensive quadratic entropy, average taxonomic distinctness, and so on. Moreover, they utilized a genetic algorithm to select the best model. Costa et al. [12] used the mean phylogenetic distance and taxonomic diversity index to extract pulmonary texture features. After obtaining the texture features, they combined a genetic algorithm and SVM to select the best training model. Finally, their proposed model was tested on the LIDC/IDRI dataset.

For traditional pattern recognition methods, manual feature extraction is required, and feature extraction is not easy. A professional radiologist uses a lot of effort to extract the relevant features, and the effectiveness of feature extraction depends on the doctor's expertise in lung cancer diagnosis and on their understanding of traditional machine learning methods. Moreover, the feature extraction process is highly subjective and time-consuming. Therefore, it is very important to develop an algorithm for automatic feature extraction.

With the development of deep learning, many researchers utilize deep learning methods to differentiate pulmonary nodules. The common classification network is convolutional neural network (CNN). To classify pulmonary nodules better, researchers usually improve the feature extractor and feature classifier of a CNN network. Wei et al. [13] proposed a Multi-crop convolution neural network (MC-CNN) to automatically extract nodule salient features. They cropped feature maps with different size and concatenated them. The multi-scale feature maps contained more semantic information, and this strategy obtained a good performance on nodule classification. To utilize spatial 3D context information, Liu et al. [14] proposed an end-to-end architecture named dense convolutional binary-tree network to classify nodules. They introduced center-crop operation into DenseNet and utilized separation and fusion operations on transition layers which were useful to enrich multi-scale features. For the above existing deep learning methods on lung cancer diagnosis, the network structure is usually relatively complex and large. Due to the limitation of medical dataset, networks are prone to overfitting and they are insufficient in capturing fine-grained features. Although there are some methods that have addressed these challenges, they were not comprehensive and could be further improved. For example, Zhu et al. [15] utilized 3D dual path networks (DPNs) to extract pulmonary nodule volume features which are effective for avoiding overfitting. However, the 3D features of pulmonary nodules are enormous and fine-grained features are not captured effectively. To capture fine-grained features, Zhang et al. [16] combined the squeeze-and-excitation attention network with aggregated residual transformations ResNeXt and applied their proposed method to the classification of malignant and benign lung nodules. The squeeze-and-excitation attention network is effective in refining features among channels, but the spatial features are not processed effectively in their work. Based on the DPNs architecture, Jiang et al. [17] added the spatial and contextual attention module into DPNs and ensembled a multi-network to classify pulmonary nodules. Their proposed attention mechanism was effective in feature refinement, but the spatial attention module could be further improved. Multi-scale, multi-resolution, and residual connection problems are not considered in their spatial attention network, which are very important for classifying lung nodules.

To extract more features of nodules, researchers tended to differentiate nodules with the fusion of deep learning features, raw nodule features, and hand-crafted nodule features. Xie et al. [18] fused gray level cooccurrence matrix (GLCM) features, Fourier shape features, and deep learning features and utilized AdaBoosted backpropagation neural network to differentiate nodules. Moreover, they fused the classification results of different classifiers. They tested their architecture on the LIDC-IDRI dataset, and the results proved that their fusion strategies were effective. Kaya et al. [19] fused deep features and hand-crafted features which contained morphological, color, or textual features of the nodules. Moreover, they utilized cascaded classifiers to differentiate the hybrid features. Jason et al. [20] fused

radiological quantitative image features and 3D deep features for patient-level nodule classification. Zhang et al. [21] fused local binary pattern-based texture features, histograms of oriented gradient-based shape features, and deep features which were extracted by 3D DPN. Finally, they utilized GBM to differentiate benign and malignant nodules. For the above hybrid features methods on lung cancer diagnosis, it was proven that combining manual features with deep learning features can further improve the accuracy of pulmonary nodules recognition [18–21], but the extraction of manual features still has the same problems as the traditional methods. To utilized the multi-attributes features and multi-scale features, Zhao et al. [22] proposed a multi-stream network for nodule classification, and they constructed a new loss function to overcome the imbalance of different attributes. However, the setting of the weight parameters between different tasks is crucial and unreasonable weight parameter settings will cause the network to fail to converge during multi-task training. However, the setting of the weight parameters between different tasks is manually set in their work, which causes the model to be unstable, and it is very difficult to debug the parameters.

Therefore, to tackle these challenges, 3D DPN is utilized as our trunk network, which is effective in avoiding overfitting. Moreover, a multi-scale spatial attention network is proposed to capture multi-scale spatial attention features. The residual attention network is used to capture multi-resolution spatial attention features, and the squeeze-and-excitation attention network is used to capture the fine-grained features among channels automatically. Finally, we combine the deep features with the raw image pixels to differentiate malignant and benign nodules.

3. Materials and Methods

3.1. Material and Preprocessing

In this paper, the employed dataset was Lung Image Database Consortium image collection (LIDC-IDRI) [23]. LIDC-IDRI is one of the largest open databases in the cancer imaging archive (TCIA) for lung cancer diagnosis, and LIDC-IDRI contains 1018 patients' chest CT scans. Each CT scan has a corresponding XML annotation file, which contains four professional thoracic radiologists' diagnosis annotations on the CT scans. The size of each scan is 512 × 512, and each XML file details the locations, boundaries, malignant level of each nodule, and so on.

The LIDC-IDRI dataset was equally and randomly split into ten folds with the LUNA16's split principle [24], 10-fold patient-level cross validation. The raw data were clipped into [−1200,600], and the values were normalized to [0,1] linearly. To reduce interference, the segmentation ground truth of LUNA16 was utilized to remove the nodules' background and the uncertain lung nodules for which the average malignant levels were equal to 3 were ignored. After data processing, we obtained 1001 3D lung nodules, in which there were 447 positive nodules and 554 negative nodules. The lung nodules of each subset were from different patients, and the detailed number of each subset is shown as Table 1.

Table 1. Detailed information of each subset.

Subset	Number of Patients	Number of Nodules	Number of Benign Nodules	Number of Malignant Nodules
Subset0	89	91	40	51
Subset1	89	106	61	45
Subset2	89	111	62	49
Subset3	89	103	57	46
Subset4	89	100	47	53
Subset5	89	91	55	36
Subset6	89	115	73	42
Subset7	89	90	62	28
Subset8	88	104	55	49
Subset9	88	90	42	48
Sum	888	1001	554	447

Nodules were cropped from the raw CT images at sizes of 32 × 32 × 32 randomly, and the nodules were enlarged to sizes of 36 × 36 × 36 by adding padding to the margin of the lung nodules. For data augmentation, we flipped the cropped data in the horizontal, vertical, z-axis directions and we set a 4 × 4 × 4 patch as 0 randomly. Our proposed architecture was tested on folds 0–9 separately, and the average experimental results are reported. Due to time and computation limitations, we conducted the ablation experiments on the 5th fold.

In addition, the data augmentation strategy beyond empirical risk minimization (MIXUP) [25] shows excellent performance on many natural image datasets, such as the ImageNet-2012, CIFAR-10, CIFAR-100, Google commands, and UCI datasets. Therefore, the strategy MIXUP was utilized to augment lung nodule data. The MIXUP strategy was only used in the training phase, and it constructed virtual training data. The MIXUP operation is described in Equation (1), in which x_i and x_j are raw input vectors, and y_i and y_j are one-hot label encodings. (x_i, y_i) and (x_j, y_j) are two examples drawn at random from our training data, and $\lambda \in [0, 1]$. The illustration of data preprocessing is shown as Figure 1.

$$\begin{cases} \widetilde{x} = \lambda x_i + (1 - \lambda) x_j \\ \widetilde{y} = \lambda y_i + (1 - \lambda) y_j \end{cases} \tag{1}$$

We used an Adam optimizer with a learning rate if 0.0001, and the momentum parameters were set as (0.5,0.999). We conducted our experiments on a workstation with a Linux system. The GPU hardware was TITANX, and the GPU memory was about 12 G. Due to the limitation in GPU memory, we set the training batch size to 4 and set the testing batch size to 2. The iteration epochs totalled 1200 for each fold.

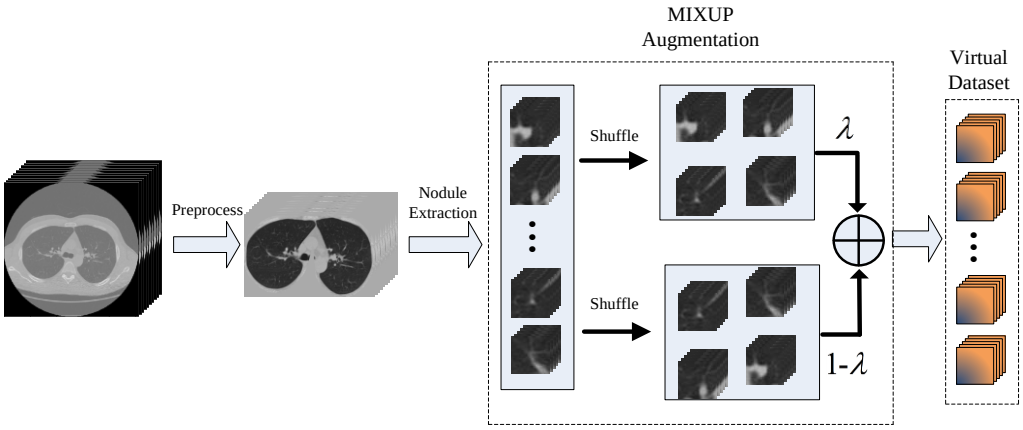

Figure 1. Illustration of the preprocessing.

3.2. Methods

In this paper, 3D DPN was utilized as our trunk network, which aggregated the residual connection and dense connection, and 3D DPN was effective in extracting a large number of features of 3D pulmonary nodules. To filter out redundant features and to capture useful fine-grained features, we utilized three types of attention networks to extract spatial attention features and channel attention features. RAN was used to extract the spatial attention features of different resolutions, MSAN was used to extract the spatial attention features of different scales, and SEN was used to extract channel attention features. After the fine-grained features were extracted, we utilized the GBM algorithm to classify nodules with deep features and raw nodule pixels. The overall architecture is shown as Figure 2, and each part of our architecture will be described in detail later.

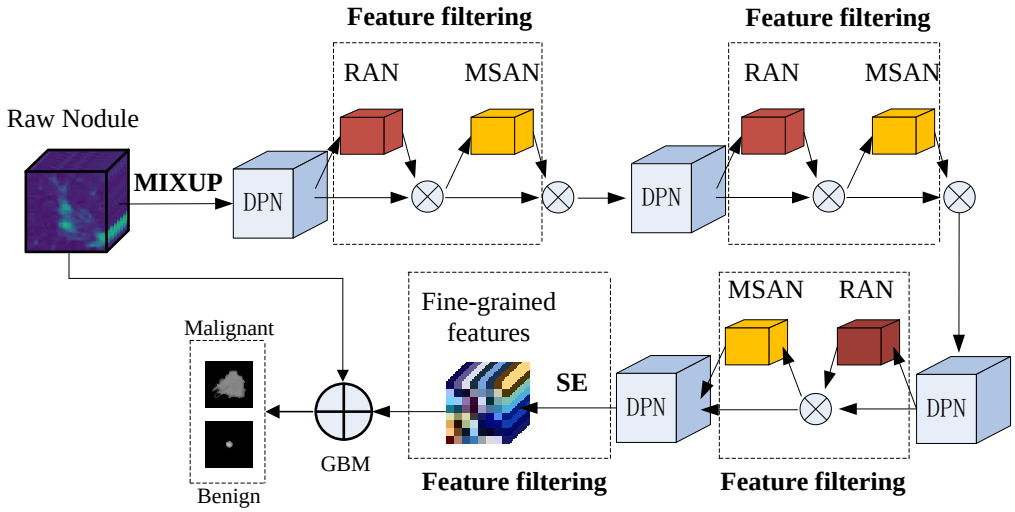

Figure 2. Illustration of the overall architecture.

3.2.1. Trunk Network

Due to the prominent recognition ability of DPN in the natural image recognition [26], 3D DPN was employed as the backbone network to extract nodule features. Dual path network integrates the residual connection and dense connection, and the dual path connection can be formulated as Equation (2), in which F represents the convolutional function, R represents the rectified linear units (RELU), y represents the feature map for dual path connection, and x represents the input raw image of dual path connection block.

$$y = R([x[:d], F(x)[:d], F(x)[d:] + x[d:]]) \quad (2)$$

The dual path connection is shown as Figure 3.

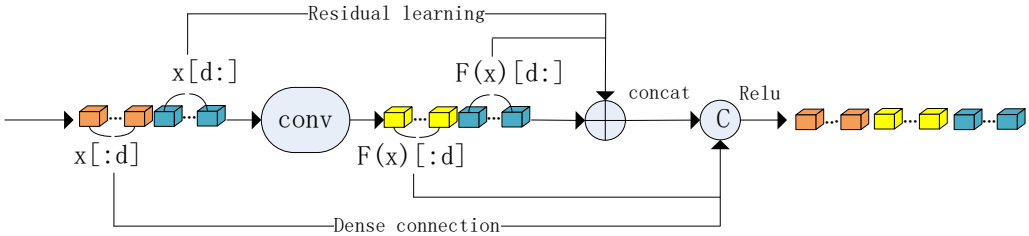

Figure 3. Illustration of the 3D dual path connection.

The parameters of DPN are shown in Table 2.

Table 2. Parameters of the dual path network (DPN) (baseline network).

Input	Layer	Channel	Blocks	Dense Depth
323 × 1	3D Conv	64	-	-
323 × 64	3D DPN blocks	192	3	16
323 × 192	3D DPN blocks	416	4	32
163 × 416	3D DPN blocks	776	10	24
83 × 776	3D DPN blocks	1408	3	128
43 × 1408	3D Average pool	1	-	-
1 × 1408	FC	1	-	-

As shown in Figure 3, the dual path network split the feature maps into two parts. One part of the feature maps was used for residual connection, and the other part of the feature

maps was used for residual learning. DPN can extract new features and combine semantic features with structural features efficiently, which is very important in classifying nodules.

As for the classifier, GBM was effective in constructing an advanced classifier. We used the GBM algorithm to differentiate malignant nodules and benign nodules with raw nodule pixels and deep features, and an illustration of the GBM is shown in Figure 4.

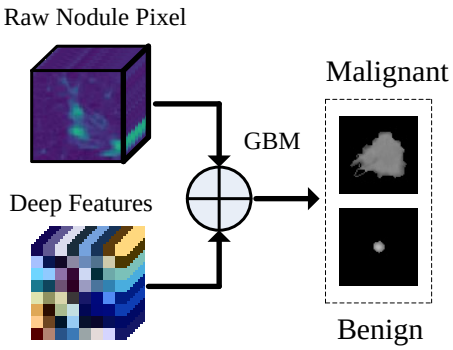

Figure 4. Illustration of the gradient boosting machine (GBM) algorithm.

3.2.2. Attention Mechanism

To capture fine-grained features, we applied a residual connection, and bottom-up and top-down structure to the attentive dual path network. The attention mechanism was named RAN. Based on the up-sample and down-sample operations, feature maps with different resolutions were generated, and the mixed features maps were generated by fusing features of different resolutions. A residual connection was helpful to capture deep feature information and to transmit structural features generated by the shallow layers of the DPN to the deep layers of the DPN. It was efficient to fuse structural features and semantic features. Besides, the RAN module can avoid overfitting to some extent. Based on the advantages of the RAN module, we can add the RAN module in the shallow, middle, and deep layers of a DPN to capture different structural features and semantic features and the multiple RAN modules do not result in a reduction in network performance. To our best knowledge, it is the first time RAN was applied to lung nodule classification. The RAN mechanism is shown in Figure 5, in which the up-sample method is trilinear interpolation, the down-sample is max-pooling, and the RB is a basic residual block. Besides, we apply the SEN module in the last layer of a DPN to capture the relationship among channels.

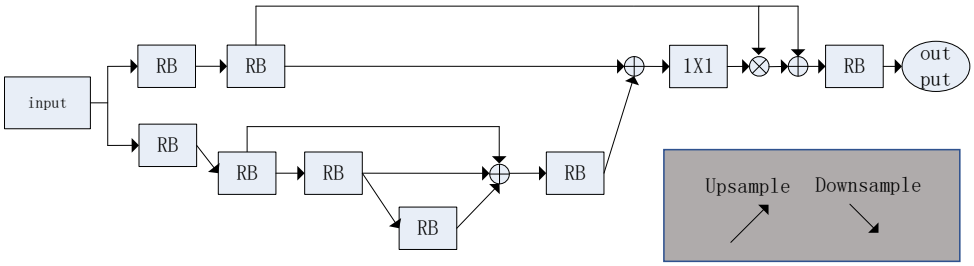

Figure 5. Illustration of the residual attention block.

Especially, we proposed a novel attention module named MSAN, which aggregates the multi-scale spatial attention mechanism and contextual attention mechanism. Based on our careful observation and experiments, we utilized two convolution paths to extract different features. One convolution path was designed with kernel size $1 \times 1 \times 1$, and the other one was designed with a dilated convolution with kernel size $3 \times 3 \times 3$ and dilation

size $2 \times 2 \times 2$. This process can be formulated as Equation (3), in which Z consists of $Z[:d]$ and $Z[d:]$, d is a hyperparameter, and we set $d = 1/2$.

$$\begin{cases} Z[d:] = W_f x \\ Z[:d] = W_g x \end{cases} \tag{3}$$

A convolution with kernel size 1×1 is efficient in extracting dense features with a small receptive field, and it is helpful to find subtle differences among nodules. A dilation convolution with kernel size 3×3 is efficient in extracting sparse features with a large receptive field, and it is helpful to capture the overall features of nodules. We chose to concatenate the two group feature maps rather than add them together. After the concatenation operation, the SE module was utilized to screen important features and to filter out unimportant features. The global average pooling operation was utilized to transform $Z \in \mathbb{R}^{H \times W \times D \times C}$ to $P \in \mathbb{R}^{1 \times 1 \times 1 \times C}$, and cth element of Z is calculated by Equation (4), in which $H \times W \times D$ is the spatial dimensions of input feature maps Z.

$$P_c = \frac{1}{H \times W \times D} \sum_{i=1}^{H} \sum_{j=1}^{W} \sum_{d=1}^{D} Z_c(i,j,d) \tag{4}$$

To capture the channel-wise dependencies, a fully connected layer was employed, and it can be expressed as Equation (5), in which δ represents the ReLU function, $W_1 \in \mathbb{R}^{C/r \times C}$, and $W_2 \in \mathbb{R}^{C \times C/r}$. r is the reduction ratio, which is used to reduce the parameters of the model.

$$M = F_{fc}(P) = \delta(W_2 \delta(W_1 P)) \tag{5}$$

The final results of the MSA block can be calculated by Equation (6), in which $O = [O_1, O_2, \ldots, O_c]$ and $F_{scale}(Z_c, M_c)$ refers to a channel-wise multiplication between the scalar M_c and the feature map $Z_c \in \mathbb{R}^{H \times W \times D}$.

$$O_c = F_{scale}(Z_c, M_c) = Z_c M_c \tag{6}$$

Our proposed MSA block is shown as Figure 6, in which DLT represents dilation convolution and AVG represents average pooling.

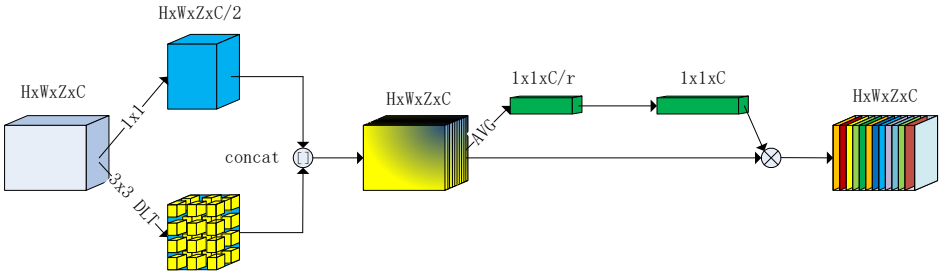

Figure 6. Illustration of a multi-scale attention block.

4. Results and Discussions

We conducted a series of experiments to evaluate the effectiveness of our proposed architecture. In this section, we will present the evaluation metrics, the dataset preprocessing, detailed experimental parameter settings, ablation experiments, and contrast experimental results with other methods.

4.1. Evaluation Metrics

To evaluate the performance of our proposed architecture on malignant and benign nodules classification, we use four indices, including accuracy (ACC), sensitivity (TPR), false positive rate (FPR), and F1-score to verify our proposed framework, and the four indices are described as follows.

- Accuracy (ACC)—the percentage of nodules that are correctly predicted.
- Sensitivity (TPR)—the percentage of correctly predicted true-positive nodules to nodules which have positive labels.
- False positive rate (FPR)—the percentage of incorrectly predicted positive nodules to nodules which have negative labels.
- F1-Score—the model's ability to balance TPR and FPR and the comprehensive ability of the model.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

$$TPR = \frac{TP}{TP + FN} \tag{8}$$

$$FPR = \frac{FP}{FP + TN} \tag{9}$$

$$F1 - score = \frac{2TP}{2TP + FP + FN} \tag{10}$$

A true positive (TP) represents the number of malignant nodules that are predicted correctly as malignant nodules; a true negative (TN) represents the number of benign nodules that are predicted correctly as benign nodules; a false positive (FP) represents the number of benign nodules that are predicted incorrectly as malignant nodules; and a false negative (FN) represents the number of malignant nodules that are predicted incorrectly as benign nodules.

In a nutshell, greater values of accuracy, sensitivity, and F1-score but lower FPR values indicate better performance. The F1-score represents the overall performance of the model, and the highest F1-score represents the best classification performance of the model.

4.2. Ablation Experiment

The ablation experiment is similar to the controlled variable method. In other words, we can control a single variable to determine whether this variable has an impact on the entire system. To verify the effectiveness of the proposed strategies, i.e., MIXUP data augmentation, RAN module, SE module, and MSAN module, we conducted a series of experiments, and the experiment details are described as follows. We use the 5th fold for validation and the remaining nine folds for training. In other words, the training set contained fold0, fold1, fold2, fold3, fold4, fold6, fold7, fold8, and fold9, and the validation set only contained fold5. The proposed strategies are described as follows.

4.2.1. MIXUP Data Augmentation

To verify the effectiveness of MIXUP data augmentation, we conducted a comparison experiment with or without MIXUP, and the results are shown in Table 2. Based on the comparison results, it is proven that MIXUP data augmentation is helpful in improving the classification performance of the model, including accuracy and sensitivity. MIXUP data augmentation constructs a virtual dataset and increases the diversity of data, which are important for improving the robustness of a model. Besides, the data augmentation operation also strengthened the model's ability to diagnose diseased nodules.

4.2.2. Residual Attention and Squeeze-and-Excitation Attention

To verify the effectiveness of the RAN and SE modules, we conducted a comparison experiment with or without the RSE and SE modules, and the results are shown in Table 2. Based on the comparison experiments results, it is proven that the residual attention module and squeeze-and-excitation module are both helpful in improving the classification performance of the model. The residual attention module is effective in capturing spatial attentive features of different resolutions, and the squeeze-and-excitation module is effective in capturing contextual attentive features, which are essential for nodule classification.

4.2.3. Multi-Scale Attention

To verify the effectiveness of our proposed MSAN module, we conducted a comparison experiment with or without the MSAN module, and the results are shown in Table 2. Based on the comparison experiments results, it is proven that our proposed multi-scale attention module is helpful in improving the classification performance of the model. Our proposed multi-scale attention network is effective in capturing the salient attentive features of different scales and concatenate the sparse features and dense features.

4.2.4. The Overall Architecture

We combine MIXUP and the various attention networks separately and test them on the dataset. Based on Table 3, we can find that, when the MIXUP operation is combined with different attention networks, the classification performance of the model will be improved. Especially, the accuracy of classification increases a lot when the MIXUP operation is combined with RAN. This represents that the spatial attention features of multi-resolution is more effective in diagnosing negative nodules than positive nodules. When MIXUP is combined with RAN and MSAN, the sensitivity increases a lot. This represents that the multi-scale attention network is effective in capturing the spatial fine-grained features of positive nodules. When the MIXUP operation is combined with MSAN, the accuracy increases a bit but sensitivity decreases a lot, which is similar to the combination of MIXUP and SE. When the three attention modules are combined with MIXUP at the same time, the model's performance is best. Based on the different combinations, we can find that the features extracted by different attention modules are complementary. When MIXUP, RAN, MSAN, and SE are all added into a baseline model DPN, the classification performance is best. Therefore, we aggregated the above three attention modules into our model and proposed an overall architecture (DPN + MIXUP + RSA + MSAN + SE), which is shown as Figure 2. The RSA module was utilized to extract different spatial attention features with different resolutions. Based on the different spatial attention features extracted, the MSAN module was utilized to extract multi-scale features and to fuse them effectively. Because of the advantages of the residual connection in the RSA module, we could apply the RSA and MSAN modules in different layers of the network, which is important when capturing different structural features and semantic features, and this operation does not cause overfitting. At the end of the network, the SE module was utilized to capture attention features between channels and to screen important channels. The proposed architecture was tested on the dataset, and the results are shown in Table 3. Based on the results, our proposed architecture is proven to be effective for nodule classification.

Table 3. Ablation experiments on the 5th fold.

Model	ACC	TPR
DPN	90.22%	81.1%
DPN + MIXUP	91.30%	91.89%
DPN + RAN	90.22%	94.59%
DPN + MSAN	91.30%	91.89%
DPN + SEN	90.22%	89.19%
DPN + MIXUP + RAN	93.48%	89.19%
DPN + MIXUP + MSAN	92.39%	83.78%
DPN + MIXUP + SEN	91.30%	86.49%
DPN + MIXUP + RAN + MSAN	92.39%	97.29%
DPN + MIXUP + RAN + MSAN + SEN	94.57%	89.19%

To evaluate the effect of the threshold, we recorded the changes in F1-score under different thresholds, which are shown in Figure 7. When the threshold is less than 0.5, the value of the F1-score increases with the increase in threshold, but when the threshold is greater than 0.5, the F1-score decreases with the increase in threshold. Therefore, we set the threshold as 0.5 in our experiments.

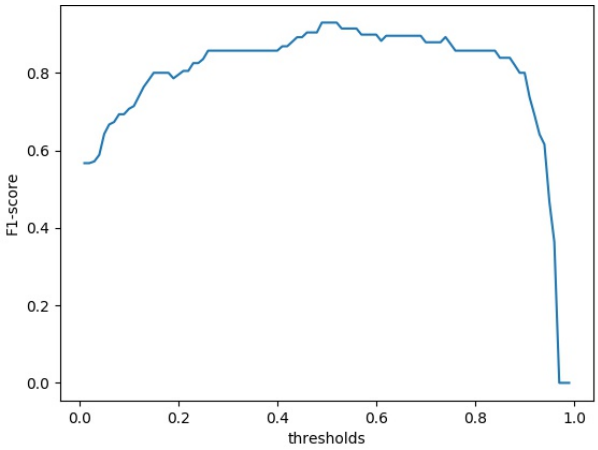

Figure 7. The effects of different thresholds on F1-score.

4.3. Comparison with State-of-the-Art Methods

We tested our proposed architecture on the ten folds and obtained the average results across the ten folds. We compared our experimental results with the existing advanced methods, including the 2D network, 3D network, and hybrid feature network. The comparison results are shown in Table 4.

Table 4. Comparison with existing methods across the ten folds.

	Year	ACC	TPR	FPR	F1-Score
Kumar et al. [27]	2015	75.01	83.35	39.0	70.44
Multi-scale CNN [28]	2015	86.84	-	-	-
Slice-level 2D CNN [29]	2016	86.70	78.60	8.8	84.43
Nodule-level 2D CNN [29]	2016	87.30	88.50	14.0	87.23
Vanilla 3D CNN [29]	2016	87.40	89.40	14.8	87.25
Multi-crop CNN [13]	2017	87.14	-	-	-
Deep 3D DPN [15]	2018	88.74	-	-	-
Nodule Size + Pixel + GBM [15]	2018	86.12	-	-	-
DeepLung [15]	2018	90.44	81.42	-	-
Xie et al. [18]	2018	89.53	-	-	-
Liu et al. [14]	2018	89.50	-	-	-
Xie et al. [30]	2019	91.60	86.52	-	-
Polat et al. [31]	2019	91.81	88.53	-	-
da Nóbrega [32]	2020	88.41	85.38	-	78.83
Attentive and ensemble CNN [16]	2020	90.24	92.04	11.06	90.45
Zhang et al. [17]	2020	91.67	-	-	-
Lima et al. [33]	2020	88.00	82.00	-	-
Liu et al. [34]	2020	90.60	83.70	-	-
Our proposed model	2020	**91.90**	**91.30**	**8.00**	**91.00**

Based on the comparison results, our model achieves the highest accuracy and F1-score and the lowest FPR. Moreover, the TPR of our model is also very high, second only to attentive and ensemble CNN (AEC) [16]. Compared to AEC, the metric TPR of our model is slightly lower, but the ACC, FPR, and F1-score are better than those of AEC. A large number of deep features are extracted through 3D DPN, which is effective in reusing features, and the original features of the raw images are captured as much as possible because of the residual connection and dense connection. Based on the rich features, the residual attention module and squeeze-and-excitation attention module are utilized to capture the spatial attention features and contextual attention features. Besides, we proposed a multi-scale attention module to capture attention features of different scales, and the different attention features are fused effectively. After the attention operation, the unimportant features were filtered out, and the salient and fine-grained features were captured, which are important

to differentiate between malignant and benign nodules. At last, based on the comparison results in Table 3, our model is proven to achieve a state-of-the-art level.

4.4. Visualization of Classification Results

To verify the performance of our proposed architecture better, we visualize the diagnosis results of some benign and malignant nodules. In Figure 8, the benign nodules are presented in the first row and the malignant nodules are presented in the second row. We make a comparison between the predicted probability of our proposed architecture, and the average diagnosis level of four radiologists and the comparison results are shown for each nodule. If the predicted probability is less than 0.5, the nodule is predicted to be a benign nodule; otherwise, it is predicted as a malignant nodule. Similarly, if the average diagnosis level of four radiologists is less than 3, the nodule is labeled as benign nodule and, if the average diagnosis level is more than 3, the nodule is labeled as malignant nodule. The visualization results show that our predicted probability of each nodule is consistent with the average diagnosis level of doctors, which proves that our proposed model is robust and that the classification results are very reliable.

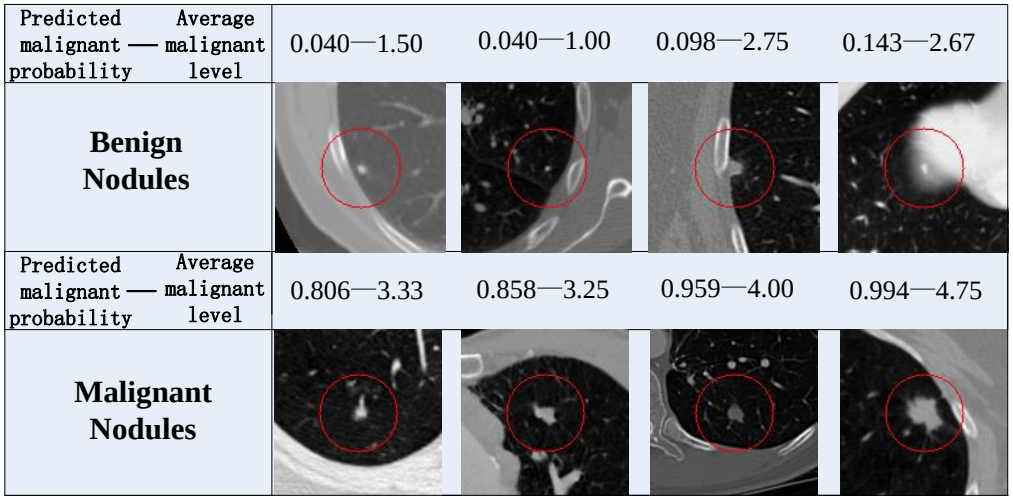

Figure 8. Visualization of benign and malignant diagnosis results: the number for each nodule shows the predicted malignant probability and the average malignancy level of four doctors.

Moreover, we add some visualization of misclassification examples which are shown in Figure 9. Based on Figure 9, our method is still insufficient for the recognition of some uncertain nodules. For example, for the first nodule on the left in Figure 9, its malignant level is diagnosed by two doctors as 4, which represents a malignant nodule, and its malignant level is diagnosed by one doctor as 3, which represents an uncertain nodule. Moreover, its malignant level is diagnosed by one doctor as 2, which represents a benign nodule. Finally, the average malignant level is 3.25, which represents the nodule as being slightly malignant and similar to an uncertain nodule for which the malignant level is 3. For such nodules, professional doctors also have differences in the diagnosis of such nodules and cannot clearly diagnose them. Such nodules often need to be further confirmed in the patient's follow-up examination. Therefore, the diagnosis of such nodules is very difficult. In the future, we hope to obtain lung cancer datasets with patient follow-up examinations, such as the National Lung Cancer Screening Trial (NLST) dataset, and we will make a further study.

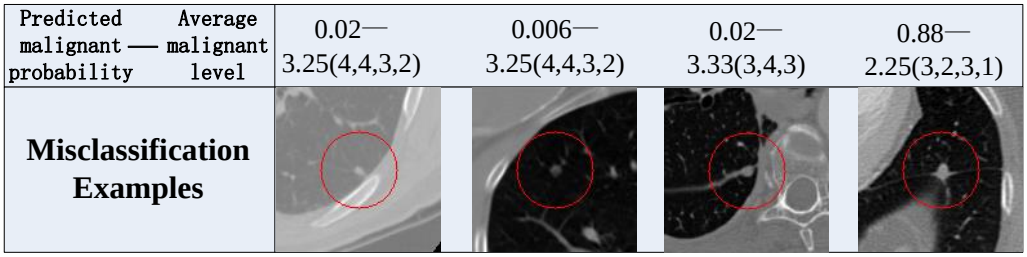

Figure 9. Visualization of misclassification examples: the number for each nodule shows the predicted malignant probability and the average malignancy level of doctors. The values in brackets refer to the diagnosis malignancy level of this lung nodule by different doctors.

4.5. Computational Complexity

To evaluate the models' computational complexities, we take a 3D image cube of size $32 \times 32 \times 32$ as the input to each model and calculate the test time of each model. All results are calculated on the same workstation mentioned before, and the results are shown in Table 5. Moreover, we train each model for one epoch and record the training time of each model in Table 6. The results show that our proposed model slightly increases the time-consumption, but the test time of each nodule is less than 0.1 s as well. In addition, the slight increase in complexity significantly improves the classification performance.

Table 5. The test time of different models.

Model Name	Test Time (s)
DPN	0.0474
DPN + RSE	0.0692
DPN + MSA	0.0594
DPN + SE	0.0413
DPN + RSE + MSA + SE	0.0916

Table 6. The training time of different models.

Model Name	Test Time (s)
DPN	107
DPN + RSE	158
DPN + MSA	140
DPN + SE	109
DPN + RSE + MSA + SE	189

5. Conclusions

In this paper, we developed a novel method for differentiating malignant and benign lung nodules. Firstly, we used the MIXUP method to construct a virtual training data, and MIXUP improved the diversity of the data. Secondly, we utilized the 3D dual-path network to extract nodule features and used the GBM algorithm to differentiate pulmonary nodules with deep features and raw nodule pixels. Thirdly, we applied the RAN and SE modules to capture spatial and contextual features and we proposed a novel multi-scale attention module to capture multi-scale attentive features. Based on the ablation experimental results in Table 2, it is proven that our proposed multi-scale attention module and the application of RAN and SEN are very effective for classifying nodules. Finally, we compared our experimental results with the state-of-the-art methods in Table 3, and the comparison results prove that our proposed model exceeds the previous state-of-the-art model. However, the results of Table 6 show that both the RAN and MSAN modules are a bit time-consuming. Although the performance of classification increases a lot, the complexity of our model is a bit higher compared to the baseline model. In Figure 9, those nodules with average malignant levels very close to 3 are very difficult to diagnose and we will further study those nodules. In the future, we will aim to lighten our network and to

fuse the nodule's other attributes features with the deep features extracted by our proposed deep learning architecture. Based on the feature fusion of our extracted deep features and other attributes features, the classification performance will achieve a new higher level.

Author Contributions: Conceptualization, K.X. and C.W.; methodology, K.X.; software, K.X and Y.J.; validation, K.X., C.W., and J.C.; formal analysis, Y.G. and Y.J.; investigation, K.X.; resources, K.X.; data curation, K.X. and J.C.; writing—original draft preparation, K.X.; writing—review and editing, K.X., C.W., J.C., and Y.G.; visualization, K.X.; supervision, K.X. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China under grant nos. 61701101, 61973093, U1713216, U20A20197 and 61973063.

Institutional Review Board Statement: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent Statement: Informed consent was obtained from all individual participants included in the study.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: https://wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI, https://luna16.grand-challenge.org/data/.

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

References

1. Siegel, R.L.; Miller, K.D.; Jemal, A. Cancer statistics. *CA Cancer J. Clin.* **2019**, *69*, 7–34. [CrossRef] [PubMed]
2. Wang, M.; Long, F.; Tang, F.; Jing, Y.; Wang, X.; Yao, L.; Ma, J.; Fei, Y.; Chen, L.; Wang, G.; et al. Autofluorescence imaging and spectroscopy of human lung cancer. *Appl. Sci.* **2017**, *7*, 32. [CrossRef]
3. Lodwick, G.S. Computer-aided diagnosis in radiology: A research plan. *Investig. Radiol.* **1966**, *1*, 72–80. [CrossRef] [PubMed]
4. Dou, Q.; Chen, H.; Yu, L.; Qin, J.; Heng, P.A. Multilevel contextual 3-D CNNs for false positive reduction in pulmonary nodule detection. *IEEE Trans. Biomed. Eng.* **2016**, *64*, 1558–1567. [CrossRef] [PubMed]
5. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt LC, UT, USA, 18–22 June 2018; pp. 7132–7141.
6. Wang, F.; Jiang, M.; Qian, C.; Yang, S.; Li, C.; Zhang, H.; Wang, X.; Tang, X. Residual attention network for image classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 3156–3164.
7. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
8. Lee, S.L.A.; Kouzani, A.Z.; Hu, E.J. Random forest based lung nodule classification aided by clustering. *Comput. Med. Imaging Graph.* **2010**, *34*, 535–542. [CrossRef]
9. Cascio, D.; Taormina, V.; Raso, G. Deep Convolutional Neural Network for HEp-2 Fluorescence Intensity Classification. *Appl. Sci.* **2019**, *9*, 408. [CrossRef]
10. Li, X.X.; Li, B.; Tian, L.F.; Zhang, L. Automatic benign and malignant classification of pulmonary nodules in thoracic computed tomography based on RF algorithm. *IET Image Process* **2018**, *12*, 1253–1264. [CrossRef]
11. de Carvalho Filho, A.O.; Silva, A.C.; de Paiva, A.C.; Nunes, R.A.; Gattass, M. Computer-aided diagnosis of lung nodules in computed tomography by using phylogenetic diversity, genetic algorithm, and SVM. *J. Dig. Imaging* **2017**, *30*, 812–822. [CrossRef]
12. de Sousa Costa, R.W.; da Silva, G.L.F.; de Carvalho Filho, A.O.; Silva, A.C.; de Paiva, A.C.; Gattass, M. Classification of malignant and benign lung nodules using taxonomic diversity index and phylogenetic distance. *Med. Biol. Eng. Comput.* **2018**, *56*, 2125–2136. [CrossRef]
13. Shen, W.; Zhou, M.; Yang, F.; Yu, D.; Dong, D.; Yang, C.; Zang, Y.; Tian, J. Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification. *Pattern Recognit.* **2017**, *61*, 663–673. [CrossRef]
14. Liu, Y.; Hao, P.; Zhang, P.; Xu, X.; Wu, J.; Chen, W. Dense Convolutional Binary-Tree Networks for Lung Nodule Classification. *IEEE Access* **2018**, *6*, 49080–49088. [CrossRef]
15. Zhu, W.; Liu, C.; Fan, W.; Xie, X. Deeplung: Deep 3d dual path nets for automated pulmonary nodule detection and classification. In Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Harvey's Casino in Lake Tahoe, Stateline, NV, USA, 12–15 March 2018; pp. 673–681.
16. Jiang, H.; Gao, F.; Xu, X.; Huang, F.; Zhu, S. Attentive and ensemble 3D dual path networks for pulmonary nodules classification. *Neurocomputing* **2020**, *398*, 422–430. [CrossRef]

17. Zhang, G.; Yang, Z.; Gong, L.; Jiang, S.; Wang, L.; Zhang, H. Classification of lung nodules based on CT images using squeeze-and-excitation network and aggregated residual transformations. *Radiol. Med.* **2020**, *125*, 1–10. [CrossRef] [PubMed]
18. Xie, Y.; Zhang, J.; Xia, Y.; Fulham, M.; Zhang, Y. Fusing texture, shape and deep model-learned information at decision level for automated classification of lung nodules on chest CT. *Inf. Fusion* **2018**, *42*, 102–110. [CrossRef]
19. Kaya, A. Cascaded classifiers and stacking methods for classification of pulmonary nodule characteristics. *Comput. Method. Progr. Biomed.* **2018**, *166*, 77–89. [CrossRef]
20. Causey, J.L.; Zhang, J.; Ma, S.; Jiang, B.; Qualls, J.A.; Politte, D.G.; Prior, F.; Zhang, S.; Huang, X. Highly accurate model for prediction of lung nodule malignancy with CT scans. *Sci. Rep.* **2018**, *8*, 1–12. [CrossRef]
21. Zhang, G.; Yang, Z.; Gong, L.; Jiang, S.; Wang, L. Classification of benign and malignant lung nodules from CT images based on hybrid features. *Phys. Med. Biol.* **2019**, *64*, 125011. [CrossRef]
22. Zhao, J.; Zhang, C.; Li, D.; Niu, J. Combining multi-scale feature fusion with multi-attribute grading, a CNN model for benign and malignant classification of pulmonary nodules. *J. Dig. Imaging* **2020**, *33*, 1–10. [CrossRef]
23. Armato, S.G.; McLennan, G.; Bidaut, L.; McNitt-Gray, M.F.; Meyer, C.R.; Reeves, A.P.; Zhao, B.; Aberle, D.R.; Henschke, C.I.; Hoffman, E.A.; 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*, 915–931. [CrossRef]
24. Kuan, K.; Ravaut, M.; Manek, G.; Chen, H.; Lin, J.; Nazir, B.; Chen, C.; Howe, T.C.; Zeng, Z.; Chandrasekhar, V. Deep Learning for Lung Cancer Detection: Tackling the Kaggle Data Science Bowl 2017 Challenge. *arXiv* **2017**, arXiv:1705.09435.
25. Zhang, H.; Cisse, M.; Dauphin, Y.N.; Lopez-Paz, D. Mixup: Beyond Empirical Risk Minimization. *arXiv* **2017**, arXiv:1710.09412.
26. Chen, Y.; Li, J.; Xiao, H.; Jin, X.; Yan, S.; Feng, J. Dual path networks. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 3–9 December 2017; pp. 4467–4475.
27. Kumar, D.; Wong, A.; Clausi, D.A. Lung nodule classification using deep features in CT images. In Proceedings of the 2015 12th Conference on Computer and Robot Vision, Washington, DC, USA, 3–5 June 2015; pp. 133–138.
28. Shen, W.; Zhou, M.; Yang, F.; Yang, C.; Tian, J. 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; pp. 588–599.
29. Yan, X.; Pang, J.; Qi, H.; Zhu, Y.; Bai, C.; Geng, X.; Liu, M.; Terzopoulos, D.; Ding, X. Classification of lung nodule malignancy risk on computed tomography images using convolutional neural network: A comparison between 2d and 3d strategies. In Proceedings of the Asian Conference on Computer Vision, Taipei, Taiwan, 20–24 November 2016; pp. 91–101.
30. Xie, Y.; Xia, Y.; Zhang, J. Knowledge-based Collaborative Deep Learning for Benign-Malignant Lung Nodule Classification on Chest CT. *IEEE Trans. Med. Imaging* **2019**, *38*, 99–1004. [CrossRef] [PubMed]
31. Polat, H.; Danaei Mehr, H. Classification of pulmonary CT images by using hybrid 3D-deep convolutional neural network architecture. *Appl. Sci.* **2019**, *9*, 940. [CrossRef]
32. da Nóbrega, R.V.M.; Rebouças Filho, P.P.; Rodrigues, M.B.; da Silva, S.P.; Júnior, C.M.D.; de Albuquerque, V.H.C. Lung nodule malignancy classification in chest computed tomography images using transfer learning and convolutional neural networks. *Neural Comput. Appl.* **2020**, *32*, 1–18. [CrossRef]
33. Lima, L.L.; Ferreira Junior, J.R.; Oliveira, M.C. Toward classifying small lung nodules with hyperparameter optimization of convolutional neural networks. *Comput. Intell.* **2020**. [CrossRef]
34. Liu, H.; Cao, H.; Song, E.; Ma, G.; Xu, X.; Jin, R.; Liu, C.; Hung, C.C. Multi-model Ensemble Learning Architecture Based on 3D CNN for Lung Nodule Malignancy Suspiciousness Classification. *J. Dig. Imaging* **2020**, *33*, 1–15. [CrossRef]