國立臺灣大學工學院工程科學及海洋工程學系

碩士論文

Department of Engineering Science and Ocean Engineering

College of Engineering

National Taiwan University

Master Thesis

結合紋理特徵與多視角殘差選擇核網路
分類肺結節良惡性

Multiview residual selective kernel networks for lung nodule classification associated with texture features

吳承哲

Cheng-Zhe Wu

指導教授：張恆華 博士

Advisor: Herng-Hua Chang, Ph.D.

中華民國 110 年 10 月

October 2021

doi:10.6342/NTU202104149

誌謝

能完成這篇論文，我要特別感謝張恆華老師，收我進到這間實驗室，並且給了我許多寶貴的意見，使我的研究更豐富多元，也要謝謝許文翰老師，在我碩班初期給予許多資源，讓我能夠毫無顧慮的學習與做研究，最後感謝家人朋友的支持以及實驗室的每一位成員對我的幫忙，謝謝大家。

doi:10.6342/NTU202104149

摘要

肺癌為世界上有著高死亡率的癌症之一，在本篇論文中，我們致力於改進電腦輔助診斷方法，使用電腦斷層影像來預測肺結節為惡性的機率。因為肺結節有著許多種不同的大小及形狀以至於難以辨識，因此我們提出了多視角殘差選擇核網路(MRSKNet)。此網路模型結合了殘差網路(ResNet)以及選擇核網路(SKNet)的優點，前者使模型能夠重複使用特徵，後者則可以自動選擇感受區的尺寸大小。使用的公開資料集為LIDC-IDRI，裡面包含了肺部電腦斷層影像，而我們篩選出877顆肺結節來做為訓練資料，其中447顆為良性、430顆為惡性，並且使用十折交叉驗證來評估我們設計的模型。在使用原始影像的實驗中達到的接受者操作特徵曲線的曲線下面積(AUC)為0.9696、準確率(accuracy)為0.9349以及靈敏度(sensitivity)為0.9346。另外，我們還使用了灰度共生矩陣(Gray-level co-occurrence matrix)、灰度行程矩陣(gray-level run length matrix)及田村(Tamura)紋理特徵來豐富輸入的影像資料。我們將原始影像跟對應的紋理特徵連接起來讓機器從紋理特徵中去學習，並且比較每一種特徵的優劣。而同質性(HOM)使我們的方法達到更好的效能，使接受者操作特徵曲線下的面積、準確率及靈敏度提高到0.9711、0.9366及0.9556。最後，我們還跟多種基本模型架構以及前人的實驗方法做比較，來驗證我們所提出的方法，而實驗結果顯示出使用我們MRSKNet加上同質性紋理特徵的分類能力優於大部分最先進的方法。

關鍵字：肺結節分類、多視角卷積神經網路、殘差學習、選擇核、紋理特徵、電腦斷層。

doi:10.6342/NTU202104149

ABSTRACT

Lung cancer is one of the leading causes of death worldwide. This thesis is dedicated to improving the computer-aided diagnosis (CAD) to predict the likelihood of malignant nodules in computed tomography (CT) images. Because the lung nodules, which have various sizes and shapes, are hard to recognize, we present a multiview residual selective kernel network (MRSKNet), which integrates the advantages of ResNet for feature reuse and SKNet for adaptive receptive field (RF) size selection. The MRSKNet is evaluated on Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI), which is a public database of lung CT images with 877 (447 benign and 430 malignant) nodules. Based on ten-fold cross validation, the experiment with original images achieve area under the receiver operating characteristic curve (AUC) of 0.9696, accuracy of 0.9349, sensitivity of 0.9346. Additionally, seven kinds of texture features calculated from the gray-level co-occurrence matrix (GLCM), gray-level run length matrix (GLRLM), and Tamura methods are exploited. We concatenated the original images with the texture features to diversify the input data. Among them, the homogeneity (HOM) improved the performance most. The AUC, accuracy, and sensitivity which were more considerable were increased to 0.9711, 0.9366, and 0.9556, respectively. Finally, we compared with other baseline models and previous works to validate our proposed method. Experimental results indicated that the classification ability of our MRSKNet with HOM outperformed the most state-of-the-art methods.

Keywords: lung nodule classification; Multiview CNN; Residual learning; Selective kernel; Texture feature; CT.

doi:10.6342/NTU202104149

CONTENTS

doi:10.6342/NTU202104149

LIST OF FIGURES

doi:10.6342/NTU202104149

doi:10.6342/NTU202104149

LIST OF TABLES

doi:10.6342/NTU202104149

doi:10.6342/NTU202104149

Chapter 1 Introduction

Cancer is one of the leading causes of death in the United States and it is also a public health problem worldwide. In 2020, lung cancer cases were the second most but the number of deaths caused by lung cancer was the greatest. There were 2.2 million people, which accounted for 11.4% among all cancers, suffering from lung cancer and the total number of deaths was 1.7 million [1], which reached a high mortality rate of 80%. Besides, the 5-year relative survival rate for lung cancer was only 21% [2]. However, National Lung Screening Trial demonstrated that the mortality of lung cancer can be reduced by 20% by earlier diagnoses with low-dose computed tomography (CT) screening [2, 3]. A white spot in the thoracic CT scans is defined as a lung nodule [4]. It is a round area and more solid than normal lung tissue, and sometimes can be a lung cancer. If it is larger than 30 mm, it is called lung mass and has a higher probability being cancerous [5]. Lung nodules can be divided into benign nodules, indeterminate nodules, and malignancy. Benign nodules are noncancerous and usually have smaller sizes and smoother contours. Malignancies are cancerous and usually have larger sizes and more variable morphologies. Different nodules are illustrated in Fig. 1-1.

Thoracic radiologists usually read thoracic CT scans to classify nodules slice by slice. This approach is time-consuming, expensive, and error prone. The work experience also affects the accuracy of differentiating benign from malignant nodules. With the development of computer technology, computer-aided diagnosis (CAD) systems have been widely used in medical image analysis, which have shown good performance on lung cancer diagnosis [6, 7]. It also promotes better quality of medical service.

LIDC-IDRI is a public database for lung cancer analysis [8]. There are 1018 cases.

doi:10.6342/NTU202104149

Those nodules whose diameters are longer than 3 mm were annotated the locations of contours and the likelihood of malignancy. As depicted in Fig. 1-1, different nodules are difficult to distinguish through human's eyes by the size, shape, or texture. Therefore, developing CAD methods to help radiologists to classify these lung nodules is a current tendency.

There are four main procedures in CAD methods for lung nodule classification [9]: (1) data collection, (2) data preprocessing, (3) feature extraction, and (4) nodule classification. Recently, thanks to rapid hardware development, deep learning has been used for image classification widely. This approach can efficiently perform feature extraction and nodule classification at the same time. The success of deep learning motivates us the use of neural networks for lung nodule classification. In our approach, we concatenated the multiple views of three-dimensional (3D) anatomical images and the handcrafted texture features as input data. First, we extracted anatomical planes (axial, coronal, and sagittal) from 3D CT images as our original input [10-14]. Then we selected seven texture features from different methods as the input [15, 16] and determined the texture feature that was most helpful to recognize the categories of nodules. In terms of models, we proposed the multiview residual selective kernel network (MRSKNet). We used the concept of the selective kernel and residual learning to build the core block, which was called the RSK block [17, 18]. To design the RSK block, we used two bottleneck convolution operations, whose kernel sizes were different to extract local receptive fields (RFs), and one identity mapping branch to avoid the degradation problem. Moreover, we deployed the soft attention mechanism to generate weights with global RFs and distribute to the above mentioned branches. Then, we used global max pooling (GMP) and concatenated the multiview features before entering a fully connected (FC) layer. Finally, the nodules were classified by the

doi:10.6342/NTU202104149

sigmoid function.

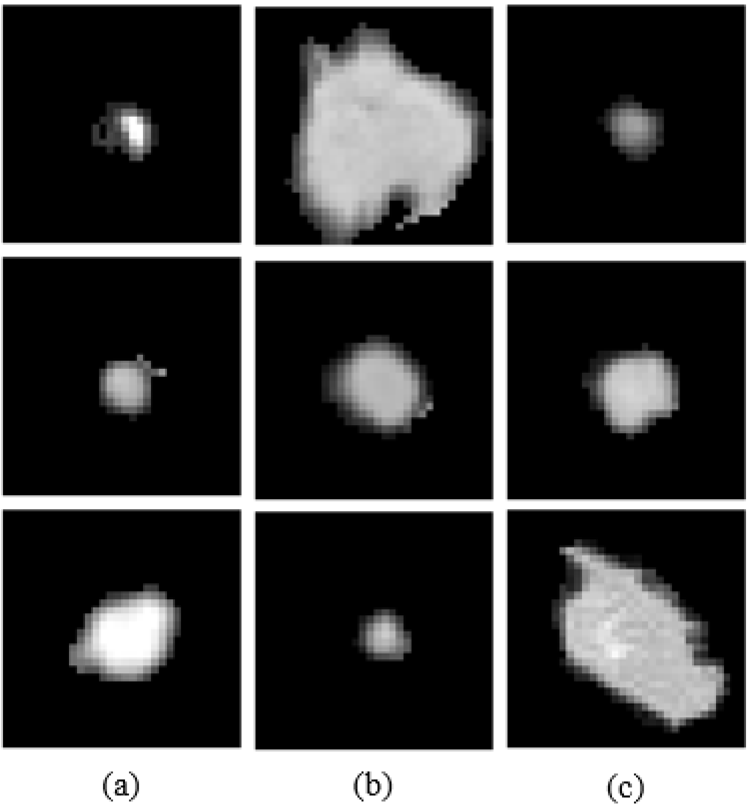


Fig. 1-1 Examples of lung nodules in the axial plane. (a) Benign nodules. (b) Malignant nodules. (c) Indeterminate nodules.

doi:10.6342/NTU202104149

Chapter 2 Related Work

In recent years, there have been many studies for classifying the nodules into benign and malignant. In terms of feature extraction, there are two main categories. One is handcrafted features by traditional image processing and the other is deep features from deep convolutional neural networks. In the part of classification, the methods are generally divided into traditional machine learning and deep convolutional neural networks (CNNs).

In most cases, handcrafted features and traditional machine learning methods were used in the same experiment. Handcrafted features, which can offer useful information for lung nodule analysis, usually include shape, texture, intensity, and morphology. They are fed to the machine learning classifier for training. Dhara et al. [19] separated benign from malignant nodules by the support vector machine (SVM) with several shape-based and texture-based features. Orozco et al. [20] used the wavelet transform to extract nodule features and the SVM for classification. de Sousa Costa et al. [21] used the mean phylogenetic distance (MPD) and taxonomic diversity index as texture feature, and applied the genetic algorithm with SVM to classify nodules. Firmino [22] extracted features by the histogram of oriented gradient (HOG) method. de Carvalho Filho et al. [23] used the phylogenetic diversity to describe lung nodules. Sasidhar et al. [16] used the Gray-level co-occurrence matrix (GLCM) to calculate the texture features of nodules and the SVM for classification. The above methods all used the SVM classifier with different feature extraction methods. Li et al. [15] also used GLCM features but the classifier is the random forest. Wu et al. [24] calculated the shape, gray, and texture of nodules and used the random forest too. Then, Rodrigues et al. [25] used the structural co-occurrence matrix-based method to extract features from nodules and

doi:10.6342/NTU202104149

compared three different classifiers: multilayer perceptron (MLP), SVM and k-nearest neighbors (KNN). Netto et al. [26] used Getis spatial autocorrelation statistics and its accumulated forms. Zinovev et al. [27] used the decision tree for classification. Lee et al. [28] used a two-step approach for feature selection and classifier ensemble. Ferreira [29] described the nodule through textures extracted from a co-occurrence matrix obtained from the nodule volume and margin sharpness extracted from perpendicular lines drawn over the borders on all nodule slices. Farahani et al. [30] extracted statistical and morphological features from nodule candidates and an ensemble of three classifiers consisting of MLP, KNN, and SVM was used to classify lung nodules.

Nowadays, deep learning approaches have been applied in computer vision successfully. Especially, using deep CNNs for image recognition has shown outstanding performances, like VGG, ResNet, and so forth [17, 31]. Besides, some researchers also used deep CNNs for medical image analysis. A significant advantage of using deep CNNs is that we do not need to extract the handcrafted features like traditional image processing methods because the deep CNNs can learn discriminative features from data directly. Hence, we can use an end-to-end model to execute feature extraction and nodule classification at the same time. There are many input formats such as 2D or 3D and segmented or non-segmented images of lung nodules.

Based on 3D and non-segmented images, Dai et al. [32] proposed a unique 3D CNN which was modified from 3D-DenseNet-40. Ren et al. [13] developed a novel manifold regularized classification deep neural network (MRC-DNN) to perform classification based on the manifold representation of lung nodule images. Zhang et al. [9] used the squeeze-and excitation network and aggregated residual transformations (SE-ResNeXt) for lung nodule classification. Fu et al. [33] referred to InceptionNet, ResNet, DenseNet and SENet and designed 3D-SE-IRNet and 3D-SE-CDNet to

doi:10.6342/NTU202104149

improve the classification ability of Inception-ResNet and CondenceNet, respectively. Liu et al. [34] thought that constructing a robust classification model with conventional deep learning-based diagnostic methods was difficult, so they proposed a multi-model ensemble learning architecture based on a 3D convolutional neural network (MMEL-3DCNN), which contains VggNet, ResNet, and InceptionNet.

Another method is based on 2D data. Lyu et al. [35] proposed a multi-level cross residual convolutional neural network (ML-xResNet), which was constructed by three-level parallel ResNets with different convolution kernel sizes to extract multi-scale features of inputs. An et al. [36] proposed a two-stage convolution neural network (2S-CNN) architecture. Besides, some researchers decomposed a 3D volume image into multiple 2D planar images. Su et al. [37] argued that a collection of 3D views could be highly informative for 3D shape recognition and proposed a new CNN architecture that combines information from multiple views of 3D data, which offered even better recognition performances.

In the lung nodule classification task, anatomical planes were used most often. Nibali et al. [14] used three 2D planar views (axial, sagittal, and coronal) instead of the full 3D volume as input and used the ResNet architecture as the basis with pretraining and curriculum learning. Sahu et al. [11] proposed a lightweight multi-section CNN for lung nodule classification. The main idea was to aggregate information from multiple cross-sections using a data driven method. Al-Shabi et al. [10] proposed a Gated-Dilated (GD) network to classify nodules. The input data were different from the above works. They used the margin coordinates of lung nodules annotated by radiologists to segment the region of interest (ROI) images and generated nodule-only images as the input. Images were used as input units instead of nodules. They also proposed a better model called Deep Local-Global network [12].

doi:10.6342/NTU202104149

There are some works using handcrafted features with DCNNs. Xie et al. [38] proposed a multiview knowledge-based collaborative (MV-KBC) deep model. They used not only multiple views (transverse, sagittal, coronal, and six diagonal planes) but also multiple appearances (overall appearance, heterogeneity in voxel values and heterogeneity in shapes). Then, the authors finetuned three pretrained ResNet-50 models to separate malignant from benign nodules. Further, they combined multiview knowledge-based collaborative learning with the semi-supervised adversarial classification (SSAC) model, which was trained by using both labeled and unlabeled data to improve accuracy of lung nodule classification [39]. In addition, there are some works that combined deep learning and traditional machine learning. Zhu et al. [40] used a 3D dual path network (3D DPN) and a gradient boosting machine (GBM) for lung nodule classification. Zhang et al. [41] used handcrafted features, deep learning, and machine learning methods. They extracted features using 3D deep DPN, local binary pattern (LBP), and HOG features and differentiated malignant from benign nodules using the GBM.

According to the above mentioned methods, data preprocessing and model construction are considerably essential for lung nodule classification. The rest of this thesis is organized as follows: Section 3 describes the prior knowledge for our studies and works. Materials and Method related to the data acquisition and model architecture is introduced in Section 4. Then, Section 5 shows the experimental details, evaluation methods, and results. Finally, Section 6 concludes this work and discusses the future work.

doi:10.6342/NTU202104149

Chapter 3 Background

3.1 CT Images

3.1.1 CT Scan

CT, also called computed axial tomography (CAT), consists of a series of X-ray images taken from different sources to produce tomographic scans [42]. The pioneering CT machines were independently invented by South Africa American physicist Allan M. Cormack and British electrical engineer Godfrey N. Hounsfield. A single CT scan is 2D and a sequence of 2D scans are used to construct a 3D image volume. Nowadays, the cross-sectional images are widely used for medical diagnosis and analysis. The reason for a CT scan is used for detecting lung cancer is that it can be used for detecting both acute and chronic changes in the lung parenchyma, but normal 2D X-rays cannot show the defects.

3.1.2 Hounsfield Unit

The Hounsfield unit (HU) also called CT number named after Godfrey N. Hounsfield is usually used in CT scans. The range of HU in CT is generally [-1024, 3071] to measure radiodensity of the human body.

The HU is derived from a linear transform of the originally measured linear attenuation coefficient, which is defined in Eq. 3-1.

$$\mathrm{HU} = 1000 \times \frac{\mu - \mu_{water}}{\mu_{water} - \mu_{air}} \qquad \text{Eq. 3-1}$$

where μ_{water} represents the transformed water and the radiodensity of distilled water

doi:10.6342/NTU202104149

is defined as zero HU at the standard pressure and temperature (STP), μ_{air} represents the transformed air and the radiodensity of air at the STP is defined as -1000 HU [43]. The HU values of other substances are showed in Table 3-1.

Table 3-1 HU list [43].

Substance	HU
Air	-1000
Lung	-700 to -600
Fat	-120 to -90
Water	0
Blood	+13 to +75
Muscle	+35 to +55
Soft tissue on contrast CT	+100 to +300
Cortical bone	+500 to +1900

3.1.3 **Database**

The database used in this work is from the Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) [8] initiated by National Cancer Institute (NCI). It consists of diagnostic lung cancer screening thoracic CT scans with annotations. It is an open source database for the development of CAD methods for lung cancer diagnosis. There are seven centers and eight medical imaging companies cooperating to create the database. It contains 1010 patients with 1018 medical cases

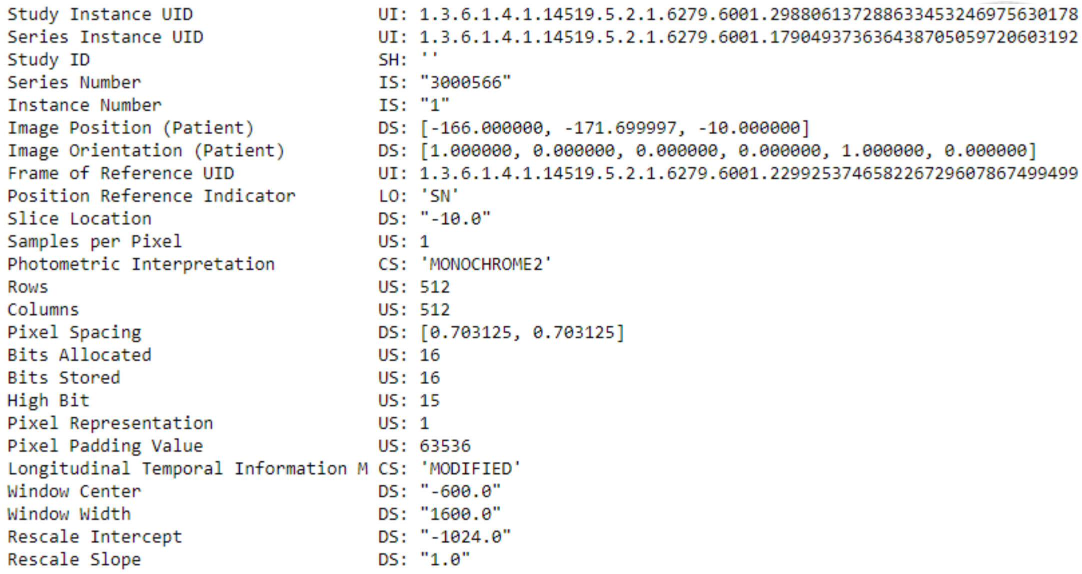
doi:10.6342/NTU202104149

with CT scans, as shown in Table 3-2. Each case contains thoracic CT scans and an XML file.

Table 3-2 The detail description of LIDC-IDRI.

Collection Statistics	updated 3/32/2012
Image size	124 GB
Modalities	CT (computed tomography) CR (computed radiography) DX (digital radiography) SEG (Segmentation)
Number of patients	1010
Number of Series	CT: 1018 CR: 53 DX: 237 SEG: 90

The CT scans are in DICOM format, which is the standard for medical images and related data. Each DICOM file consists of a header and raw data. Fig. 3-1 shows some elements in a DICOM file. The XML file records the results of a two-phase image annotation process by four experienced thoracic radiologists. In the first blinded-read phase, radiologists reviewed CT scans individually and marked lesions. Then, they reviewed their own marks and compared with anonymized marks from other radiologists to obtain a final annotation.



```
Study Instance UID                    UI: 1.3.6.1.4.1.14519.5.2.1.6279.6001.298806137288633453246975630178
Series Instance UID                   UI: 1.3.6.1.4.1.14519.5.2.1.6279.6001.179049373636438705059720603192
Study ID                              SH: ''
Series Number                         IS: "3000566"
Instance Number                       IS: "1"
Image Position (Patient)              DS: [-166.000000, -171.699997, -10.000000]
Image Orientation (Patient)           DS: [1.000000, 0.000000, 0.000000, 0.000000, 1.000000, 0.000000]
Frame of Reference UID                UI: 1.3.6.1.4.1.14519.5.2.1.6279.6001.229925374658226729607867499499
Position Reference Indicator          LO: 'SN'
Slice Location                        DS: "-10.0"
Samples per Pixel                     US: 1
Photometric Interpretation            CS: 'MONOCHROME2'
Rows                                  US: 512
Columns                               US: 512
Pixel Spacing                         DS: [0.703125, 0.703125]
Bits Allocated                        US: 16
Bits Stored                           US: 16
High Bit                              US: 15
Pixel Representation                  US: 1
Pixel Padding Value                   US: 63536
Longitudinal Temporal Information M CS: 'MODIFIED'
Window Center                         DS: "-600.0"
Window Width                          DS: "1600.0"
Rescale Intercept                     DS: "-1024.0"
Rescale Slope                         DS: "1.0"
```

Fig. 3-1 Example of partial elements in a DICOM file

The lesions are classified into three categories ("nodule" > or = 3mm, "nodule" < 3 mm and "non-nodule" > 3mm). Only the nodules greater than 3 mm are annotated the likelihood of malignancy. There are five levels: Level 1: highly unlikely for cancer. Level 2: moderately unlikely for cancer. Level 3: indeterminate likelihood. Level 4: moderately suspicious for cancer. Level 5: highly suspicious for cancer.

3.2 Image Processing

3.2.1 Anatomical plane

An anatomical plane used to describe the location of structures or the direction of movements is a hypothetical plane used to transect the body. There are three planes that are commonly used: axial, coronal, and sagittal as shown in Fig. 3-2. The axial plane divides the body into upper and lower portions, the coronal into front and back portions, and the sagittal into left and right portions [44]. Nowadays, volume acquisition CT has become routine and easy. The three different planes of axial, coronal, and sagittal planes can reconstruct the 3D data entirely. Viewing the anatomy and pathology in these three

doi:10.6342/NTU202104149

planes is considerably helpful when evaluating the disease of a patient [45].

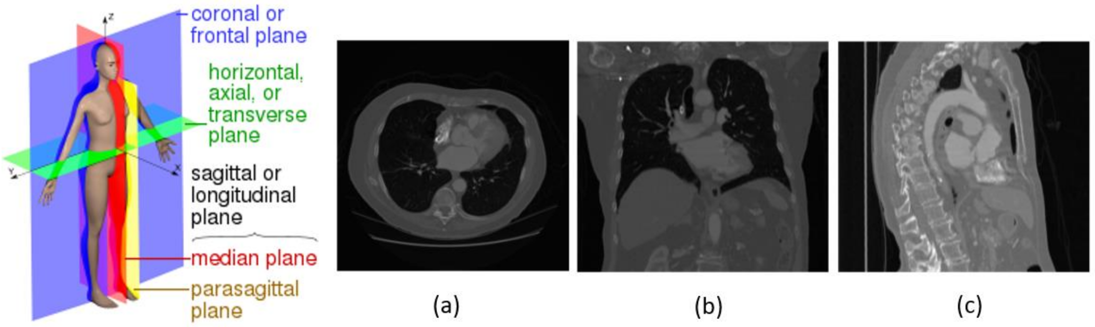

coronal or frontal plane
horizontal, axial, or transverse plane
sagittal or longitudinal plane
median plane
parasagittal plane
(a)
(b)
(c)

Fig. 3-2 Anatomical planes [44] for viewing lung CT scans. (a) axial plane. (b) coronal plane. (c) sagittal plane.

3.2.2 GLCM

GLCM was proposed by Haralick et al. in 1973 [46]. GLCM is computed from the statistical distribution of observed combinations of intensities at specified positions relative to each other in the image [47]. It is used as a method for texture analysis with many applications especially in medical image analysis [48].

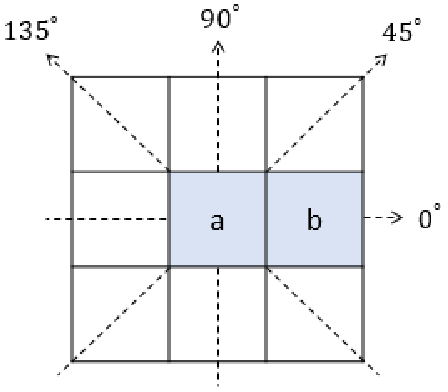
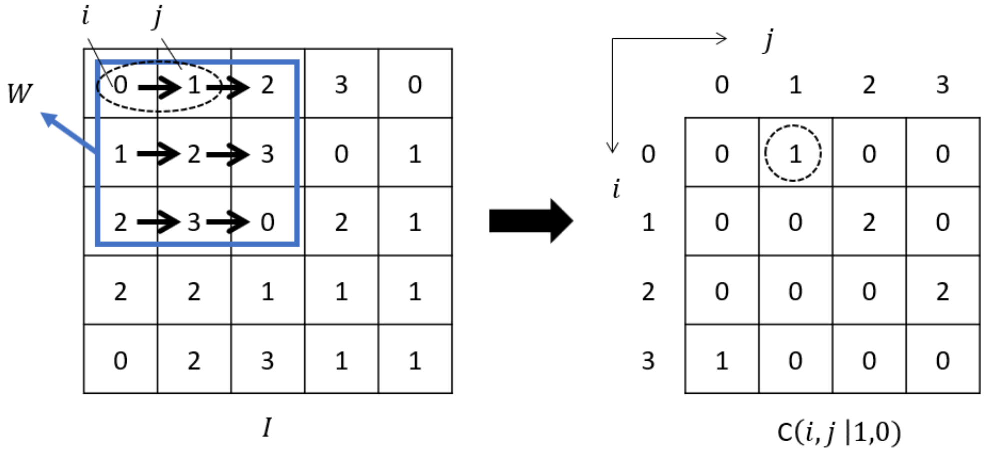


Fig. 3-3 The relationship of adjacent pixels (d= 1) with different angles.

To calculate GLCM, we need to first determine three parameters: the size of the sliding window, displacement distance (d), and angle (θ). The common position relationship is shown in Fig. 3-3 where a, b is an adjacent pixel pair with $d = 1$. A GLCM is a matrix whose size is equal to the gray level of the image. The value of matrix $C(i,j|\Delta_x,\Delta_y)$ is the relative frequency of two pixels separated by a distance (Δ_x, Δ_y), which is defined by (d,θ). The occurrence is based on their intensity: one with intensity ‘i’ and the other with intensity ‘j’. Given a $X \times Y$ gray-level image I with an intensity range $[0, L-1]$, its GLCM size is $L \times L$. We can obtain the GLCM with a $\mathrm{K} \times \mathrm{K}$ sliding window W using

$$C(i,j|\Delta_x,\Delta_y) = \begin{cases} \sum_{x=0}^{K-1}\sum_{y=0}^{K-1} 1, & \text{if } W(x,y) = i \text{ and } W(x+\Delta_x, y+\Delta_y) = j \\ 0, & \text{otherwise} \end{cases} \qquad \text{Eq. 3-2}$$

where $W(x,y)$ is a gray-level value. In the following example, we set the size of the sliding window as 3×3, $\mathrm{d} = 1$, and $\theta = 0°$ to calculate the GLCM, as shown in Fig. 3-4.



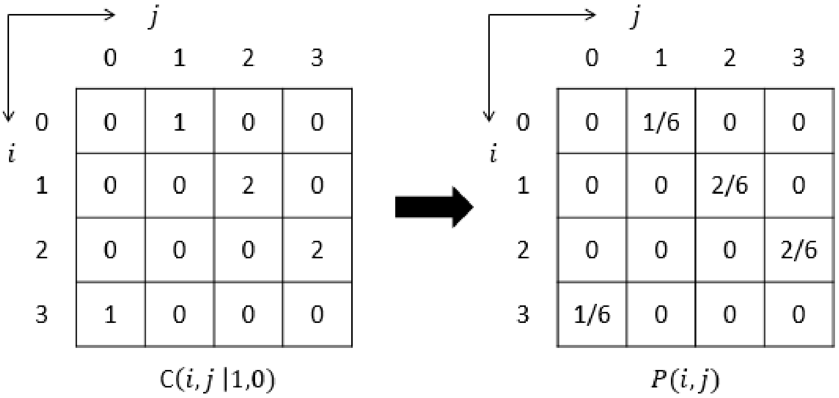
0	1	2	3	0
1	2	3	0	1
2	3	0	2	1
2	2	1	1	1
0	2	3	1	1

	0	1	2	3
0	0	1	0	0
1	0	0	2	0
2	0	0	0	2
3	1	0	0	0

Fig. 3-4 Example of the GLCM construction in the specified sliding window.

Finally, we normalize the above matrix to get the probability matrix $P(i,j)$:

$$P(i,j) = \frac{C\left(i,j\middle|\Delta_x,\Delta_y\right)}{\sum_{i=0}^{X-1}\sum_{j=0}^{Y-1} C\left(i,j\middle|\Delta_x,\Delta_y\right)} \qquad \text{Eq. 3-3}$$



i \ j	0	1	2	3
0	0	1	0	0
1	0	0	2	0
2	0	0	0	2
3	1	0	0	0

C$(i,j\,|1{,}0)$

i \ j	0	1	2	3
0	0	1/6	0	0
1	0	0	2/6	0
2	0	0	0	2/6
3	1/6	0	0	0

$P(i,j)$

Fig. 3-5 Probability matrix computation.

Fig. 3-5 illustrates a probability matrix computation example. After getting the probability matrix, statistical features are acquired by different approaches according to different requirements. There are totally twenty-eight texture features based on GLCM [46]. In our experiments, we only used entropy (ENT) and homogeneity (HOM).

(1) Entropy (ENT)

$$\text{ENT} = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1} P(i,j)\left(-ln\big(P(i,j)\big)\right) \qquad \text{Eq. 3-4}$$

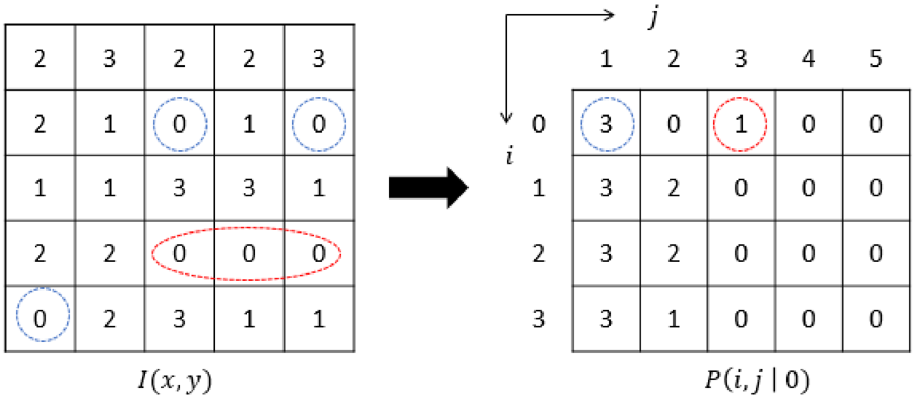
(2) Homogeneity (HOM)

$$\text{HOM} = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1} \frac{P(i,j)}{1+(i-j)^2} \qquad \text{Eq. 3-5}$$

3.2.3 GLRLM

Gray-level run length matrix (GLRLM) is a simpler and more objective texture analysis method. Generally, if the texture is coarser the gray-level run-length is longer, so we can extract texture features by different gray-level run-lengths. Proposed by Galloway in 1975 [49], GLRLM is also used in medical image analysis widely.

The GLRLM element $p(i,j \mid \theta)$ is the number of elements j with the intensity i in a particular direction. For example, Fig. 3-6 shows a 5×5 image with 4 gray levels on the left. By setting $\theta = 0°$, the size of GLRLM is 4×5. There are 3 independent pixels whose intensity is 0 that represents the run-length 1, so $p(0,1) = 3$. Then, there are 3 adjacent pixels whose intensity is all 0 that represents the run-length 3, so $p(0,3) = 1$, as shown in Fig. 3-6.



2	3	2	2	3
2	1	0	1	0
1	1	3	3	1
2	2	0	0	0
0	2	3	1	1

$I(x,y)$

i \ j	1	2	3	4	5
0	3	0	1	0	0
1	3	2	0	0	0
2	3	2	0	0	0
3	3	1	0	0	0

$P(i,j \mid 0)$

Fig. 3-6 GLRLM construction

Particularly, one matrix and two vectors are computed based on the GLRLM.

(1) Gray-level run-length pixel number matrix

$$p_p(i,j) = p(i,j) \cdot j \qquad \text{Eq. 3-6}$$

This matrix records the number of pixels in each run-length j.

(2) Gray-level run-number vector

$$p_g(i) = \sum_{j=1}^{R} p(i,j) \qquad \text{Eq. 3-7}$$

where R is the maximum run-length. This vector represents the summation of the j direction.

(3) Run-length run-number vector

$$p_r(j) = \sum_{i=0}^{L} p(i,j) \qquad \text{Eq. 3-8}$$

where L is the range of the gray-level. This vector represents the summation of the i direction.

There are 11 features derived from the above matrix and vectors, and 4 features are used in this thesis.

(1) Gray level non-uniformity (GLN)

$$\mathrm{GLN} = \frac{1}{n_r}\sum_{i=1}^{L}\left(\sum_{j=1}^{R} p(i,j)\right)^2 = \frac{1}{n_r}\sum_{i=1}^{L} p_g\,(i)^2 \qquad \text{Eq. 3-9}$$

where n_r is the summation of the number of run-length:

doi:10.6342/NTU202104149

$$n_r = \sum_{i=1}^{L} \sum_{j=1}^{R} p(i,j)$$

Eq. 3-10

(2) Run-length non-uniformity (RLN)

$$\mathrm{RLN} = \frac{1}{n_r} \sum_{j=1}^{R} \left(\sum_{i=1}^{L} p(i,j) \right)^2 = \frac{1}{n_r} \sum_{j=1}^{R} p_r\,(j)^2$$

Eq. 3-11

(3) Run percentage (RP)

$$\mathrm{RP} = \frac{n_r}{n_p}$$

Eq. 3-12

where n_p is the number of all pixels that are considered:

$$n_p = L \times R$$

Eq. 3-13

(4) Short run emphasis (SRE)

$$\mathrm{SRE} = \frac{1}{n_r} \sum_{i=1}^{L} \sum_{j=1}^{R} \frac{p(i,j)}{j^2} = \frac{1}{n_r} \sum_{j=1}^{R} \frac{P_r(j)}{j^2}$$

Eq. 3-14

3.2.4 Tamura feature

There a six Tamura texture features [50] based on human vision and psychology, such as Coarseness, Contrast, Directionality, Line-likeness, Regularity, and Roughness. The first three features are associated with human vision and more distinctive than the last three ones. In our study, we are only interested in the Coarseness feature.

Coarseness is used to measure the grain size of image features. A bigger coarse value represents a coarser image. The algorithm is described in the following:

(1) Calculate the average intensity in a $2^k \times 2^k$ region, where $\mathrm{k} \in [0, \dots, 5]$, centered at (x, y) using:

$$A_k(x,y) = \frac{1}{2^{2k}} \sum_{i=x-2^{k-1}}^{x+2^{k-1}-1} \sum_{j=y-2^{k-1}}^{y+2^{k-1}-1} I(i,j) \qquad \text{Eq. 3-15}$$

where $I(i,j)$ is the gray-level value of pixel (i,j).

(2) For every pixel (x, y), calculate the difference of symmetric points in the horizontal and vertical directions.

Horizontal direction:

$$E_{k,h}(x,y) = |A_k(x + 2^{k-1}, y) - A_k(x - 2^{k-1}, y)| \qquad \text{Eq. 3-16}$$

Vertical direction:

$$E_{k,v}(x,y) = |A_k(x, y + 2^{k-1}) - A_k(x, y - 2^{k-1})| \qquad \text{Eq. 3-17}$$

(3) Find the best range K for every pixel from the highest output value:

doi:10.6342/NTU202104149

$$S_{best}(x,y) = 2^K \qquad \text{Eq. 3-18}$$

where K, derived from the following equation, is the range of the maximum difference in the horizontal or vertical direction:

$$K = max\left(E_{k,h}(x,y), E_{k,v}(x,y)\right) \qquad \text{Eq. 3-19}$$

(4) Calculate the average of S_{best} to get the Coarseness F_{crs}:

$$F_{crs} = \frac{1}{M \times N}\sum_{i=1}^{M}\sum_{j=1}^{N} S_{best}(i,j) \qquad \text{Eq. 3-20}$$

where M and N denote the width and length of the image, respectively.

3.3 Deep Learning

3.3.1 Convolutional neural network

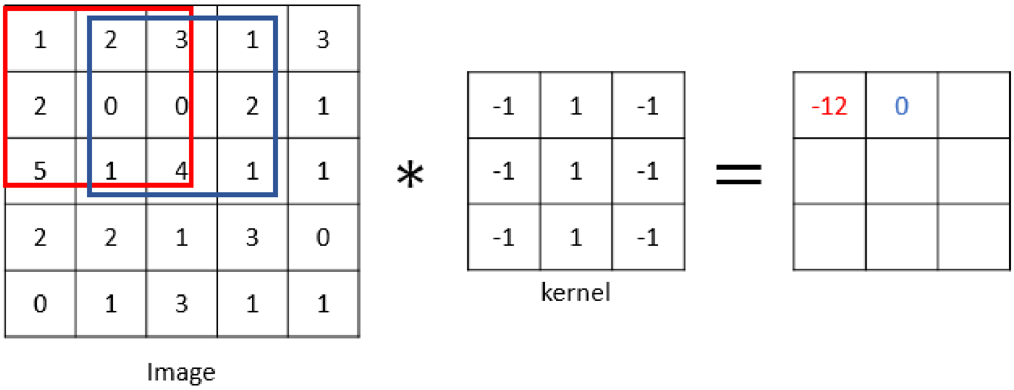
The convolutional neural network (CNN) was inspired by biological neural networks because the connectivity pattern between neurons is similar to the organization of the animal visual cortex. There are two main reasons to explain why the convolution is suitable for image recognition: (1) Some patterns are much smaller than the whole image, so a neuron does not have to see the whole image to discover the pattern. Through the convolution operation, it can extract the local features from a whole image in a local RF whose size is defined by the kernel size. (2) Weight sharing is also a significant advantage of the convolution layer. It makes the CNNs require

doi:10.6342/NTU202104149

lower computational resources and less time than the deep neural networks (DNNs) with only fully connected (FC) layers.

The convolution operation is shown in Fig. 3-7. Each convolution kernel which is also called filter is convolved across the width and height of the input images. It employs an element-wise product between two matrices and summarizes the results from the multiplication.

Generally, the convolution layer is a feature extraction method. Compared to traditional image classification algorithms, CNN uses relatively little pre-processing because CNN can automatically learn how to optimize the filter through training. Recently, plenty CNN models have been applied in image, video, and audio recognition, recommender system, natural language processing, image classification, image segmentation, and medical image analysis [51].



1	2	3	1	3
2	0	0	2	1
5	1	4	1	1
2	2	1	3	0
0	1	3	1	1

Image

-1	1	-1
-1	1	-1
-1	1	-1

kernel

-12	0	

Fig. 3-7 Example of discrete convolution operations.

3.3.2 **Multiview CNN**

In the field of computer vision, most methods for image recognition were based on 2D images in the beginning. Afterwards, 3D images were also used for image recognition because some researchers thought that there was more spatial information in a 3D image. However, the 3D images tended to be high-dimensional to cause the curse of dimensionality which made the classifier prone to overfitting and took a lot of time while training. Su et al. [37] proposed a novel CNN architecture, as shown in Fig. 3-8, which combines multiple views of a 3D shape into a single and compact shape descriptor and outperformed the classifier built directly on the 3D images. They found that a collection of 2D views could be highly informative for 3D shape recognition, which was applied to the emerging CNN architecture.

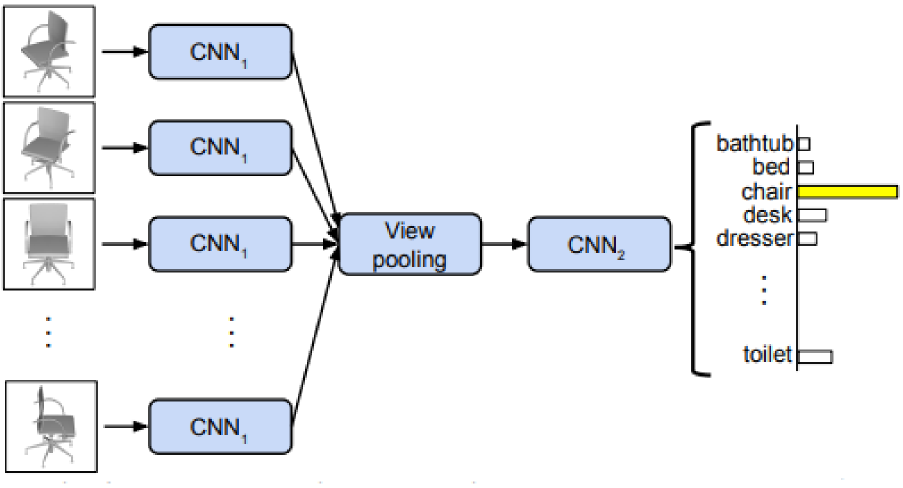


Fig. 3-8 MVCNN [37].

3.3.3 Residual Learning

Because of the degradation problem, which means that not all systems are easy to optimize, a deeper model has larger training error than the shallower counterpart. To solve this problem, He et al. [17] proposed a deep residual learning framework. The residual block is shown in Fig. 3-9. To reduce and increase dimensions easily and decrease the computational complexity, they also designed the bottleneck architecture

doi:10.6342/NTU202104149

as shown on the right of Fig. 3-10.

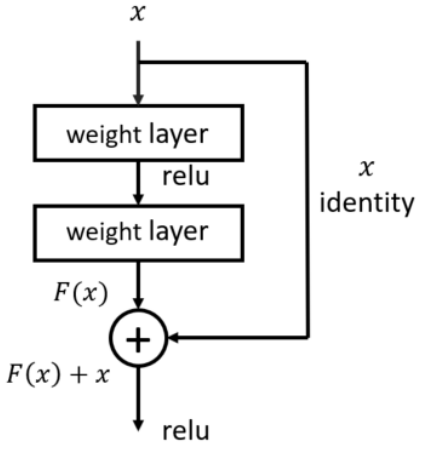


Fig. 3-9 Residual block [17].

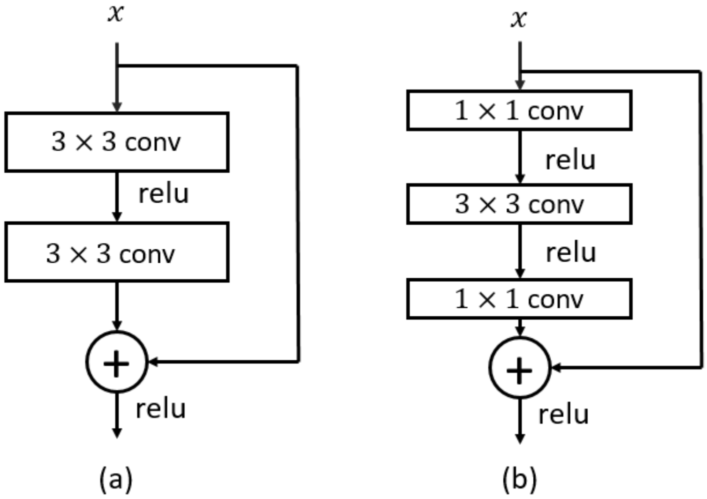


Fig. 3-10 (a) Basic residual block and (b) Bottleneck residual block [17].

Instead of making the stacked layers fit a direct underlying mapping, the authors let these layers fit a residual mapping. Assuming that $\mathrm{H(x)}$ is an original underlying mapping with an input x. The authors let the stacked nonlinear layers fit another mapping (residual mapping) with

doi:10.6342/NTU202104149

$$F(x) := H(x) - x \qquad \text{Eq. 3-21}$$

where $\mathrm{F(x)}$ is the same as the original underlying mapping. Then the original mapping would be recast into $\mathrm{F(x) + x}$. According to the result of experiments, it is easier to optimize the residual mapping than the original mapping in DCNNs.

The reformulation of $\mathrm{F(x) + x}$ can be employed by feedforward neural networks with a shortcut connection, as shown in Fig. 3-9. In [17], the shortcut connections simply execute the identity mapping and their outputs are added to the outputs of the stacked layers. The identity shortcut connection does not add extra computational complexity and parameters.

3.3.4 Selective Kernel

It is well-known that the RF sizes of neurons in the same region are different in the visual cortex, so the neurons are able to acquire multi-scale spatial information during the same process. Therefore, in a standard CNN, the multi-scale convolutional architecture has been adopted widely, such as InceptionNets. However, Li et al. [18] used adaptive changing of RF sizes. There were many experiments proving that the RF sizes of neurons are not fixed but modulated by stimulus. In [18], they presented a nonlinear method to aggregate information from multiple kernels to realize the adaptive RF sizes of neurons. They built a state-of-the-art model called Selective Kernel Network (SKNet). The core architecture is SK Convolution, which makes the neurons adaptively adjust their RF sizes. It is an automatic selection operation among multiple kernels with different kernel sizes. There are three operations: Split, Fuse, and Select. Fig. 3-11 shows a two-branch case. The Split operation contains multiple paths with different kernel sizes, which correspond to different RF sizes of neurons, to obtain the

doi:10.6342/NTU202104149

local information. The Fuse operation combines and aggregates the information from multiple paths to obtain a global information for selection weights. The Select operation aggregates the feature maps from different kernel sizes according to the selection weights.

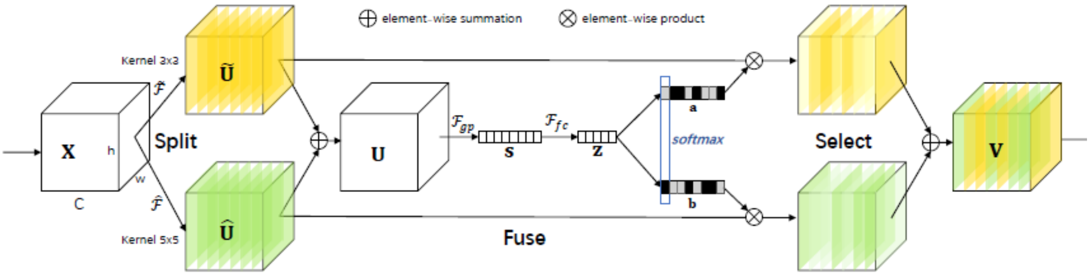


Fig. 3-11 SK Convolution [18].

3.3.5 Activation Function

The basic unit of neural networks is a neuron, whose behavior can be expressed as a linear function. In the real world, data may be various and complicated, so we need a more complicated system to fit data other than a linear system. An input is processed by a nonlinear transformation through an activation function. Therefore, an activation function is applied to neural networks using:

$$f(x) = f\left(\sum_{i} w_i x_i + b\right) \qquad \text{Eq. 3-22}$$

where f denotes a nonlinear function, w_i is the weight of input x_i, and b is the bias.

In the network architecture, Rectified Linear Unit (ReLU) is the most used activation function, which is formulated as:

$$f(x) = max(0, x) \quad \text{Eq. 3-23}$$

We can easily realize ReLU from Fig. 3-12. The slope of ReLU is 1 for all positive values and 0 for negative values. The advantages of ReLU are sparse activation and efficient computation. The most crucial part is that it makes gradient descent and back propagation efficient to prevent gradient vanishing.

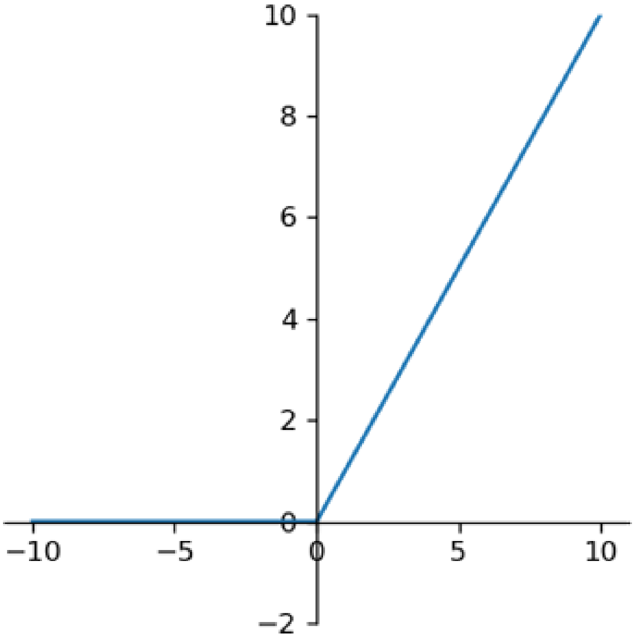


Fig. 3-12 Plots of ReLU.

In the classification task, networks should output the probability to determine which class the input belongs to. Although the sigmoid function may cause gradient vanishing (not suitable in the model), it is usually applied to the last layer to map the prediction to probability. According to Eq. 3-24 and Fig. 3-13, the output values of the sigmoid function are between 0 and 1.

doi:10.6342/NTU202104149

$$f(x) = \frac{1}{1 + e^{-x}} \quad \text{Eq. 3-24}$$

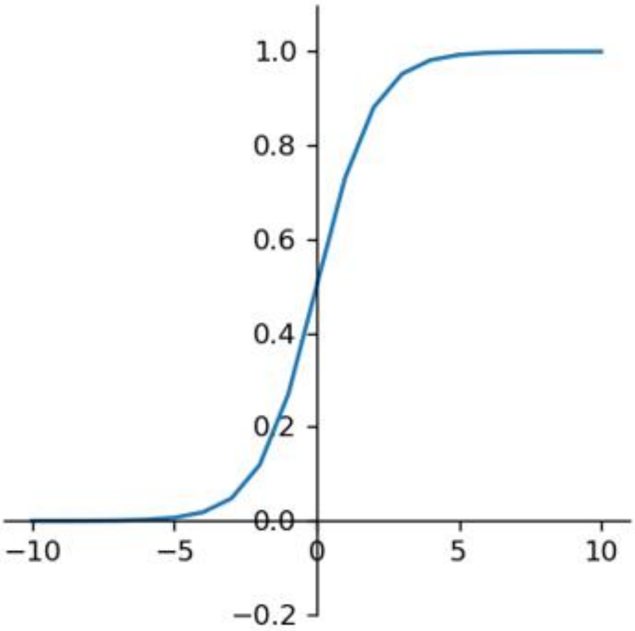


Fig. 3-13 Plots of the sigmoid function.

3.3.6 **Loss function**

The loss function is the core base in deep learning. It is an approach of evaluating how well the model fits the dataset. In the optimization problem, we usually look for the minimization of a loss function, so selecting an appropriate loss function is extremely significant.

In the classification task, the Cross Entropy (CE) loss is used mostly. It measures the difference between the estimated probability distribution and the true distribution, which is expressed as:

doi:10.6342/NTU202104149

$$\mathrm{CE} = -\sum_{i}^{C} y_i log(\hat{y}_i) \qquad \text{Eq. 3-25}$$

where C is the number of classes, y_i is the ground truth and $\hat{y}_i$ is the prediction. It is suitable for multiple classification. When C is 2, which means that there are only two classes, the equation is rewrote in Eq. 3-26, which is called the Binary Cross Entropy (BCE) loss. It is only used in the binary classification task.

$$\mathrm{BCE} = -\frac{1}{N}\sum_{n=1}^{N} y_n log\hat{y}_n + (1 - y_n)log(1 - \hat{y}_n) \qquad \text{Eq. 3-26}$$

where N is the number of data, y_n is the ground truth and $\hat{y}_n$ is the prediction. If y_n and $\hat{y}_n$ are equal, BCE would be 0, otherwise BCE is a positive number. When the difference between y_n and $\hat{y}_n$ is higher, the loss is higher too.

doi:10.6342/NTU202104149

Chapter 4 Materials and Methods

In this section, we describe the proposed methods for lung nodule classification in detail. There are three main subsections: (1) data acquisition from LIDC-IDRI and data preprocessing which includes nodule segmentation; (2) the proposed model architecture; (3) configuration settings.

4.1 Data acquisition

4.1.1 Data collection

The database we used is LIDC-IDRI [8], which is a public database. There are 1018 cases and 1010 patients. Each case consists of a series of thoracic images and one XML file, which stores annotation and related information such as the UID number, nodule position, and likelihood of malignancy by four radiologists. We obtained the data from The Cancer Imaging Archive (TCIA) [52] and used the LIDC nodule size report [53] from Vision and Image Analysis Group of Cornell University. It provides multiple research groups to use the same size-selected subset of nodules. We chose the nodules which were found by at least 3 radiologists. Nodules were annotated the malignancy levels from 1 to 5 by every radiologist, and we took the median of the malignancy levels in each case. The rating less than 3 is considered a benign nodule. On the contrast, the rating bigger than 3 is considered a malignant nodule. Besides, we discarded the nodules whose median ratings equal to 3 in the binary classification. Finally, we used 698 subjects which contain 1386 nodules. There are 447 benign, 430 malignant, and 509 indeterminate nodules. For binary classification, only 877 nodules were used.

Among the nodules we selected, the two largest sizes of nodules are 32.684 mm and 32.286 mm and the rest nodule sizes are between 3 mm to 32 mm. The distribution of the selected nodule sizes is shown in Fig. 4-1. According to this figure, benign nodules usually have smaller sizes than malignant nodules.

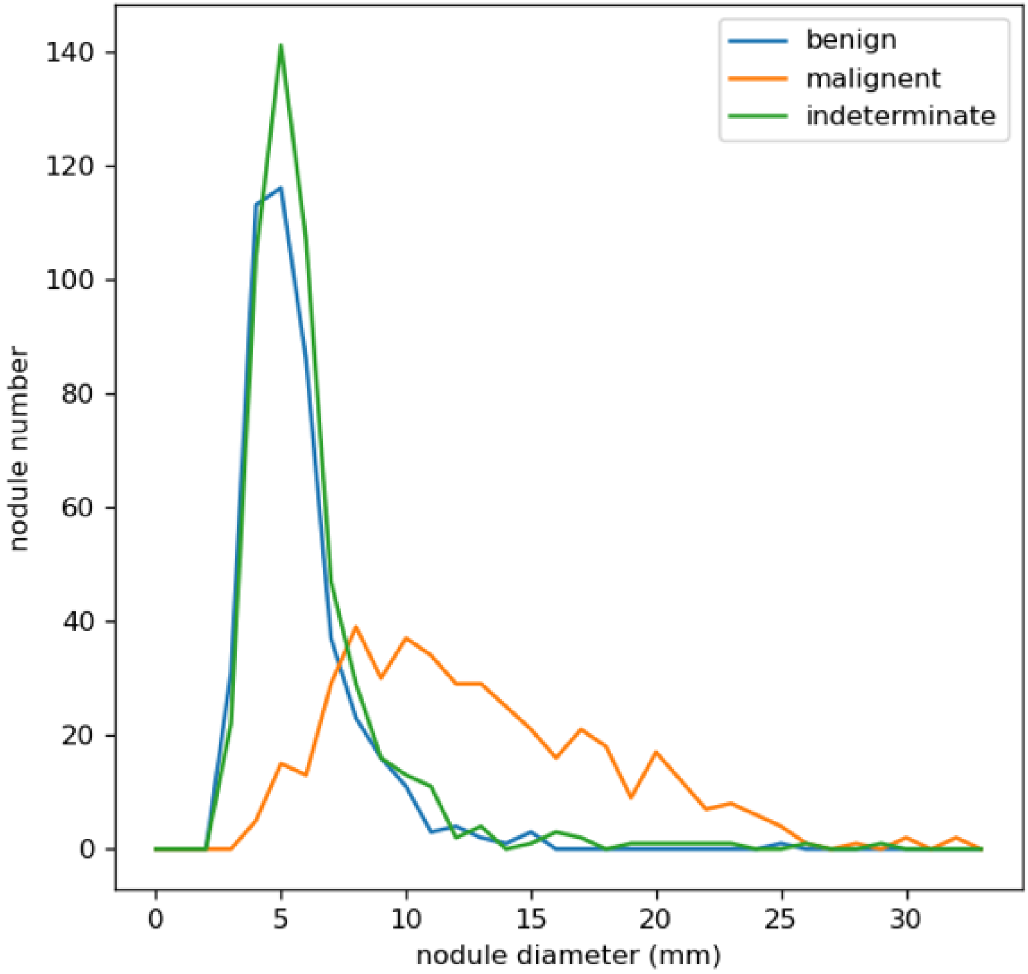


Fig. 4-1 The distribution of the selected nodule sizes.

4.1.2 **Data Preprocessing**

The preprocessing steps involved the extraction of nodules from the lung CT images. Each case contained a series of DICOM files and one XML file which recorded

doi:10.6342/NTU202104149

the positions of nodules and their malignancy level. To acquire the lung CT images, we set out-of-scan pixels from -2000 to 0 and rescaled the pixel values to the HU using the following equation:

$$CT(x,y) = \sum \left(m \times p(x,y)\right) + b \qquad \text{Eq. 4-1}$$

where $p(x,y)$ is an image from the DICOM file directly, $CT(x,y)$ is the image with the HU, m is the rescale slope which is 1 and b is the intercept which is usually -1024 but seldom 0 from the header of the DICOM file. Then we have the images with HU and clipped the range of the HU values to [-1024, 3071].

These lung CT images indicating nodules have the same size of 512×512, but their pixel spacings and slice thicknesses are different. The distribution of the pixel spacing and slice thickness is shown in Fig. 4-2 and Fig. 4-3, respectively. Therefore, in our experiments, all images were resampled to a specified isotropic resolution of $1 \times 1 \times 1\ mm^3$ per voxel by using the cubic spline interpolation. Then we clipped the range of the HU values to [-1000, 400] and normalized them to [0, 1] [12], which is easier for computers to learn and recognize. The formula of normalization is:

$$X_{normalize} = \frac{X - X_{min}}{X_{max} - X_{min}} \qquad \text{Eq. 4-2}$$

where X is the CT image after clipping, X_{max} and X_{min} are the maximum and minimum values in the CT image, and $X_{normalize}$ is the normalized CT image. Next, we picked the lung images with nodules and produced the nodule-shape mask with the coordinates of nodules' contours. We set the pixel values inside the nodule 1 because

doi:10.6342/NTU202104149

the background pixel values are 0. Then, we performed the element-wise product of the lung CT images and their respective masks to segment the nodules from the whole lung, as shown in Fig. 4-4. The background region is too large comparing to the longest nodule diameter, which is 32.684 mm according to Fig. 4-1. To better capture most of selected nodules, we extracted the ROI with a square of $32 \times 32\ mm^2$ [9, 10, 12, 13, 41], as shown in Fig. 4-5. Each nodule has a series of 2D images, which are used to generate a set of 3D images with a shape of $32 \times 32 \times 32$. Finally, to generate multiple views, we employ three 2D anatomical planes, which are axial, coronal, and sagittal planes, from the 3D image volumes as our input [10, 12, 14]. The size of each plane is 32×32, as shown in Fig. 4-6. We saved these images as png format, which consists of 8-bit unsigned integers.

To diversify the input images, we extracted texture features from the above input images, which are called the original images, as shown in Fig. 4-7. Based on the GLCM, we extracted the ENT and HOM features from the original images, as shown in Fig. 4-8 and Fig. 4-9. Based on the GLRLM, we generated the GLN, RLN, RP, and SRE features, as shown in Fig. 4-10 ~ Fig. 4-13. On the basis of Tamura method, Coarseness feature was produced, as shown in Fig. 4-14. While calculating the texture features, we used the 3×3 filter and zero padding to traverse the whole images with stride=1. Afterwards, we concatenated the texture features and the original images in their color channels, as demonstrated in Fig. 4-15. Finally, there were seven kinds of input combinations corresponding to seven texture features.

doi:10.6342/NTU202104149

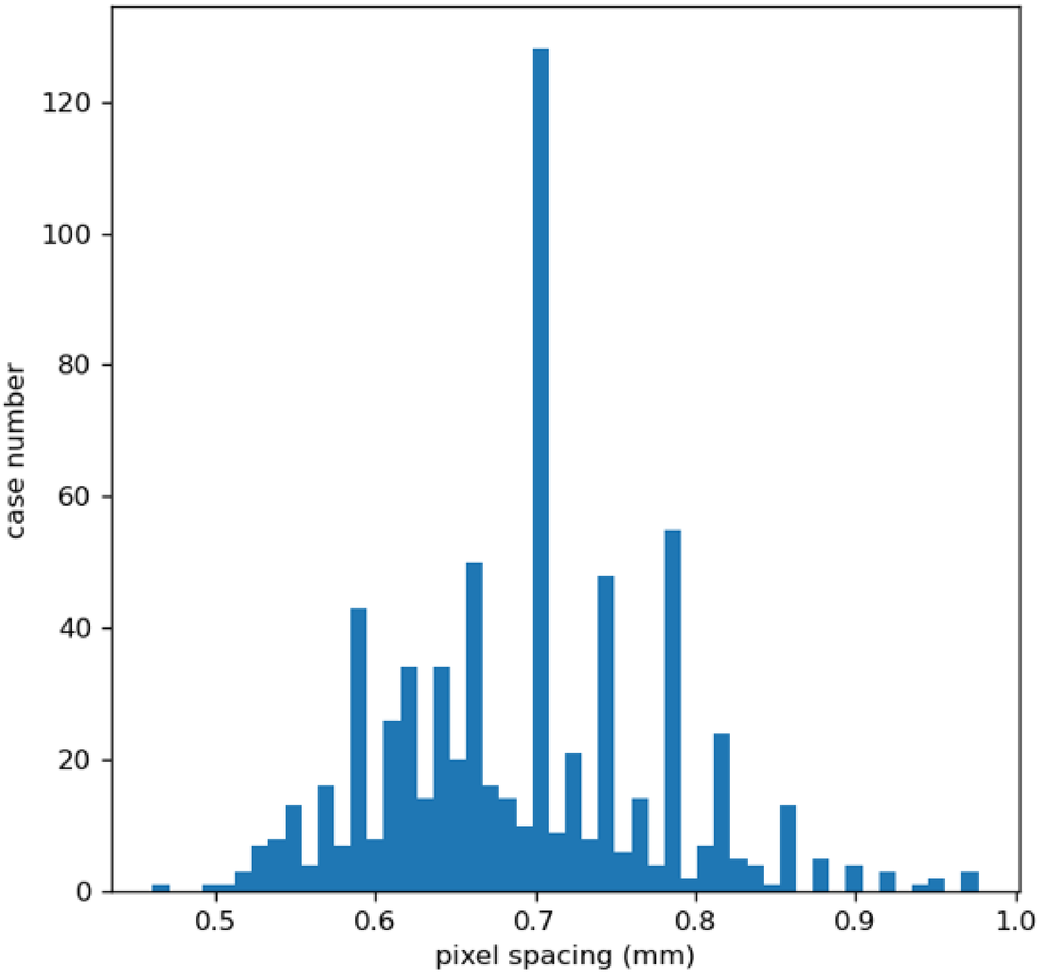


Fig. 4-2 The distribution of the pixel spacing.

doi:10.6342/NTU202104149

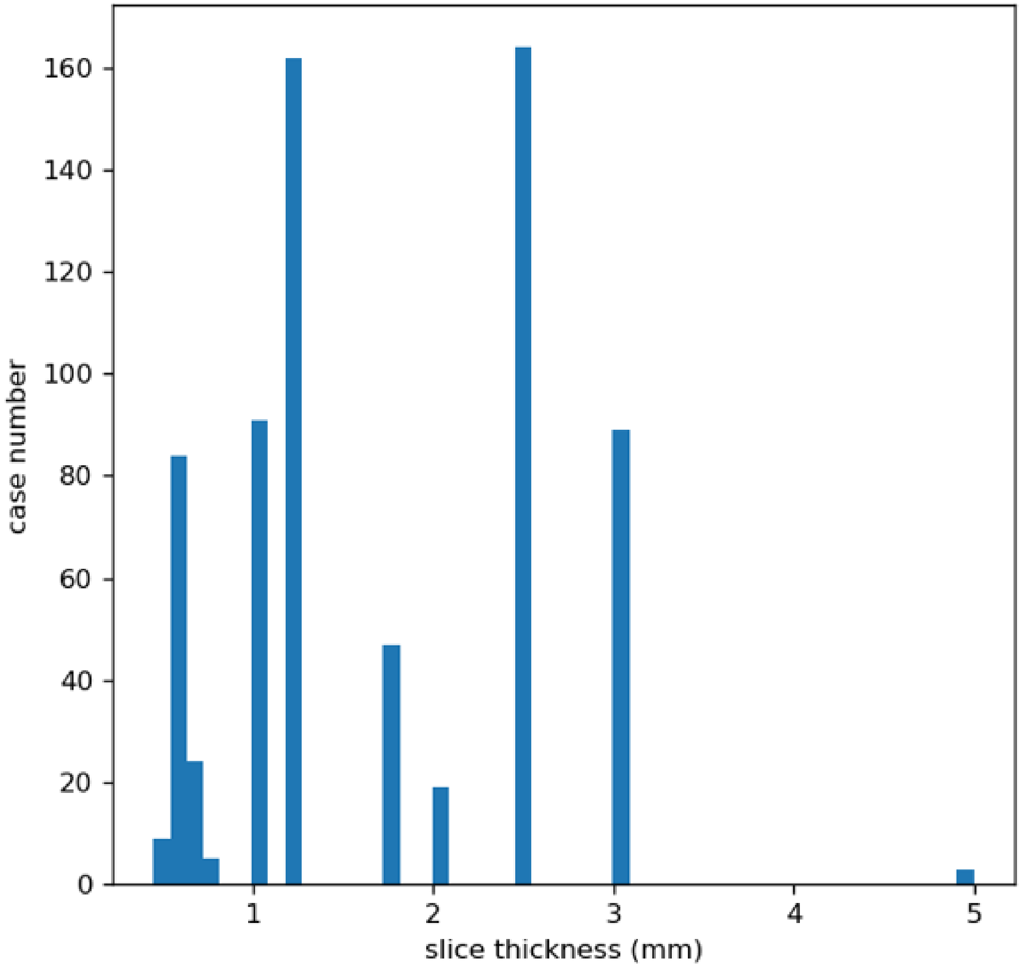


Fig. 4-3 The distribution of the slice thickness.

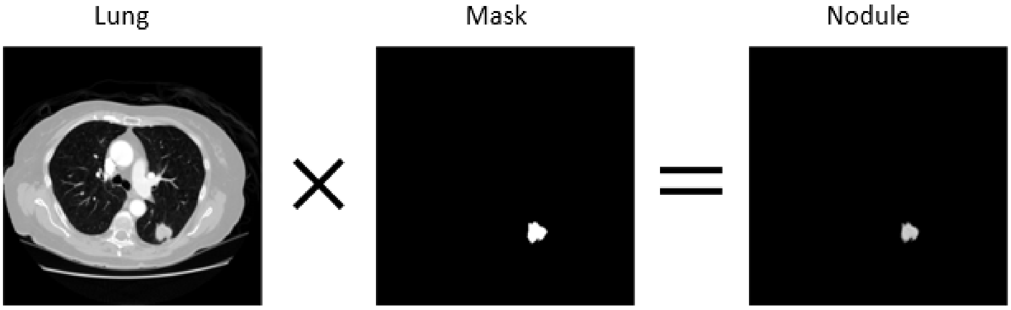


Fig. 4-4 Nodule segmentation.

doi:10.6342/NTU202104149

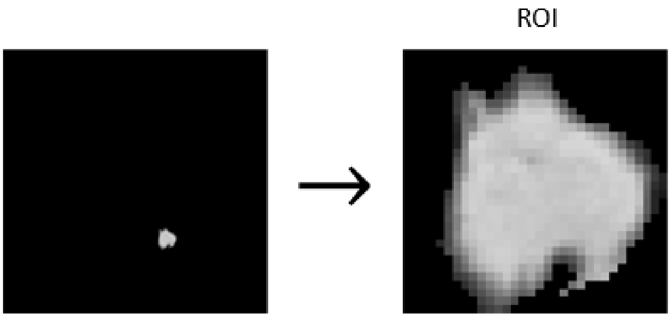


Fig. 4-5 ROI extraction

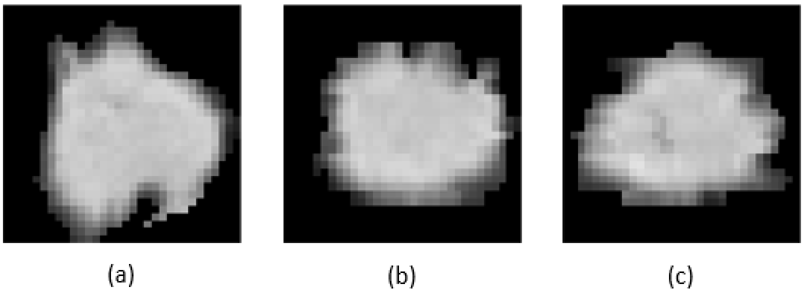


Fig. 4-6 Extracted lung nodule images for recognition.

(a) Axial plane (b) Coronal plane (c) Sagittal plane.

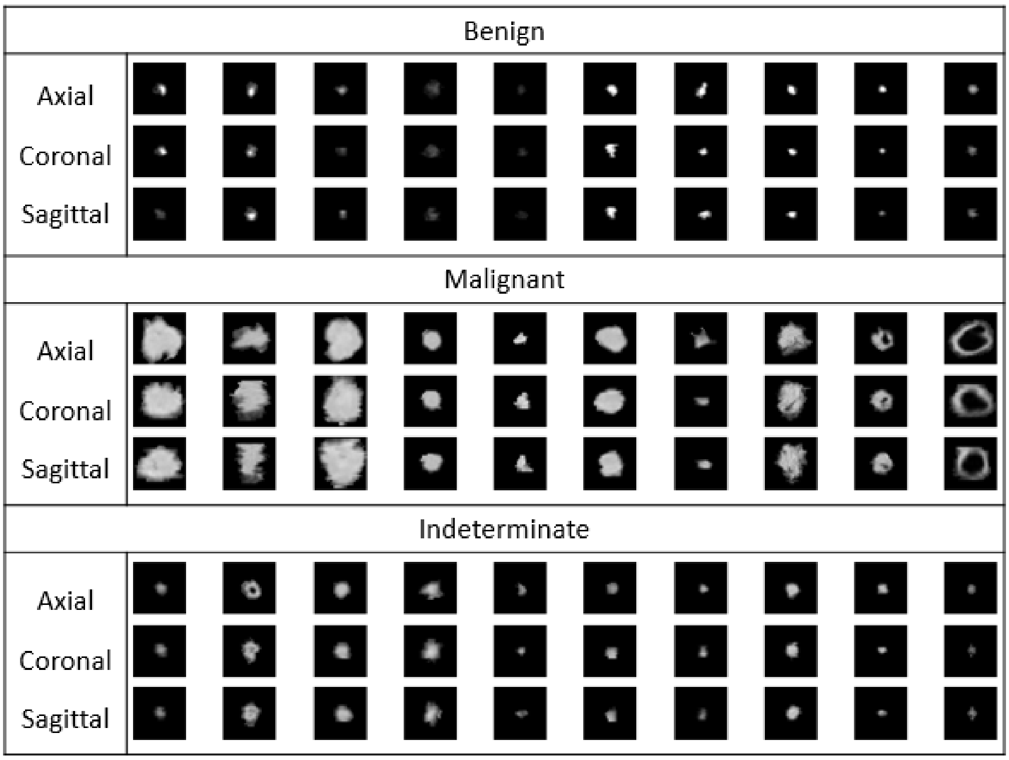


Fig. 4-7 Examples of the original images.

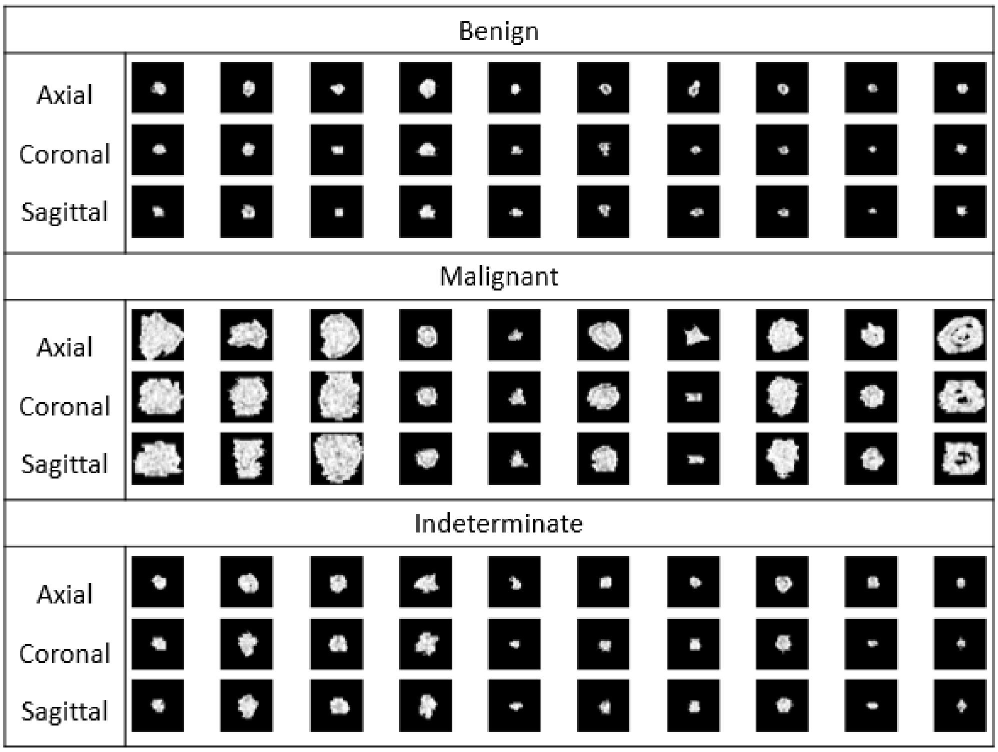


Fig. 4-8 ENT features of Fig. 4-7.

doi:10.6342/NTU202104149

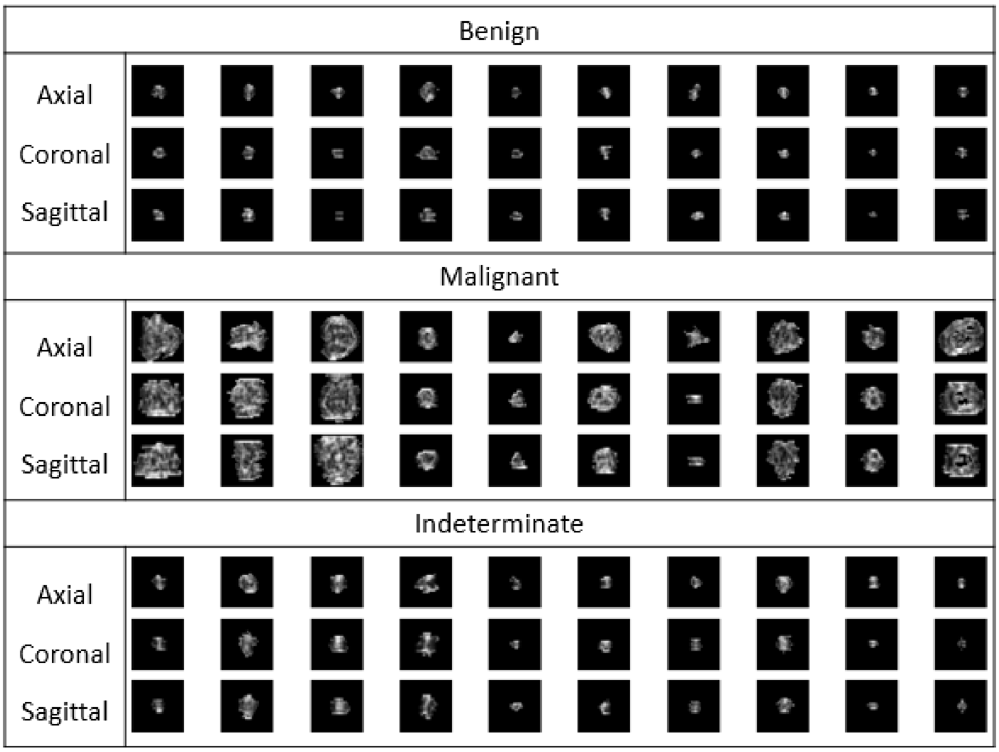


Fig. 4-9 HOM features of Fig. 4-7.

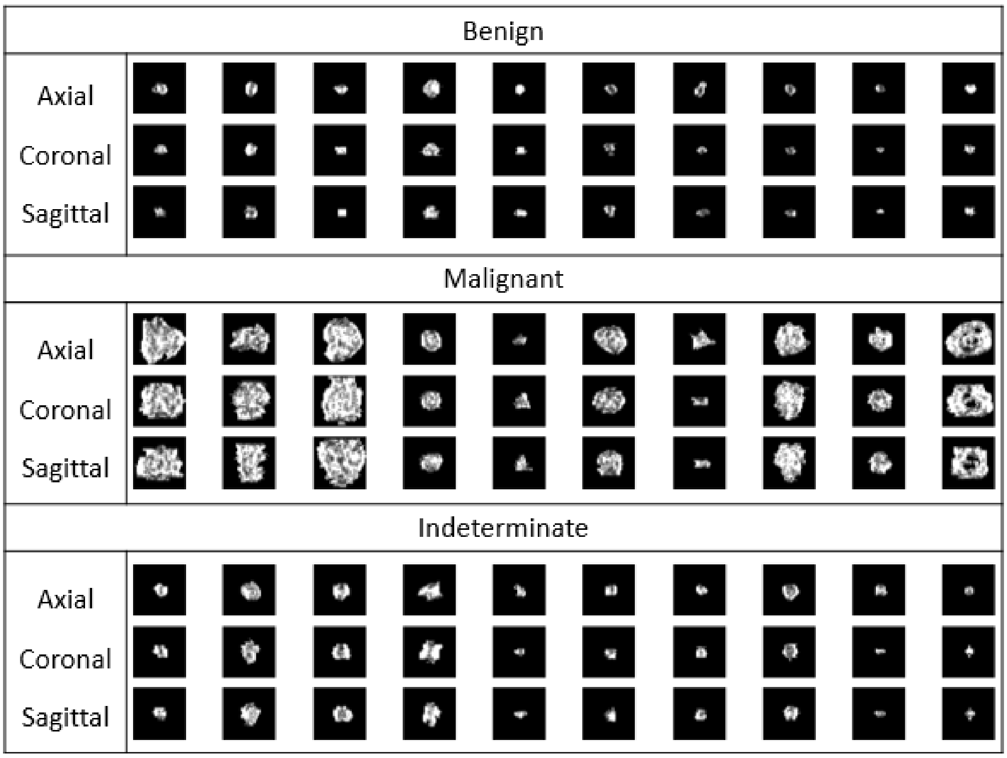


Fig. 4-10 GLN features of Fig. 4-7.

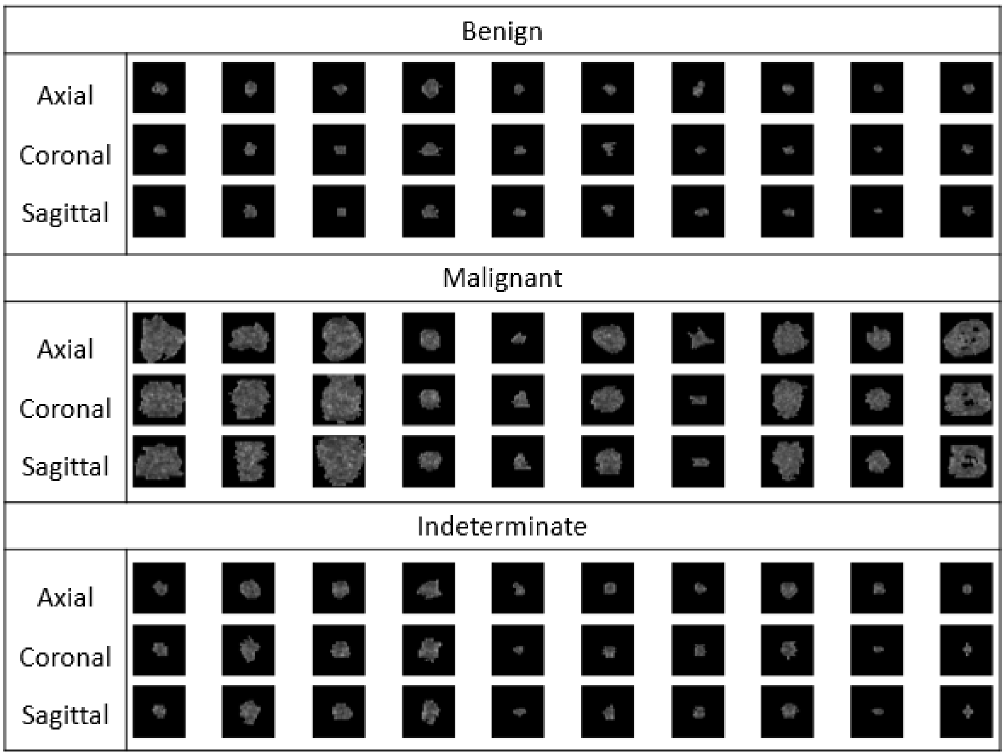


Fig. 4-11 RLN features of Fig. 4-7.

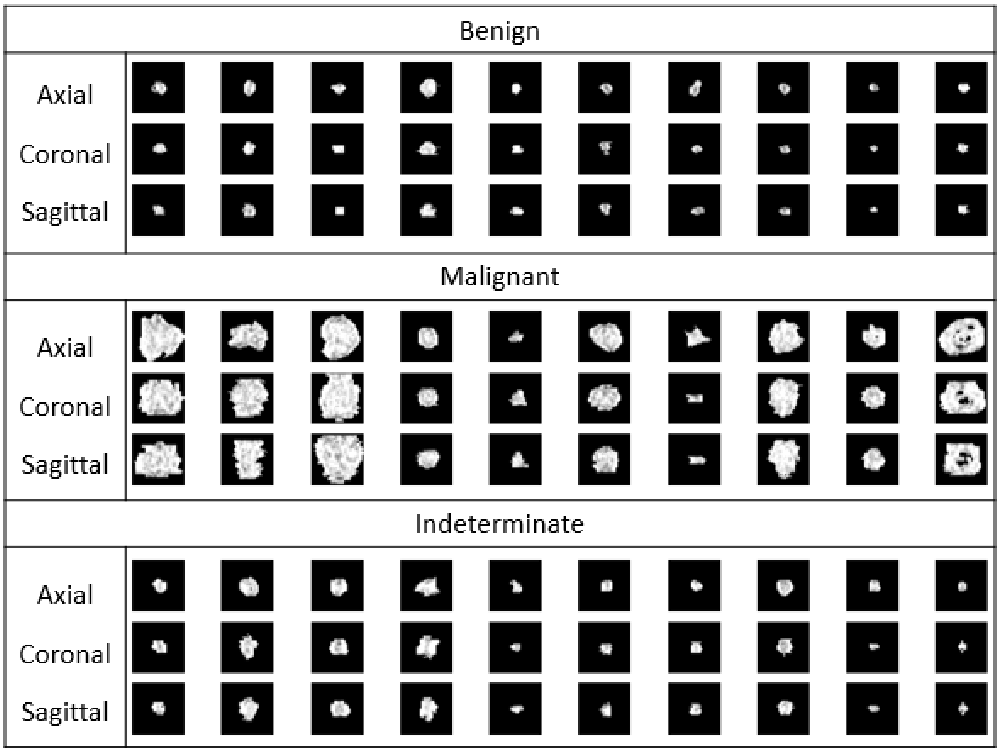


Fig. 4-12 RP features of Fig. 4-7.

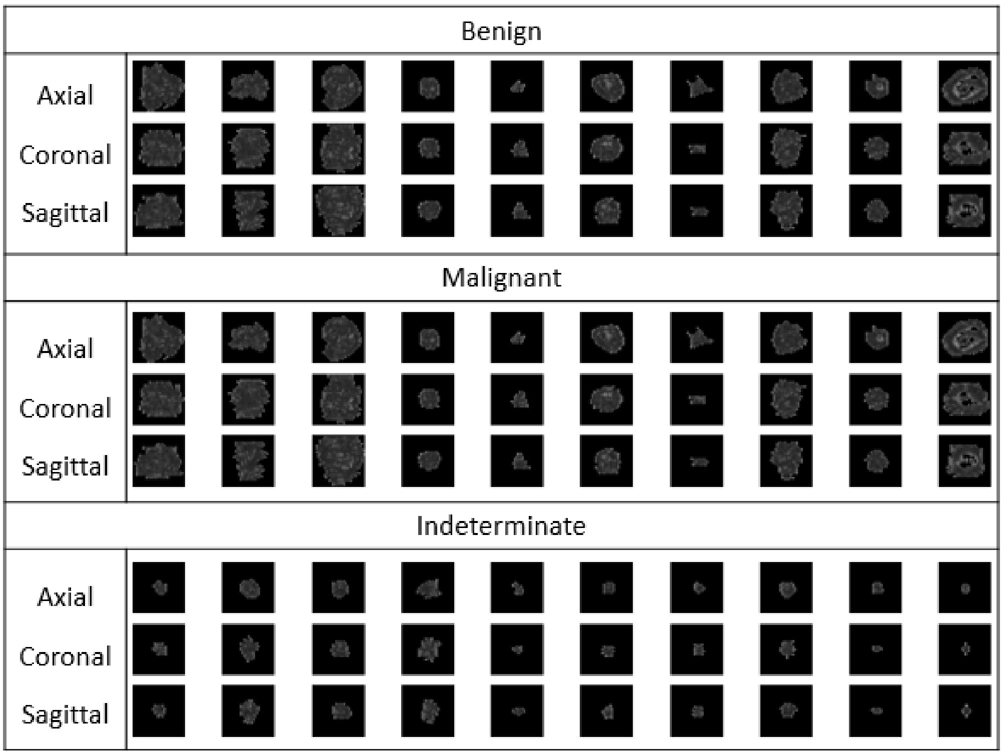


Fig. 4-13 SRE features of Fig. 4-7.

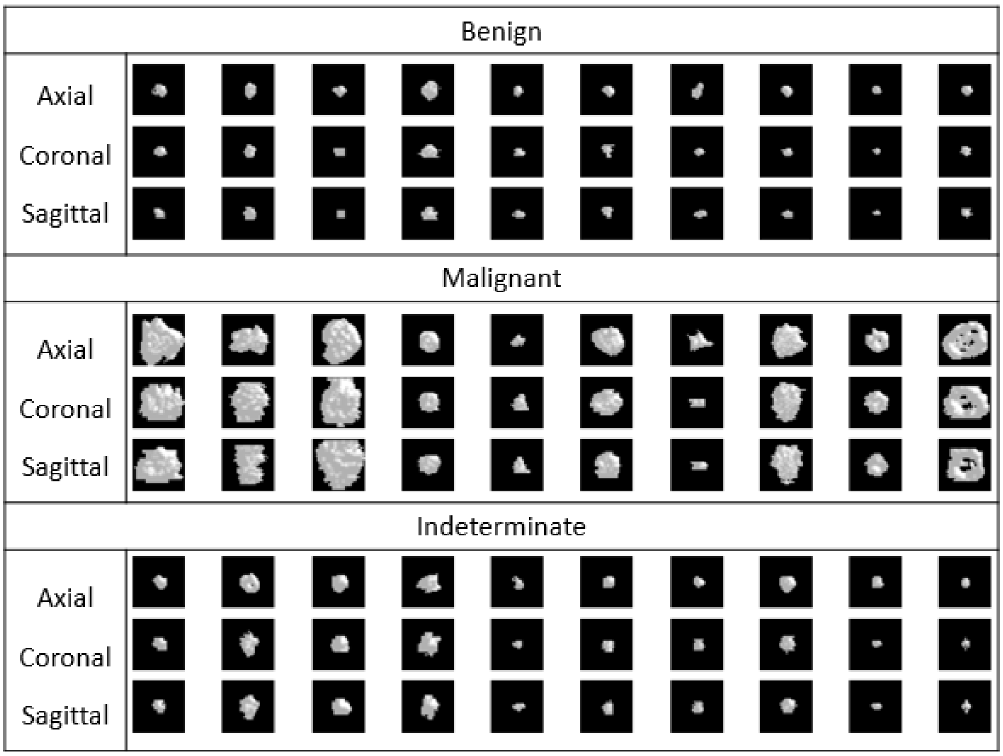


Fig. 4-14 Coarseness features of Fig. 4-7.

doi:10.6342/NTU202104149

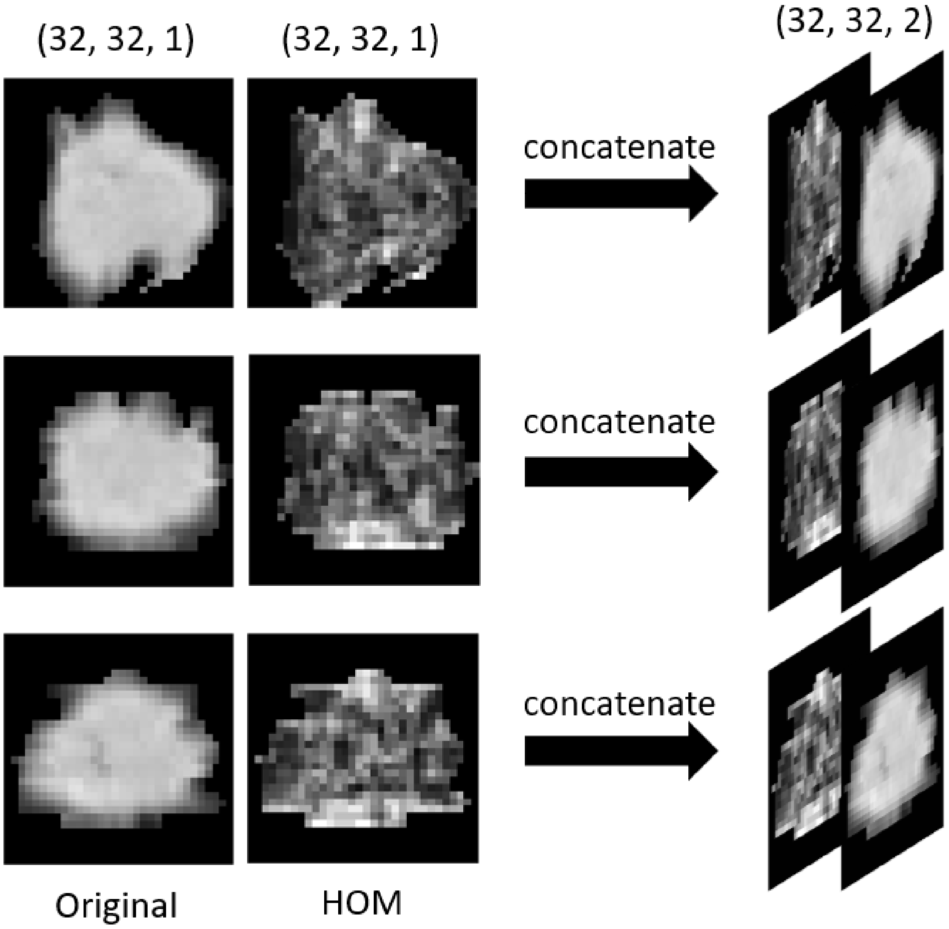


Fig. 4-15 Concatenation of the original images and texture features.

4.1.3 Data Augmentation

Deep learning, which is a data driven method, requires training with abundant data to perform well on test data. Medical images are hard to acquire and there are not enough data mostly and so is our case. After data screening and preprocessing, there are only 877 nodules in our dataset. We applied seven rotation angles (i.e., 45°, 90°, 135°, 180°, 225°, 270°, and 315°) to each nodule image, as shown in Fig. 4-16. Then we employed horizontal flip on these nodule images, as shown in Fig. 4-17, and rotate them again. Finally, the number of nodules increased by sixteen times after data augmentation.

doi:10.6342/NTU202104149

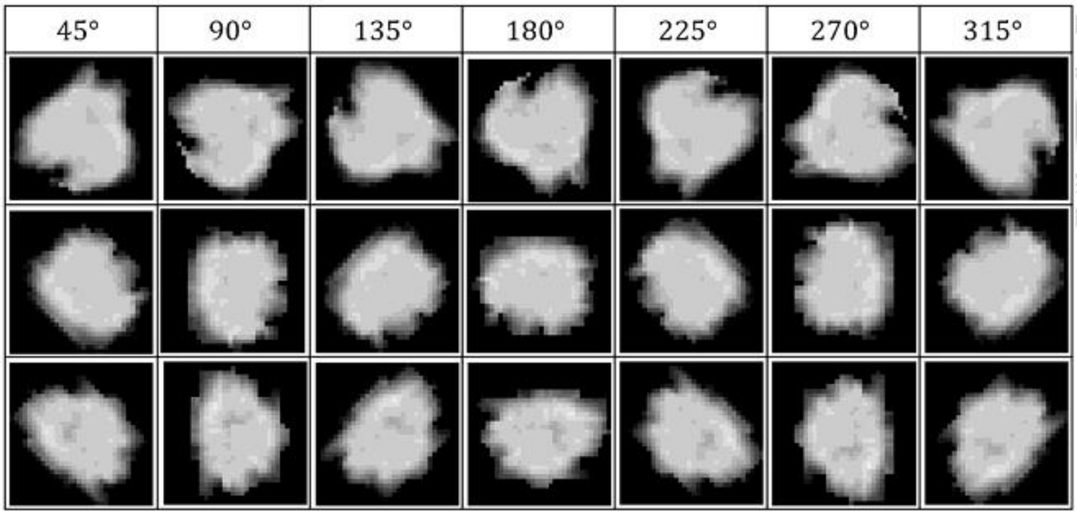


Fig. 4-16 Rotation examples.

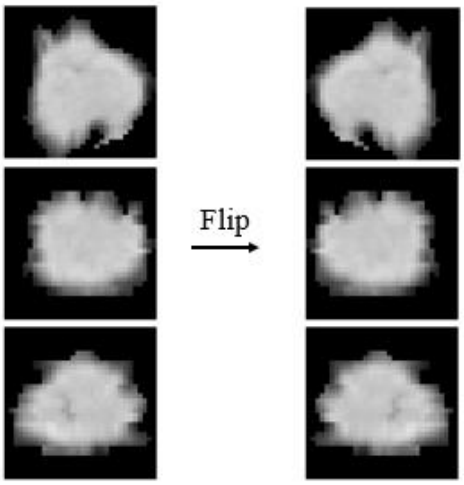


Fig. 4-17 Flip examples.

4.2 Model Architecture

4.2.1 RSK Block

Lung nodules generally have different scales and various morphologies. A larger filter is difficult to extract subtle features from small objects. For large nodules, a small filter's RFs are too small to capture the overall information of nodules. Therefore, we propose the residual selective kernel (RSK) block to extract diverse features from nodules with two different filters.

The idea of the RSK block is from the “SK convolution” in SKNet [18]. The SK convolution is used instead of the normal 3×3 convolution in a bottleneck residual block of ResNet [17], as shown in Fig. 3-10 (b). In our approach, we combined the design concepts of the residual learning and selective kernel. We employed the RSK block through three operations: Split, Fuse, and Select. In the basic architecture, the part of Split consists of two sequential convolution operations with different kernel sizes. We even added one more branch which is the identity mapping to the RSK block so that there are three branches in the beginning of the network: (1) Conv3, (2) Conv5, and (3) identity mapping. The architecture can learn the weights of all features and multiply the weights and the features at last. The RSK block is shown in Fig. 4-18. It made neurons to adaptively adjust their RF sizes based on multiple scales of input information during inference.

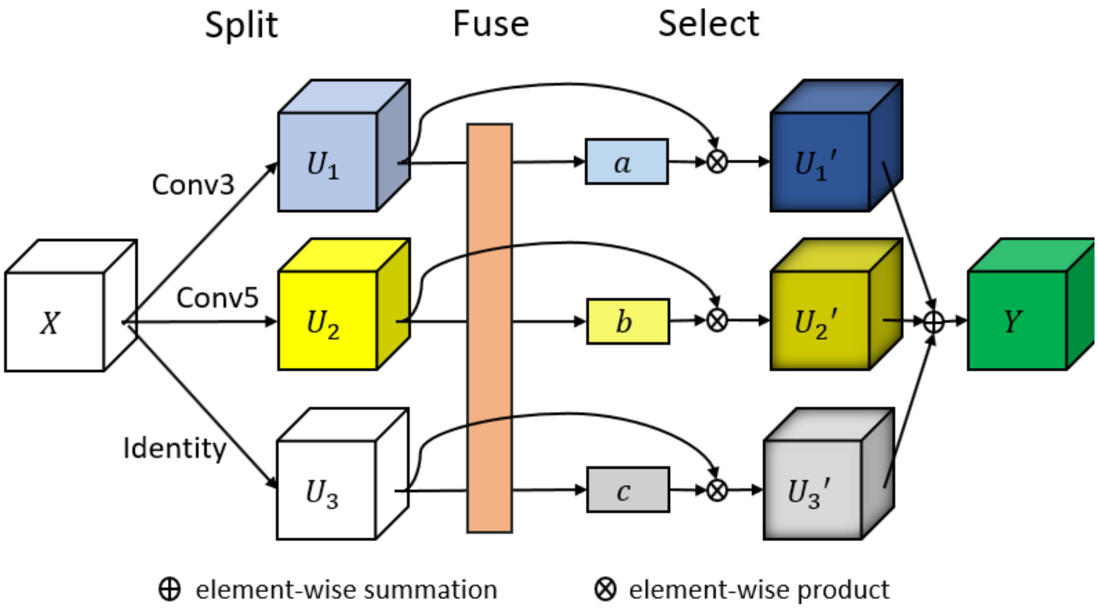


Fig. 4-18 RSK block.

The first part is Split. The input feature map is denoted as X $\in R^{H\times W\times C}$, where H, W, and C are the height, width, and number of channels of the input X. We conducted transformations consisting of two sequential convolution operations (Conv3 and Conv5) and one identity mapping shown in Fig. 4-18. This makes $X \rightarrow U_i \in R^{H'\times W'\times C'}$, where H', W', and C' are the height, width, and number of channels after the transformations and the subscript i is the number of the branches. This convolution operator is similar to the residual bottleneck block. The Conv3 process comprises 1×1 convolution, 3×3 convolution, and 1×1 convolution, as shown below:

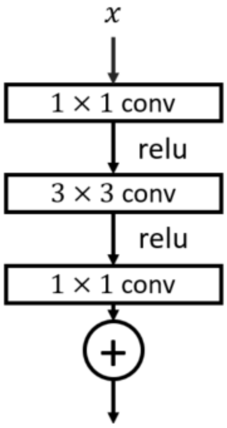


Fig. 4-19 Bottleneck.

And we built the Conv5 process by changing the kernel size of the middle convolution to 5. The channels in the first 1×1 convolutions are halved to reduce the computational complexity and then increased to the number of the output channels C' in the last 1×1 convolution. The first two convolutions are followed by the same batch normalization and ReLU activation function and the last convolution is followed by the batch normalization only. Besides, when the input channels and the output channels are different, we use the 1×1 convolution followed by the batch normalization to employ the identity mapping. Therefore, there are three branches in

doi:10.6342/NTU202104149

our design architecture.

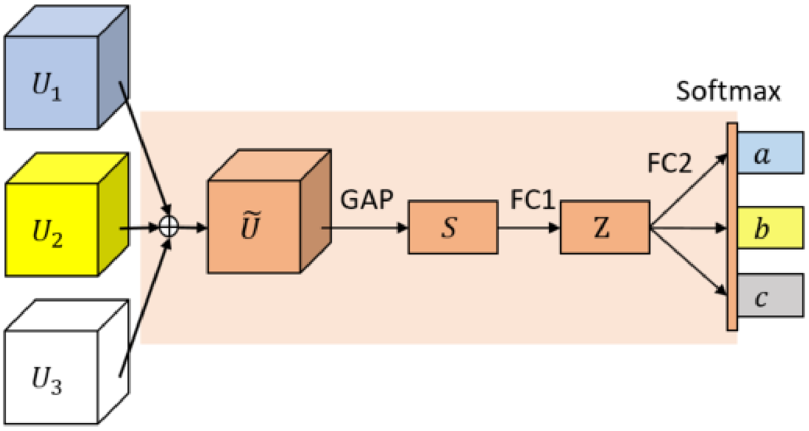


Fig. 4-20 Fuse operator

The second part is Fuse which is the orange region in Fig. 4-18 and we show the detail architecture in Fig. 4-20. Our objective is to make neurons to adaptively adjust their RF sizes by using gates to control the information flowing from the three branches, which contain different feature maps, into the next layer. To attain this goal, the gates need to consolidate the feature maps from all branches. First, we applied the ReLU activation function to the feature maps which are from the previous step, and then fused them via the element-wise summation:

$$\widetilde{U}(h,w,c) = \sum_{i=0}^{N} U_i(h,w,c) \qquad \text{Eq. 4-3}$$

where N is the number of branches. Then we used global average pooling (GAP) to obtain the global information and flattened the feature maps to generate the channel-wise statistics:

doi:10.6342/NTU202104149

$$S(c) = GAP\left(\widetilde{U}(h,w,c)\right) = \frac{1}{H' \times W'} \sum_{h=1}^{H'} \sum_{w=1}^{W'} \widetilde{U}(h,w,c) \qquad \text{Eq. 4-4}$$

where $S \in R^{C'}$ is the output of GAP. Afterwards, we employed two simple FC layers to get compact features using

$$Z = \mathrm{FC}(S) \qquad \text{Eq. 4-5}$$

where $\mathrm{Z} \in R^{d\times 1}$ is the output of the first FC layer and d denotes the adjustable number of channels in the first FC layer. To determine d, we used a reduction ratio r to control its value:

$$d = C'/r \qquad \text{Eq. 4-6}$$

where C' is the number of output channels. This makes our model have better precision and adaptive selections. Then, the second FC layer divides Z into three branches.

The last part is Select. This step uses soft attention across channels to adaptively select different scale feature maps. After the second FC layer, we applied the softmax function to the channel-wise digits to get the soft attention weights a, b, and c corresponding to the three feature maps, which are from the Split step:

$$\mathrm{a} = \frac{e^{AZ}}{e^{AZ} + e^{BZ} + e^{CZ}} \qquad \text{Eq. 4-7}$$

doi:10.6342/NTU202104149

$$b = \frac{e^{BZ}}{e^{AZ} + e^{BZ} + e^{CZ}} \qquad \text{Eq. 4-8}$$

$$c = \frac{e^{CZ}}{e^{AZ} + e^{BZ} + e^{CZ}} \qquad \text{Eq. 4-9}$$

where $A, B, C \in R^{C' \times d}$ are the outputs from the second FC layer and $a, b, c \in R^{C' \times 1}$ are the soft attention weights for the feature maps U. To do element-wise product, we reshaped the dimension of the soft attention weights to $a, b, c \in R^{H' \times W' \times C'}$ which is the same as the feature maps U. The final feature map Y is obtained via the element-wise summation of the soft attention weights multiplying the feature maps from each branch using

$$Y = U_1' + U_2' + U_3' = a \otimes U_1 + b \otimes U_2 + c \otimes U_3 \qquad \text{Eq. 4-10}$$

where $Y \in R^{W' \times H' \times C'}$ is the output of the RSK block.

4.2.2 MRSKNet

Since the input size is small, we build a smaller model architecture which has better efficiency for our task than classic ResNet or SKNet [17, 18]. The classifier recognizes different objects based on both semantic and detailed features [35]. A small filter is able to acquire more detail information from the inputs but lack of semantic features. A large filter is on the contrary. Our proposed model used the SK technique which extracts multi-scale features to solve this problem. It can even adjust different RF sizes adaptively.

We called our model "MRSKNet" because we combined the concept of the multiview CNN, residual learning, and SK. In terms of the model architecture, we referred to the design of [14, 18, 37]. In our design, we thought that anatomical planes were for the multiple views of 3D images in medical image analysis. Using the multiview inputs not only preserved the spatial information of initial 3D images but also decreased the dimension of inputs to accelerate training. Due to the three different anatomical planes, Fig. 4-21 shows our proposed model with three branches before entering the concatenation layer.

In each branch, the main architecture contains four RSK blocks. The 2D dropout layers are after the second and the fourth RSK blocks to avoid overfitting. The reason we used the 2D dropout layer is that adjacent pixels within feature maps are strongly correlated, so the 2D dropout layer, which randomly zeroes out the entire channels would aid to promote independence between the feature maps. Then we used GMP to flatten these features and concatenate each feature before entering the FC layer. Eventually, the sigmoid function is used to predict the likelihood of the malignancy p. Table 4-1 presents the best settings of MRSKNet in our experiment with original images and Table 4-2 presents the best settings of MRSKNet in our experiment with texture features, where the output column shows the size of the output feature map in the layer, M denotes the number of branches, channels and stride are the hyperparameters of convolution in RSK block, and the hyperparameter r denotes the reduction ratio to decrease the number of channels of the first FC layer in RSK blocks. The difference of model architecture between the two experiments is the dropout rate.

doi:10.6342/NTU202104149

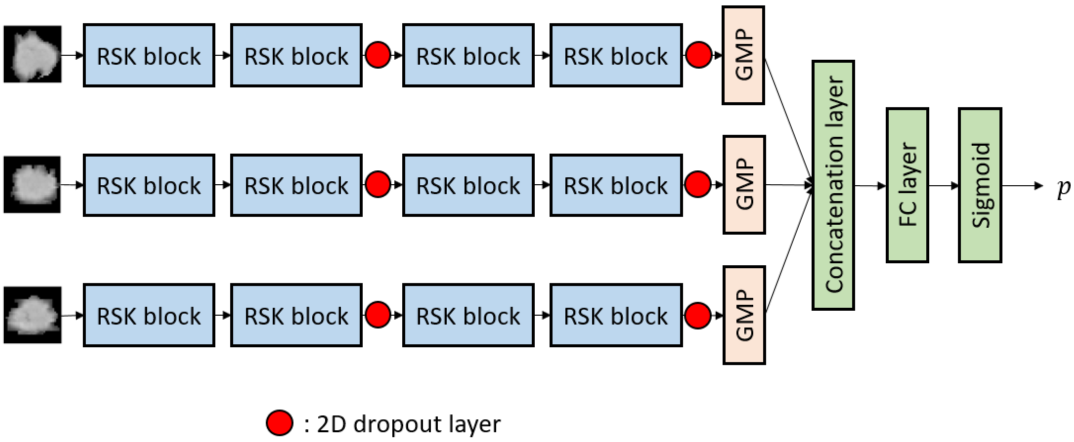


Fig. 4-21 MRSKNet diagram.

Table 4-1 The architecture of MRSKNet in the experiment with original images.

Layer	output	MRSKNet
RSK block	16 × 16	M=2, channels=32, stride=2, r=8
RSK block	16 × 16	M=2, channels=32, stride=1, r=8
Dropout		dropout rate =0.1
RSK block	8 × 8	M=2, channels=64, stride=2, r=8
RSK block	8 × 8	M=2, channels=64, stride=1, r=8
Dropout		dropout rate =0.1
Output layer	1 × 1	1 × 1 GMP, 1D FC layer, sigmoid
Parameter number		114K

doi:10.6342/NTU202104149

Table 4-2 The architecture of MRSKNet in the experiment with texture features.

Layer	output	MRSKNet
RSK block	16×16	M=2, channels=32, stride=2, r=8
RSK block	16×16	M=2, channels=32, stride=1, r=8
Dropout		dropout rate =0.2
RSK block	8×8	M=2, channels=64, stride=2, r=8
RSK block	8×8	M=2, channels=64, stride=1, r=8
Dropout		dropout rate =0.2
Output layer	1×1	1×1 GMP, 1D FC layer, sigmoid
Parameter number		114K

doi:10.6342/NTU202104149

Chapter 5 Experimental Results

5.1 Implementation detail

Our model was implemented in Pytorch [54], which is one of the most popular python frameworks to employ deep learning. The machine we used contains 128 GB memory, sixteen Intel Core i7-10700F CPUs, and Ubuntu 18.04.5 LTS operation system. To speed up the training of the neural network, the machine is also equipped with one Nvidia GeForce RTX 3090 GPU, whose specification is shown in Table 5-1.

The MRSKNet was trained with a batch size of 256 for 50 epochs. The initial learning rate was set to 1×10^{-4} and declined to 1×10^{-5} after half of the epochs. We set a higher initial learning rate so that the gradient descent was faster in the early stages of training. Then, we decreased the learning rate to avoid overshooting and to find the global minimum easier. We used Adam as our optimizer with the default hyperparameters β_1 and β_2 which were 0.9 and 0.999, respectively. The loss function is BCE for this binary classification task. Then, we initialized all the weights with the Kaiming uniform distribution $\mu\left(-\sqrt{k}, \sqrt{k}\right)$ [55], which is more suitable for the model containing the ReLU activation function to avoid the problem of gradient vanishing and exploding, where n is the number of input channels multiplying kernel size in a convolution layer or just the number of input channels in a FC layer.

$$k = \frac{1}{n} \qquad \text{Eq. 5-1}$$

doi:10.6342/NTU202104149

Table 5-1 Specification of GeForce RTX 3090.

Specification	GeForce RTX 3090
CUDA Cores	10496
Tensor Cores	328
ROPs	112
Base Core Clock	1.4 GHz
Boost Core Clock	1.7 GHz
Memory Clock	9750 MHz
Memory	24 GB
Memory Bandwidth	936 GBps
Memory Bus Width	384 bit
Single Precision	29284 GFLOPS
Double Precision	458 GFLOPS
Transistors	28.3 B
Architecture	GA102

5.2 Evaluation method

We evaluated our model by the ten-fold cross validation method. In our dataset, there were 447 benign nodules and 430 malignant nodules. The number of nodules was 877. The training data and validation data were split into 9:1 randomly. In order balance the data distribution, we kept that the ratio of benign nodules roughly 51% and the other

doi:10.6342/NTU202104149

was malignant nodules in each fold. After the data augmentation process, the number of nodules increased by sixteen times. Then, we chose the augmentation data for training and non-augmentation data for validation. The data distribution for ten-fold cross validation is shown in Table 5-2. Finally, we chose the best result in each fold and average them to be our final performance.

Table 5-2 The data distribution in each fold.

Fold 1~7	Training	Benign: 6432 Malignant: 6192 Total: 12624
	Validation	Benign: 45 Malignant: 43 Total: 88
Fold 8~10	Training	Benign: 6448 Malignant: 6192 Total: 1264
	Validation	Benign: 44 Malignant: 43 Total: 87

Although we distributed the benign and malignant nodule images equally in each fold, the numbers of benign and malignant nodules were still different. In the case of imbalanced data, we could not just evaluate the performance of the model by using accuracy, which is the most common metric in the classification task. To evaluate the performance of our proposed model which was the average performance of the ten validation folds, we also used the area under the receiver operating characteristic (ROC) curve (AUC), sensitivity, specificity, and precision metrics. The ROC curve shows the relation between true positive rate (TPR) which is sensitivity and false positive rate (FPR) which is $1 -$ specificity. These evaluation metrics have been used widely in binary classification from different aspects. These metrics are defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{Eq. 5-2}$$

$$\text{AUC} = \frac{\sum_{Sample_i \in positive} rank_{sample_i} - \frac{M(M+1)}{2}}{M \times N} \tag{Eq. 5-3}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{Eq. 5-4}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{Eq. 5-5}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{Eq. 5-6}$$

where TP represents true positives, TN represents true negatives, FP represents false positives, and FN represents false negatives. In the formula of AUC metric, M and N are the number of positive and negative samples, respectively and the rank is determined by the probability of sample prediction in an ascending order that means the rank of the sample with minimum probability is 1 and the rank of sample with maximum probability is M. We thought that diagnosing a malignant nodule as a benign one is a serious problem, which can cause treatment delay. Hence, sensitivity which is also called recall is a considerably important metric. The AUC metric is properly used to evaluate the performance of the classifier model with balanced and even imbalanced data. When the ROC curve is closer to the top-left side, the AUC is higher and it indicates that the model has higher TPR and lower FPR. Therefore, the main metrics which we used to evaluate models were AUC, accuracy, and sensitivity.

5.3 Results

We not only proposed a new model architecture but also extracted seven kinds of texture features from the original images to improve the performance. There were totally eight subsets which were the original images (only contained the anatomical planes) along with other seven texture features (ENT, HON, GLN, RLN, RP, SRE, and Coarseness), which were combined with the original images. We tried several configurations in the experiment with the subset of the original images and applied the best architecture to other seven subsets. Then, we also finetuned some hyperparameters to make the performance better in the experiment with the texture features. The final configurations were shown in Table 4-1 and Table 4-2.

5.3.1 Performance of the proposed method

According to the exploration of the model configurations which is illustrated in the ablation study section, we used the most suitable configurations for our proposed MRSKNet. To evaluate our model, the ten-fold cross validation was used in our experiments. In the experiment with the original images, the model configuration is shown in Table 4-1. The result is revealed in Table 5-3 and the ROC curve of each fold is shown in Fig. 5-1.

Table 5-3 The performance of the experiment with the original images.

AUC	Accuracy	Sensitivity	Specificity	Precision
0.9696	0.9349	0.9346	0.9356	0.9395

doi:10.6342/NTU202104149

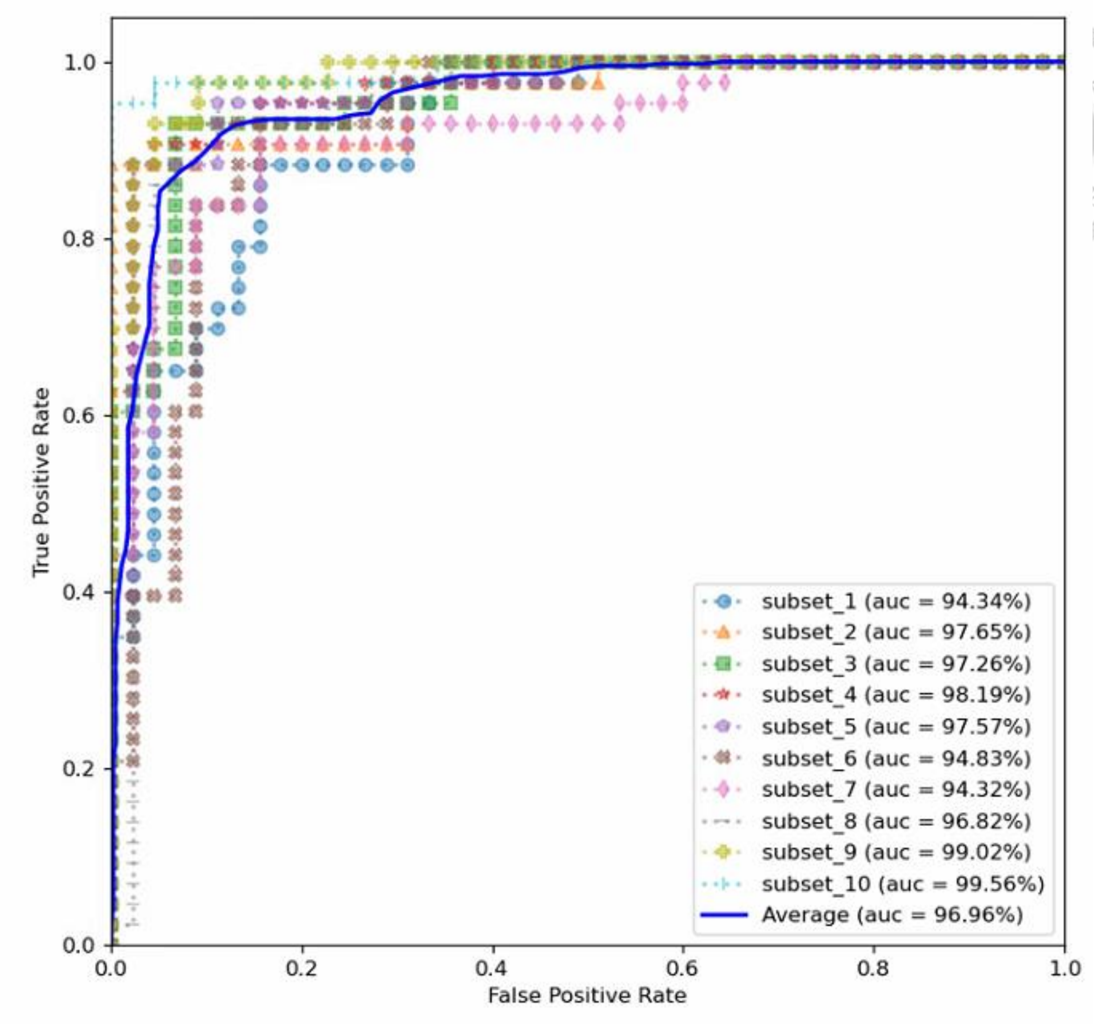


Fig. 5-1 ROC curves of ten-fold cross validation in the experiment with original images.

In addition to the original images as the input, we extracted the texture features from the original images and combined them as the input images (refer to Fig. 4-15). There were 7 kinds of texture features used in our experiments: ENT, HOM, GLN, RLN, RP, SRE, and Coarseness. We also used the same model architecture but finetuned the model hyperparameters. Finally, only the dropout rate was changed to 0.2. Then we got the new configuration which is shown in Table 4-2. From the results in Table 5-4, we noticed that the experiment with HOM improved the performance significantly. The AUC, accuracy, and sensitivity achieved 0.9711, 0.9366, and 0.9556, respectively.

doi:10.6342/NTU202104149

Besides, the highest Specificity is the experiment with ENT and the highest precision is the experiment with the original images. However, these two metrics are less important. Fig. 5-2 shows the ROC curves of different texture feature combination experiments and Fig. 5-3 shows the ROC curves of the experiment which combines with HOM. The blue solid line is the average result of 10-fold cross validation.

Table 5-4 Comparisons of the performance between different texture feature combination experiments.

Dataset	AUC	Accuracy	Sensitivity	Specificity	Precision
Original	0.9696	0.9349	0.9346	0.9356	**0.9395**
ENT	0.9669	0.9242	0.9139	**0.9365**	0.9314
HOM	**0.9711**	**0.9366**	**0.9556**	0.9177	0.9222
GLN	0.9675	0.9159	0.9209	0.9144	0.9114
RLN	0.9691	0.9324	0.9498	0.9150	0.9200
RP	0.9661	0.9254	0.9305	0.9187	0.9228
SRE	0.9696	0.9349	0.9405	0.9246	0.9253
COA	0.9692	0.9323	0.9320	0.9333	0.9320

doi:10.6342/NTU202104149

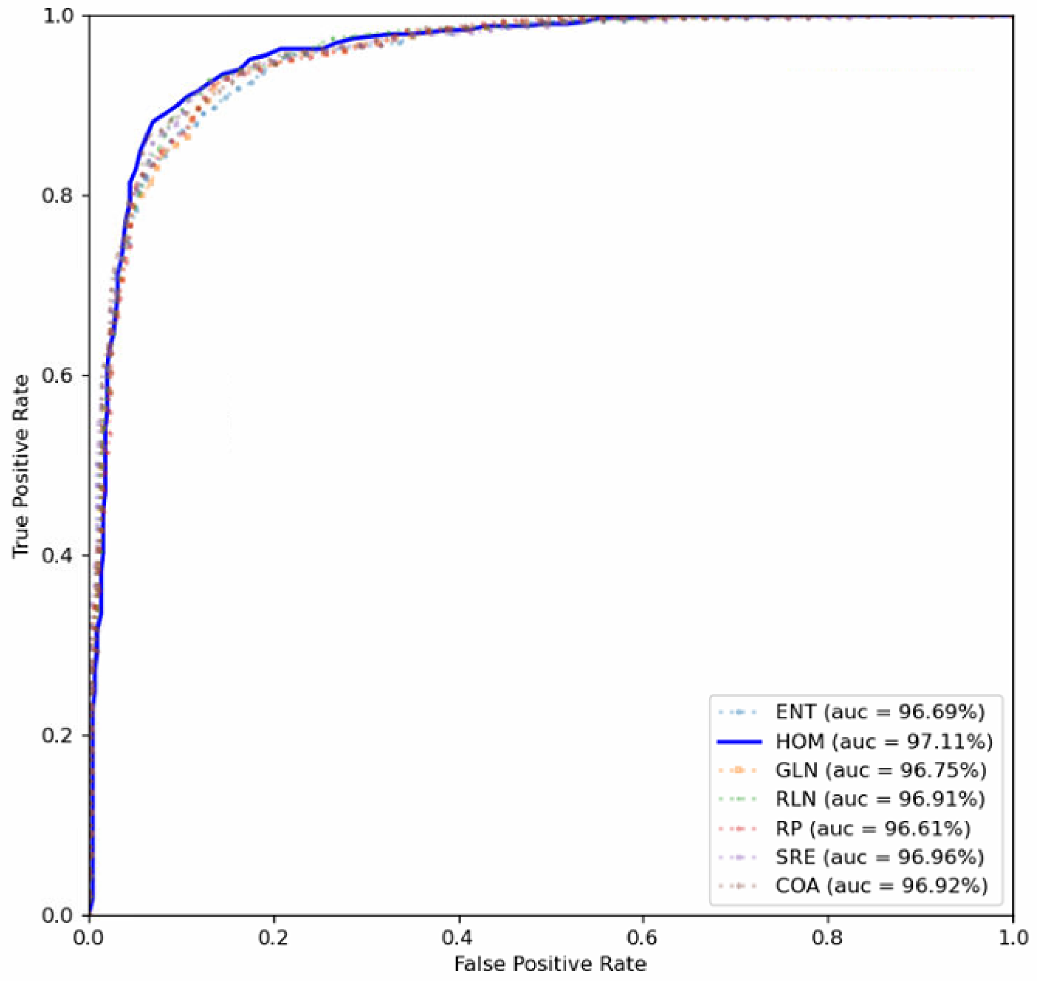


Fig. 5-2 ROC curves of different texture feature combination experiment.

doi:10.6342/NTU202104149

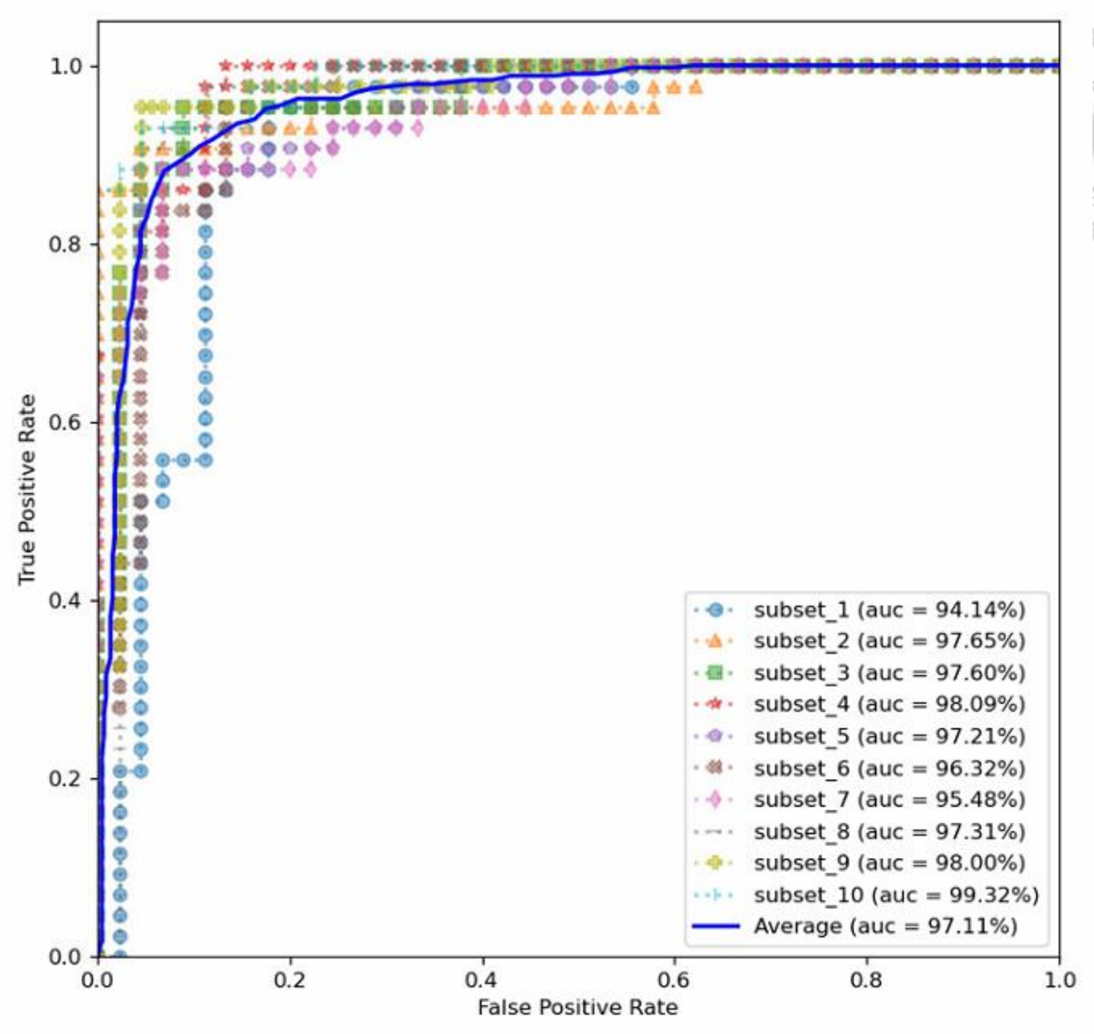


Fig. 5-3 ROC curves of the experiment which combines with HOM.

5.3.2 Ablation study

Because we used AUC, accuracy, and sensitivity to evaluate model performance, we would comprehensively select a better result. In this subsection, we only used the subset of the original images. We compared one configuration at a time and fixed other configurations the same as Table 4-1. In terms of the model architecture, we varied the numbers of RSK blocks and the number of channels of each layer. All RSK blocks have three branches. We tried 2 blocks with 32 channels, 4 blocks with 32 and 64 channels, and 6 blocks with 32, 64, and 128 channels. The number of channels were doubled

doi:10.6342/NTU202104149

every other two blocks. Table 5-5 indicated that the model with 4 RSK blocks had the best performance.

Table 5-5 Performance comparisons of different numbers of SK blocks and channels of each layer MRSKNet in the experiment with the original images.

Numbers of RSK blocks	Channels of each layer	AUC	Accuracy	Sensitivity
2	(32, 32)	0.9679	0.9269	0.9297
4	(32, 32), (64, 64)	**0.9696**	**0.9349**	0.9346
6	(32, 32), (64, 64), (128, 128)	0.9602	0.9288	**0.9356**

Since our proposed MRSKNet had the overfitting problem without dropout, we tried three kinds of the dropout rate. The comparison results are shown in Table 5-6. According to the results, using dropout layers could indeed improve the generalization of our proposed model. The dropout rate could be adjusted depending the situation. In our task, we adopted the dropout rate 0.1 and achieved the best performance.

doi:10.6342/NTU202104149

Table 5-6 Performance comparison between different dropout rates in MRSKNet in the experiment with the original images.

dropout rate	AUC	Accuracy	Sensitivity
0	0.9615	0.9327	0.9334
0.1	**0.9696**	**0.9349**	**0.9346**
0.2	0.9683	0.9293	0.9230
0.3	0.9661	0.9220	0.9122

Subsequently, we compared the hyperparameter r ,which is a reduction ratio to control the channels in the first FC layer in the RSK block. Similar to [18], the adjustable channels were set smaller than the output channels of RSK block to make the model more efficient. Finally, we acquired the best performance with $r = 8$, as shown in Table 5-7.

Table 5-7 Performance comparisons of the reduction ratio r in the experiment with the original images.

r	AUC	Accuracy	Sensitivity
2	0.9694	0.9247	0.9119
4	0.9681	0.9243	0.9210
8	**0.9696**	**0.9349**	**0.9346**
16	0.9686	0.9259	0.9266

The following is to finetune the hyperparameters in the experiment with HOM which is the best texture feature to improve the classification ability. Fist, we changed

doi:10.6342/NTU202104149

the dropout rate and the result is shown in Table 5-8. On the whole, the dropout rate = 0.2 is a better choice in the experiment with HOM.

Table 5-8 Performance comparison between different dropout rates in MRSKNet in the experiment with HOM.

dropout rate	AUC	Accuracy	Sensitivity
0	0.9657	0.9242	0.9523
0.1	0.9693	0.9350	**0.9596**
0.2	**0.9711**	0.9366	0.9556
0.3	0.9701	**0.9396**	0.9292

Finally, we adjusted the different reduction ratio r. From Table 5-9, we choose $r = 8$ which is the same as the experiment with the original in the experiment with HOM. After finetuning, we acquired the best classification ability as our final result.

Table 5-9 Performance comparisons of the reduction ratio r in the experiment with HOM.

r	AUC	Accuracy	Sensitivity
2	0.9653	0.9167	0.8963
4	0.9695	0.9299	0.9338
8	0.9711	**0.9366**	**0.9556**
16	**0.9729**	0.9290	0.9319

In this part, we designed two baseline models to compare with our proposed MRSKNet on the input data with HOM. One is the most classical model ResNet, which

doi:10.6342/NTU202104149

consists of four bottleneck residual blocks as shown in Fig. 3-10, which we call Basic ResNet. The other was SKNet, whose hyperparameters M and r are identical to the setting in our proposed MRSKNet and it also consists of four bottleneck blocks with the SK convolution which is called SK block as shown in Fig. 3-11. We call it Basic SKNet. To build these two baseline models, we only replaced the RSK block to the bottleneck residual block or the SK block and maintain the model architecture and configuration the same as Table 4-2, such as the channels in each block, the position of the dropout layer and the dropout rate. Table 5-10 illustrates that our proposed model architecture outperformed these baseline models.

Table 5-10 Performance comparisons between our proposed MRSKNet and two basic baseline models.

Model	AUC	Accuracy	Sensitivity
Basic ResNet	0.9689	0.9227	0.9276
Basic SKNet	0.9689	0.9317	0.9374
MRSKNet	**0.9711**	**0.9366**	**0.9556**

5.3.3 Performance comparison

In this section, we chose some previous works which used the same database to compare with our proposed method which is the experiment with HOM. Table 5-11 reports the comparison between our proposed method and the state-of-the-art methods, which were based on handcrafted features and traditional machine learning classifiers. Among these works, Sasidhar et al. [16] and Li et al. [15] also used the GLCM as our proposed method, but they chose the SVM and RF classifiers to classify the nodules.

doi:10.6342/NTU202104149

Though handcrafted features could represent the structure of lung nodules and useful information, but there still existed some limitations. For example, these works only used one to three kinds of handcrafted features that were not diverse enough so that it might have worse generalization ability. According to Table 5-11, our proposed method had the best AUC, accuracy, sensitivity, and precision.

Table 5-12 summarizes the comparison between our proposed method and the state-of-the-art methods based on deep learning. These model architectures were generally based on classical CNN models. In deep learning methods, the input data were divided into 2D and 3D types. Some works even used 2D multiview images instead of 3D volume images such as Nibali et al. [14], Al-Shabi et al. [10, 12], and our proposed method. The main advantage of deep learning was the ability to learn high discriminative features from raw data directly and automatically. Therefore, there are more and more researchers using deep learning to analyze medical images. It is more efficient, and accurate, and has better generalization ability. In our proposed method, we combined both handcrafted features and deep learning features. The performance of our proposed method outperformed most deep learning methods, except specificity. Our accuracy is the secondary, but our AUC is better than the work which achieved the highest accuracy. Moreover, the sensitivity of our proposed method is even three percent higher than other methods. Overall, our proposed method has the best AUC, sensitivity, precision, and the secondary accuracy.

doi:10.6342/NTU202104149

Table 5-11 Comparison between our proposed method and state-of-the-art methods based on handcrafted features and traditional machine learning classifiers.

Author	Year	Nodules (benign, malignant)	AUC	Accuracy	Sensitivity	Specificity	Precision
Netto et al. [26]	2012	198 (99, 99)	-	0.8179	0.8282	0.8077	-
*Orozco et al. [20]	2015	106 (59, 47)	0.805	-	0.9090	0.7391	0.82
*Dhara et al. [19]	2016	542 (279, 263)	0.9465	-	-	-	-
de Carvalho Filho et al. [23]	2017	1405 (1011, 394)	0.921	0.9252	0.931	0.9226	-
*Sasidhar et al. [16]	2017	-	-	0.92	-	-	-
de Sousa Costa et al. [21]	2018	1405 (1011, 394)	0.94	0.9181	0.9342	0.9121	-
*Li et al. [15]	2018	1000	0.95	0.90	0.92	0.83	-
*Ferreira et al. [29]	2018	1171 (745, 426)	0.858	0.800	0.702	0.856	-
*Wu et al. [24]	2019	614 (294, 320)	0.9702	0.9237	0.9428	0.9027	-
Proposed method	2021	877 (447, 430)	0.9711	0.9366	0.9556	0.9177	0.9222

* indicates that the work uses the GLCM feature as input data.

doi:10.6342/NTU202104149

Table 5-12 Comparison between our proposed method and state-of-the-art methods based on deep learning.

Author	Year	Nodules (benign, malignant)	AUC	Accuracy	Sensitivity	Specificity	Precision
Nibali et al. [14]	2017	831 (421, 410)	0.9459	0.8990	0.9107	0.8864	0.8935
*Zhu et al. [40]	2018	1004 (450, 554)	-	0.9044	-	-	-
*Dai et al. [32]	2018	1011	0.9690	0.9147	0.9126	0.9167	-
Xia et al. [38]	2018	1945 (1301, 644)	0.9570	0.9160	0.8652	0.9400	-
Xia et al. [39]	2019	1945 (1301, 644) + 1839 unlabeled	0.9581	0.9253	0.8494	0.9628	-
*Fu et al. [33]	2019	1186 (650, 536)	-	0.8993	0.8334	0.9105	-
Al-Shabi et al. [10]	2019	848 (442, 406)	0.9514	0.9257	0.9221	-	0.9185
Al-Shabi et al. [12]	2019	848 (442, 406)	0.9562	0.8864	0.8866	-	0.8738
Zhang et al. [41]	2019	1004 (450, 554)	0.9687	0.9378	-	-	-
*Zhang et al. [9]	2020	1004 (450, 554)	0.9563	0.9167	-	-	-
*Ren et al. [13]	2020	1226 (795, 431)	-	0.90	0.81	0.95	-
*Liu et al. [34]	2020	1268 (863, 405)	0.939	0.906	0.837	0.939	-
Lyu et al. [35]	2020	#	0.9705	0.9219	0.9210	0.9150	-
Proposed method	2021	877 (447, 430)	0.9711	0.9366	0.9556	0.9177	0.9222

* indicates that the work uses 3D data to be the input for model.
# indicates that this paper used 22001 nodules (10752 benign nodules and 11249 malignant nodules), but we thought it might be the number of nodule images.

doi:10.6342/NTU202104149

Chapter 6 Conclusions

In this thesis, we developed a new CAD method for lung nodule classification to distinguish malignant nodules from benign ones in CT images. To recognize the lung nodules with various sizes and shapes, we combined the advantages of the residual learning and selective kernel technique to develop a brand-new model architecture MRSKNet. The RSK block improve the efficiency and effectiveness of object recognition by adaptively selecting different kernels and even the identity mapping with a soft attention mechanism. It also can analyze the shape and size of the nodules by the global features and the density and structure of the nodules by the local features. In addition, we used the idea of multiview and adopted anatomical planes to be the input data. In our experiments, we used the LIDC-IDRI database to train our MRSKNet and evaluate our proposed method. The experimental results with the original images achieved AUC of 0.9696, accuracy of 0.9349, sensitivity of 0.9346, specificity of 9356, and precision of 0.9395, which already outperformed some of the state-of-the-art methods. Then, we employed GLCM, GLRLM and Tamura texture features associated with the original images to improve classification ability. Among these texture features, the experiment with HOM feature outperformed other features and achieved better performance. Although the specificity and precision were declined a little, the AUC, accuracy, and sensitivity which were more considerable were increased to 0.9711, 0.9366, and 0.9556, respectively.

In the future, developing an unsupervised method for this task can be considered because annotating the labels for every image is time-consuming. Hence, with unlabeled data, more and more data can be used to improve the generalization ability and performance.

doi:10.6342/NTU202104149

REFERENCE

1. Ferlay, J., et al., *Global Cancer Observatory: Cancer Today. Lyon, France: International Agency for Research on Cancer.* 2020.
2. Siegel, R.L., et al., *Cancer Statistics, 2021.* CA Cancer J Clin, 2021. **71**(1): p. 7-33.
3. Medicine, N.L.S.T.R.T.J.N.E.J.o., *Reduced lung-cancer mortality with low-dose computed tomographic screening.* 2011. **365**(5): p. 395-409.
4. Marianna Sockrider, M.J.A.J.o.R. and C.C. Medicine, *What is a Lung Nodule?* 2016. **193**(7): p. I.
5. Hansell, D.M., et al., *Fleischner Society: glossary of terms for thoracic imaging.* 2008. **246**(3): p. 697-722.
6. Awai, K., et al., *Pulmonary nodules: estimation of malignancy at thin-section helical CT—effect of computer-aided diagnosis on performance of radiologists.* 2006. **239**(1): p. 276-284.
7. Awai, K., et al., *Pulmonary nodules at chest CT: effect of computer-aided diagnosis on radiologists' detection performance.* 2004. **230**(2): p. 347-352.
8. Armato III, S.G., et al., *The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans.* 2011. **38**(2): p. 915-931.
9. Zhang, G., et al., *Classification of lung nodules based on CT images using squeeze-and-excitation network and aggregated residual transformations.* 2020: p. 1-10.
10. Al-Shabi, M., H.K. Lee, and M.J.I.A. Tan, *Gated-dilated networks for lung nodule classification in CT scans.* 2019. **7**: p. 178827-178838.
11. Sahu, P., et al., *A lightweight multi-section CNN for lung nodule classification and malignancy estimation.* 2018. **23**(3): p. 960-968.
12. Al-Shabi, M., et al., *Lung nodule classification using deep local–global networks.* 2019. **14**(10): p. 1815-1819.
13. Ren, Y., et al., *A manifold learning regularization approach to enhance 3D CT image-based lung nodule classification.* 2020. **15**(2): p. 287-295.
14. Nibali, A., et al., *Pulmonary nodule classification with deep residual networks.* 2017. **12**(10): p. 1799-1808.
15. Li, X.-X., et al., *Automatic benign and malignant classification of pulmonary nodules in thoracic computed tomography based on RF algorithm.* 2018. **12**(7): p. 1253-1264.
16. Sasidhar, B., et al. *Automatic classification of lung nodules into benign or malignant using SVM classifier*. in *Proceedings of the 5th International Conference on Frontiers in Intelligent Computing: Theory and Applications*. 2017. Springer.
17. He, K., et al. *Deep residual learning for image recognition*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
18. Li, X., et al. *Selective kernel networks*. in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019.
19. Dhara, A.K., et al. *Classification of pulmonary nodules in lung CT images using shape and texture features*. in *Medical Imaging 2016: Computer-Aided Diagnosis*. 2016. International Society for Optics and Photonics.

doi:10.6342/NTU202104149

20. Orozco, H.M., et al., *Automated system for lung nodules classification based on wavelet feature descriptor and support vector machine.* 2015. **14**(1): p. 1-20.

21. de Sousa Costa, R.W., et al., *Classification of malignant and benign lung nodules using taxonomic diversity index and phylogenetic distance.* 2018. **56**(11): p. 2125-2136.

22. Firmino, M., et al., *Computer-aided detection (CADe) and diagnosis (CADx) system for lung cancer with likelihood of malignancy.* 2016. **15**(1): p. 1-17.

23. de Carvalho Filho, A.O., et al., *Computer-aided diagnosis of lung nodules in computed tomography by using phylogenetic diversity, genetic algorithm, and SVM.* 2017. **30**(6): p. 812-822.

24. Wu, W., et al., *Malignant-benign classification of pulmonary nodules based on random forest aided by clustering analysis.* 2019. **64**(3): p. 035017.

25. Rodrigues, M.B., et al., *Health of things algorithms for malignancy level classification of lung nodules.* 2018. **6**: p. 18592-18601.

26. Netto, S.M.B., et al., *Analysis of directional patterns of lung nodules in computerized tomography using Getis statistics and their accumulated forms as malignancy and benignity indicators.* 2012. **33**(13): p. 1734-1740.

27. Zinovev, D., et al., *Predicting panel ratings for semantic characteristics of lung nodules.* 2010.

28. Lee, M.C., et al., *Computer-aided diagnosis of pulmonary nodules using a two-step approach for feature selection and classifier ensemble construction.* 2010. **50**(1): p. 43-53.

29. Ferreira, J.R., M.C. Oliveira, and P.M.J.J.o.d.i. de Azevedo-Marques, *Characterization of pulmonary nodules based on features of margin sharpness and texture.* 2018. **31**(4): p. 451-463.

30. Farahani, F.V., et al., *Hybrid intelligent approach for diagnosis of the lung nodule from CT images using spatial kernelized fuzzy c-means and ensemble learning.* 2018. **149**: p. 48-68.

31. Simonyan, K. and A.J.a.p.a. Zisserman, *Very deep convolutional networks for large-scale image recognition.* 2014.

32. Dai, Y., et al., *Incorporating automatically learned pulmonary nodule attributes into a convolutional neural network to improve accuracy of benign-malignant nodule classification.* 2018. **63**(24): p. 245004.

33. Fu, J. *Application of modified inception-resnet and condensenet in lung nodule classification*. in *3rd International Conference on Computer Engineering, Information Science & Application Technology (ICCIA 2019)*. 2019. Atlantis Press.

34. Liu, H., et al., *Multi-model Ensemble Learning Architecture Based on 3D CNN for Lung Nodule Malignancy Suspiciousness Classification.* 2020. **33**(5): p. 1242-1256.

35. Lyu, J., X. Bi, and S.H.J.S. Ling, *Multi-level cross residual network for lung nodule classification.* 2020. **20**(10): p. 2837.

36. An, Y., et al. *Lung Nodule Classification using A Novel Two-stage Convolutional Neural Networks Structure'*. in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2019. IEEE.

37. Su, H., et al. *Multi-view convolutional neural networks for 3d shape recognition*. in *Proceedings of the IEEE international conference on computer vision*. 2015.

38. Xie, Y., et al., *Knowledge-based collaborative deep learning for benign-malignant lung nodule classification on chest CT.* 2018. **38**(4): p. 991-1004.

39. Xie, Y., J. Zhang, and Y.J.M.i.a. Xia, *Semi-supervised adversarial model for benign–malignant lung nodule classification on chest CT.* 2019. **57**: p. 237-248.
40. Zhu, W., et al. *Deeplung: Deep 3d dual path nets for automated pulmonary nodule detection and classification*. in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. 2018. IEEE.
41. Zhang, G., et al., *Classification of benign and malignant lung nodules from CT images based on hybrid features.* 2019. **64**(12): p. 125011.
42. Wikipedia, c., *CT scan*, in *Wikipedia, The Free Encyclopedia.*
43. Wikipedia, c., *Hounsfield scale*, in *Wikipedia, The Free Encyclopedia.*
44. Wikipedia contributors, *Anatomical plane*, in *Wikipedia, The Free Encyclopedia.*
45. Mayo, J.R.J.J.o.t.i., *CT evaluation of diffuse infiltrative lung disease: dose considerations and optimal technique.* 2009. **24**(4): p. 252-259.
46. Haralick, R.M., et al., *Textural features for image classification.* 1973(6): p. 610-621.
47. Mohanaiah, P., et al., *Image texture feature extraction using GLCM approach.* 2013. **3**(5): p. 1-5.
48. Wikipedia contributors, *Co-occurrence matrix*, in *Wikipedia, The Free Encyclopedia.*
49. Galloway, M.M.J.N.S.R.T.R.N., *Texture analysis using grey level run lengths.* 1974. **75**: p. 18555.
50. Tamura, H., et al., *Textural features corresponding to visual perception.* 1978. **8**(6): p. 460-473.
51. Wikipedia contributors, *Convolutional neural network*, in *Wikipedia, The Free Encyclopedia.*
52. Clark, K., et al., *The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository.* 2013. **26**(6): p. 1045-1057.
53. Reeves, A.P. and A.M. Biancardi, *The Lung Image Database Consortium (LIDC) Nodule Size Report.*
54. Paszke, A., et al., *Automatic differentiation in pytorch.* 2017.
55. He, K., et al. *Delving deep into rectifiers: Surpassing human-level performance on imagenet classification*. in *Proceedings of the IEEE international conference on computer vision*. 2015.

doi:10.6342/NTU202104149