Identification of lung nodules based on combining multi-slice CT images and clinical information

Yuto Nishitaki
Kyushu Institute of Technology, 1-1 Sensui-cho, Tobata-ku, Kitakyushu, 804-8550, Japan

Tohru Kamiya
Kyushu Institute of Technology, 1-1 Sensui-cho, Tobata-ku, Kitakyushu, 804-8550, Japan

Shoji Kido
Osaka University, 2-2 Yamadaoka, Suita, Osaka, 565-0871, Japan

Email: nishitaki.yuuto699@mail.kyutech.jp

Abstract

Although chest CT scans are an effective means of diagnosing lung cancer, there are still problems such as the heavy burden on physicians. To solve this problem, computer-aided diagnosis (CAD) systems are being introduced. Conventional CAD systems are based on a method that uses only image information. In this study, we propose a method for identifying nodular shadows that integrates a composite image created from multi-slice CT images and clinical information such as the patient's age, sex, and medical history in the medical record. The proposed method extracts features from the multi-section CT and clinical information, respectively, integrates the features, and then performs binary classification of nodules and vessels using a classifier. The proposed method achieved a very high accuracy of Accuracy=0.983, TPR=0.987, and FPR=0.018.

Keywords: Computer Aided Diagnosis, Deep Learning, Clinical Information, Multimodal, Multi-slice CT

1. Introduction

Malignant neoplasms are the leading cause of death worldwide, and lung cancer is one of the most serious of these diseases; according to a 2022 report by the International Agency for Research on Cancer (IARC), more than 2.4 million new cases of lung cancer will be reported worldwide each year, making it the leading cause of cancer death [1]. In Japan, malignant neoplasms have been the leading cause of death since 1981, and by 2022, lung cancer will be the leading cause of death in men, the second leading cause in women, and the leading cause in both men and women combined [2]. Because lung cancer progresses rapidly, and survival decreases significantly as the disease progresses, early detection and early treatment are important issues. When lung cancer is suspected, a chest X-ray is performed first, and if abnormalities are found, a chest CT scan is widely used as a more precise examination. However, the large number of CT images generated in a single examination places a heavy burden on the physician reading the images. In addition, the possibility of variations in diagnostic accuracy and missed cases due to differences in the skills and experience of the reading physicians has also been pointed out. To address these issues, computer-aided diagnosis (CAD) systems are needed to reduce the burden on the reading physician and improve diagnostic accuracy. Conventional CAD systems are based on a method that uses only image information [3], [4]. However, in actual clinical practice, diagnosis is made by considering background information such as the patient's age, gender, and medical history. For this reason, CAD systems that integrate image and clinical information have been studied in recent years. In this paper, we propose a method for identifying pulmonary nodular shadows by combining multi-slice CT images and clinical information from medical records.

2. Methodology

The overall process is shown in Figure 1. First, image features and clinical information are extracted from the CT images obtained from the examination and the clinical information of patient, respectively. Next, these features are integrated and finally classified into two classes, nodules and vessels, using a classifier.

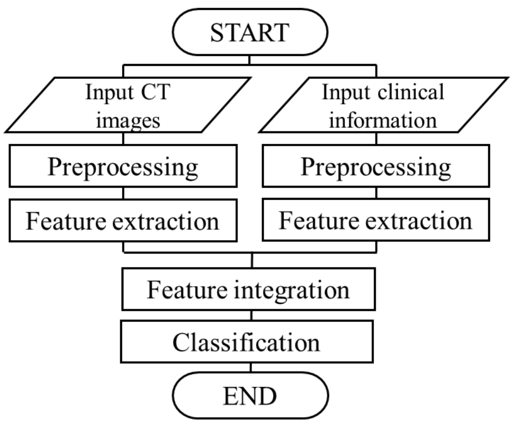


Fig.1. Flow of the proposed method.

2.1. *Image feature*

In this paper, to obtain 3D features of the target with low computational cost, a synthetic image using three images (axial, sagittal, and coronal cross-sections) is fed into an image feature extraction model. For feature extraction, CoAtNet [5], which combines CNN to capture local features and Self-Attention to capture global features, is used as the base model, and an improved CoAtNet that introduces a Convolutional Block Attention Module (CBAM) [6] is used. The CBAM was introduced to estimate attention maps along two different dimensions, channel and spatial, with the goal of highlighting important features and suppressing unnecessary information.

2.2. *Clinical information feature*

In this paper, gender, age, and medical history are used as medical record information. The six types of medical history used are colorectal cancer, stomach cancer, other cancers, tuberculosis, pneumonia, and other lung-related diseases. As a preprocessing step, gender is converted to one-hot coding with 1 for males and 0 for females. Similarly, history is converted to one-hot coding with 1 if the disease is present and 0 if it is absent. Age is standardized to reduce the effect of outliers and to equalize the scales of the explanatory variables. After these preprocessing steps, a multi-layer perceptron consisting of three layers is used to extract clinical information features. Details of the model used are shown in Table 1.

2.3. *Feature Integration*

The acquired 512-dimensional image features and 128-dimensional clinical information features are horizontally combined by concatenating layers to form a 640-dimensional feature vector. Self-Attention with Scaled Dot-Product Attention [7] is then applied to the combined features to highlight important relationships between different modalities. This feature vector is fed into a four-layer multilayer perceptron, which performs a two-class classification of nodules and blood vessels. A summary of the model used is shown in Table 2.

Table 1. Clinical information extraction model

Layer	Input	Output	Activation function
Input	8		-
Dropout.1	-	-	-
Linear.1	8	128	ReLU
Dropout.2	-	-	-
Linear.2	128	256	ReLU
Dropout.3	-	-	-
Linear.3	256	128	softmax
Output		128	-

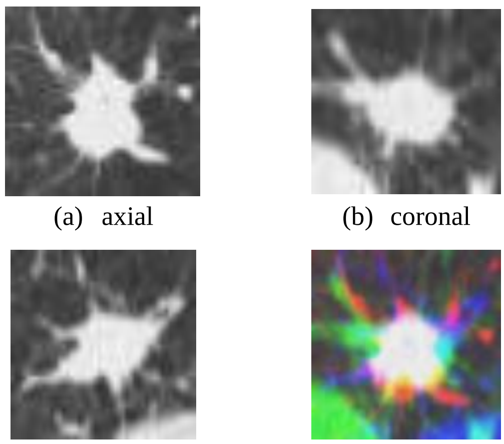
Table 2. Classification model

Layer	Input	Output	Activation function
Input	640		-
Dropout.1	-	-	-
Linear.1	640	1024	ReLU
Dropout.2	-	-	-
Linear.2	1024	512	ReLU
Dropout.3	-	-	-
Linear.3	512	256	ReLU
Dropout.4 Linear.4	256	2	softmax
Output		2	-

3. Experiment

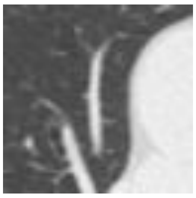
3.1. *Dataset*

In this paper, we use three images from chest CT images: axial, sagittal, and coronal cross-sections to obtain 3D features of the object while minimizing computational complexity. CT images with a nodule of 20 mm or less were used as input images. For the axial section, the slice plane with the largest nodule was selected from a set of CT image data obtained from a chest CT examination. The area of the nodule was manually segmented and modified under the guidance of a physician. In addition, because the region of interest was small, it was cropped from 512 × 512 pixels to 112 × 112 pixels and resized as needed. For the coronal and sagittal sections, the center coordinates of the region of interest were obtained from the axial section, and each image was cropped to 112 × 112 pixels based on these coordinates. The same procedure was used to generate images of blood vessels. The nodule images of 157 nodules from 73 cases were used as positive images, and the vessel images of 325 vessels from 73 cases were used as negative images. The images are shown in Figure 2 and Figure 3.

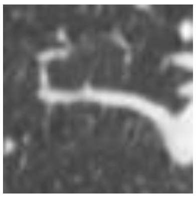


(a) axial (b) coronal (c) sagittal (d) fusion image

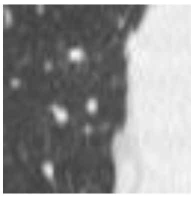
Fig.2. Dataset (lung nodule)



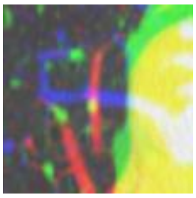
(e) axial



(f) coronal



(g) sagittal



(h) fusion image

Fig.3. Dataset (blood vessel)

Table 3. Experimental results (image only)

Model (Input image)	Accuracy	TPR	FPR
Improved ResNet [8] (Axial image)	0.938	0.917	0.052
CoAtNet (Axial image)	0.935	0.923	0.061
Proposed model (Axial image)	0.950	0.949	0.049
Proposed model (Fusion image)	0.977	0.980	0.025

Table 4. Experimental results (image + clinical information)

Method	Accuracy	TPR	FPR
Method1	0.968	0.968	0.032
Proposed method	0.983	0.987	0.018

3.2. *Data augmentation*

Deep learning is prone to overlearning when the amount of data is small. Therefore, it is necessary to improve the generalization performance by data expansion. In this paper, the training data was rotated from 15 to 345 degrees in 15-degree increments and flipped horizontally and vertically to increase the data volume 23-fold.

3.3. *Evaluation function*

The metrics used are Accuracy, which is the proportion of correctly classified images in all data, True Positive Rate (TPR), which is the proportion of images correctly identified as positive when the correct answer is positive, and False Positive Rate (FPR), which is the proportion of images incorrectly identified as positive when the correct answer is negative.

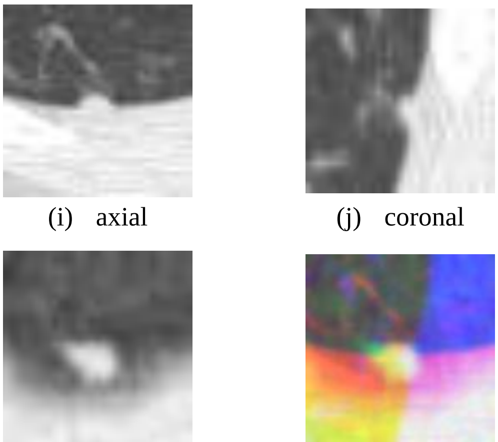
4. Results and Discussion

Table 3 shows the classification results using only images as input. The image feature extraction models used were Improved ResNet, which was proposed in a previous study [8], and CoAtNet, which was the base model before improvement, and the proposed model that introduced CBAM into CoAtNet. Improved ResNet is a ResNet34 [9] with a residual block just before, a convolutional autoencoder is added to ResNet34, and an SE block [10] is introduced within the residual block. The classification results using image and clinical information are shown in Table 4. All the proposed models were used as image feature extraction models. Method 1 is a method in which image features and clinical information are integrated and then directly input into a classification model. In the proposed method, the features transformed by Self-Attention are input into the classification model after feature integration.

Table 3 shows that the proposed method achieves higher accuracy than the other methods in the image-only experiment. This may be due to the introduction of CBAM, which calculates the importance in both the channel and spatial directions for the feature map, thereby highlighting important features. Another contributor to the improved accuracy may be the use of composite images generated from cross-sectional images in three directions, which allows accurate detection of different shape features: spherical or elliptical pulmonary nodules and vessels with elongated tubular or branched structures, while reducing computational complexity.

Table 4 shows that when additional information about the patient was learned in combination with image information, the proposed method using Self-Attention achieved higher accuracy than the method using image information alone. This result can be attributed to the fact that the introduction of Self-Attention was able to dynamically highlight correlations between important features and also effectively integrate complementary information between different modalities.

Examples of nodules misclassified by the proposed method are shown in Figure 4. In this case, the nodule was misclassified because the area occupied by lung tissue other than the nodule was large in the image, and it was difficult to accurately capture the shape features of the target. In addition, the limited number of images of nodules invaded by other lung tissue in the dataset used may have contributed to the misclassification. Future research needs to collect and expand image data showing different types of tissue erosion.



(i) axial (j) coronal (k) sagittal (l) fusion image

Fig.4. Misclassified image

5. Conclusion

In this paper, we proposed a method to identify pulmonary nodular shadows by combining multi-slice CT images and clinical information from medical records. The proposed method achieved a very high accuracy of Accuracy=0.983, TPR=0.987, and FPR=0.018, confirming the effectiveness of the proposed method.

In future studies, image data showing different types of tissue erosion will be collected and tested again. In addition, since image cropping is currently done manually, we hope to automate image generation using temporal subtraction imaging [11] or similar methods in the future.

References

1. International Agency for Research on Cancer, Cancer Today: Data Visualization Tools for Exploring the Global Cancer Burden in 2022, https://gco.iarc.fr/today/en.
2. Foundation for Promotion of Cancer Research, Cancer Statistics 2024, 2024.
3. Tamai, et al., "Detection of Nodular Shadows from Chest CT Images Using Deep Learning," Proceedings of the Biomedical Fuzzy Systems Society Annual Conference, vol. 33, pp. 26–29, 2020.
4. Li, et al., "Development of a Classification Model for Detecting Nodular Shadows in Pulmonary CT Images," Proceedings of the 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 1P4-J-10-01, 2019.
5. Z. Dai et al., "CoAtNet: Marrying Convolution and Attention for All Data Sizes", Advances in Neural Information Processing Systems 34, 2020.
6. S. Woo et al., “CBAM: Convolutional Block Attention Module”, European Conference on Computer Vision, pp.3-19, 2018.
7. A. Vaswani et al., "Attention Is All You Need", Advances in Neural Information Processing System, vol. 30, pp5998-6008, 2017.
8. N. Honda et al., “Identification of abnormal tissue from CT images using improved ResNet34”, International Conference on Control, Automation and Systems, pp.532-536, 2022.
9. K. He, X. Zhang, S. Ren, J. Sun, “Deep residual learning for image recognition”, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.770-778, 2016.
10. J. Hu, L. Shen, S. Albanie, G. Sun, E. Wu, “Squeeze-and-excitation networks”, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.7132-7141, 2018.
11. A. Kano et al., “Digital image subtraction of temporally sequential chest images for detection of interval change”, Medical Physics, Vol.21, No.3, pp.453-461, 1994.

Authors Introduction

Mr. Yuto Nishitaki

He received his Bachelor’s degree in Engineering in 2023 from the Faculty of Engineering, Kyushu Institute of technology in Japan. He is currently a master student in Kyushu Institute of Technology, Japan.

Tohru Kamiya, Prof. Ph.D

He received his B.A. degree in Electrical Engineering from Kyushu Institute of Technology in 1994, the Master and Ph.D. degree from Kyushu Institute of Technology in 1996 and 2001, respectively. He is a professor at the Department of Mechanical and Control Engineering at Kyushu Institute of Technology. His research interests are focused on image processing and medical application of image analysis.

Shoji Kido M.D. Ph.D.

He received his M.D. degree from Osaka University in 1988. He received his Ph.D. degrees in Medicine and Information Science from Osaka University in 1992 and 1999, respectively. He is a guest professor of Osaka University Institute for Radiation Science and Osaka University Graduate School of Medicine. His research interests are focused on the use of artificial intelligence in radiology.