

# Detection of Lung Nodules from Temporal Subtraction CT Image Using Elastic Net-Based Features Selection

Natsuho Baba<sup>1</sup>, Tohru Kamiya<sup>1</sup>, Takashi Terasawa<sup>2</sup>, Takatoshi Aoki<sup>2</sup>

<sup>1</sup> Kyushu Institute of Technology, 1-1 Sensui, Tobata, Kitakyushu, Fukuoka, 804-8550, Japan

<sup>2</sup> University of Occupational and Environmental Health, 1-1, Iseigaoka, Yahatanishi, Kitakyushu, Fukuoka, 807-8555, Japan

Email: baba.natsuho781@mail.kyutech.jp, kamiya@cntl.kyutech.ac.jp

## Abstract

CT (computed tomography) is mainly used to diagnose lung cancer. Many CT images impose a heavy burden on visual screening, so a CAD (computer-aided diagnosis) system is expected to reduce the burden. In this paper, we propose an image analysis method to detect lung nodules from chest CT images using machine learning techniques. The best results were obtained for the method using LightGBM with feature reduction by Elastic Net.

*Keywords:* Computer Aided Diagnosis, Machine Learning, Temporal Subtraction Technique, Radiomics, Elastic Net.

## 1. Introduction

Lung cancer is the most frequently diagnosed cancer worldwide and the leading cause of cancer-related death. Currently, it is primarily diagnosed through CT (computed tomography). Many CT images generated for a patient are burden for radiologists to interpret. Furthermore, there are inherent errors in radiologists' interpretation of medical images, and not all data and image information can be discovered [1]. To address these issues, CAD (computer-aided diagnosis) systems are being developed to assist in the interpretation and diagnosis of medical images.

The temporal subtraction system, one of the CAD system, subtracts a previous image from a current one to highlight interval changes. In the temporal subtraction image, normal tissues such as blood vessels that do not change over time can be removed. In addition, it can enhance the interval changes between the previous and current images including newly appearing lesions and changes of existing lesions [2].

In this paper, we propose a method to extract and classify lung cancer lesions from chest CT images with the goal of reducing the radiologists' workload and improving detection accuracy. First, we generate temporal subtraction images and extract candidate lesions that may include lung cancer or artifacts. Next, radiomics features are extracted from these lesions, and, finally, machine learning is used to classify lesions within these lesions.

## 2. Methodology

In this study, the proposed method is applied to 21 cases, with images obtained using a multidetector-row CT (MDCT) scanner. Each case consists of a set of current and previous CT images of the same subject, with the

temporal subtraction technique used to identify initial candidate lesions containing lung cancer or subtraction artifacts. Next, machine learning is applied to classify lesions within these initial lesions as lung cancer or non-cancer, incorporating radiomics features.

### 2.1. The temporal subtraction technique

The temporal subtraction technique involves creating a subtraction image by performing a difference operation between the current and previous images. The temporal subtraction image removes normal structures, such as blood vessels, that are commonly present in the present and past. The image highlights newly appearing lesions and changes in existing lesions over time.

### 2.2. Extraction of initial candidate regions from the temporal subtraction images

Due to the presence of many artifacts in the temporal subtraction images, mask processing is applied to extract initial candidate regions. In this study, only candidate lesion areas larger than 5 mm were targeted, and all other areas were background filled by mask processing. Machine learning was then used to classify whether these lesions were nodules or artefacts such as blood vessels.

### 2.3. Extraction of features

Radiomics features are extracted from the mask image using the selected lesion candidate regions. These features capture information such as tumor texture, intensity, heterogeneity, and shape from the chosen regions, enabling the automatic quantification and extraction of detailed features on a large scale [3]. In this study, shape and texture radiomics features were extracted using the open-source package Pyradiomics [4].

## 2.4. Feature selection

In this paper, 1026 radiomics features were computed from candidate lesion regions. However, with such a large number of features, there is an increased risk of overfitting, especially given the relatively limited amount of data. To address this, it is crucial to select only the most significant features, reducing redundant features and noise. Here, feature selection was performed using Elastic Net [5], [6], [7], a method that combines the strengths of both Lasso [5], [8] and Ridge regression [5], [9].

- (i) Lasso (Least Absolute Shrinkage and Selection Operator)

Lasso [5], [8] is a powerful technique that performs both regularization and feature selection. During feature selection, variables with non-zero coefficients after reduction are retained as part of the model. Additionally, Lasso enhances model interpretability by eliminating variables that are not related to the response variable.

- (ii) Ridge regression

Ridge regression [5], [9] like Lasso, is another powerful technique that performs regularization and feature selection. In this process, many predictors have non-zero coefficients and are extracted from a normal distribution. Ridge Regression is especially effective when there are many predictors with small individual effects, as it helps stabilize coefficient estimation in linear models with many correlated variables.

- (iii) Elastic Net

Elastic Net [5], [6], [7] is a penalized linear regression model that applies a combination of Lasso and Ridge regression penalties to the loss function during training. It performs both feature selection, a characteristic of Lasso, and coefficient stabilization, a characteristic of Ridge regression.

## 2.5. Machine learning classification

In this paper, the features selected by Elastic Net are used as explanatory variables in supervised learning, and the machine learning methods XGBoost and LightGBM are applied for comparison and validation.

- (i) XGBoost (Extreme Gradient Boosting)

XGBoost [10], [11] is a classification and regression method based on gradient boosting decision trees. It offers strong generalization capabilities, high scalability, and fast computation. The model is built through boosting with decision trees as the base learners.

- (ii) LightGBM (Light Gradient Boosting Machine)

LightGBM [12], [13] similar to XGBoost, is a classification and regression method based on gradient boosting decision trees. Compared to traditional gradient boosting methods, LightGBM

achieves faster learning speeds by using less memory, thus reducing computational complexity in parallel learning.

## 3. Experiments and Results

### 3.1. Evaluation

The evaluation metrics used are accuracy, TPR (True Positive Rate), FPR (False Positive Rate), and AUC (Area Under the Curve). As shown in Table 1, cases classified as lesions are considered positive cases, while cases classified as structures like blood vessels and the chest wall, representing normal tissues, are considered negative cases. The formulas for calculating accuracy, TPR, and FPR are provided in equations (1), (2), and (3). AUC represents the area under the receiver operating characteristic curve.

$$Accuracy = \frac{TN + TP}{TP + FP + FN + TN} \quad (1)$$

$$TPR = \frac{TP}{TP + FN} \quad (2)$$

$$FPR = \frac{FP}{TN + FP} \quad (3)$$

### 3.2. Results

In this study, temporal subtraction images were generated from 21 lung cancer cases acquired with an MDCT scanner. A total of 231 candidate lesion images obtained from these subtraction images were used as input data for machine learning classification. In this dataset, nodules that appeared over time were classified as positive lesions, while artifacts such as blood vessels and bones were classified as negative lesions. XGBoost and LightGBM were used individually for classification. Table 2 compares the methods without feature selection, with feature selection by Lasso (based on the previous method [14]), and with feature selection by Elastic Net. From Table 2, the best results were achieved using feature selection with Elastic Net and classification with LightGBM.

Example images from the experiment are shown in Figure 1. In this figure, (a) is the current image, (b) is the previous image, (c) is the temporal subtraction image, and (d) is the mask image generated using the temporal subtraction image. Arrows indicate lung cancer areas and image (d) shows the area is enhanced as a white area.

Table 1. Evaluation basis

	Predicted Positive	Predicted Negative
True Label Positive	True Positive	False Negative
True Label Negative	False Positive	True Negative

Table 2. Identification result (Acc. : Accuracy)

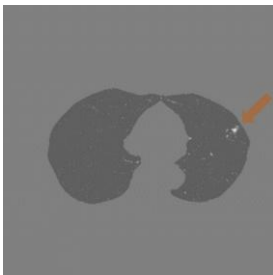
		TPR	FPR	Acc.	AUC
XGBoost	No feature reduction	67.29	30.60	67.11	0.721
	Lasso	67.90	26.40	70.12	0.786
	Elastic Net	83.60	21.04	80.96	0.836
LightGBM	No feature reduction	77.77	20.33	77.96	0.799
	Lasso	75.47	20.54	76.64	0.792
	Elastic Net	81.69	21.41	79.24	0.835



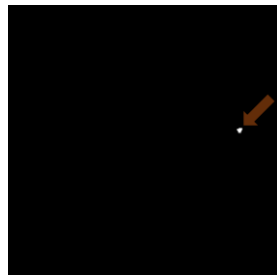
(a) The current image



(b) The previous image



(c) The temporal subtraction image



(d) The mask image

Fig.1 Example images from the experiment

#### 4. Discussion

Table 2 shows that overall accuracy is higher when feature selection is performed using Elastic Net compared to when it is not applied. This is because Elastic Net combines the coefficient shrinkage of ridge regression and the feature selection capability of Lasso, effectively removing redundant features. Comparing the previous method [14] with the Elastic Net-based feature selection, the proposed method extracted features from similar wavelet transform statistics and sub-bands, preserving important features even in the presence of high correlations among predictors. This confirms that Elastic Net feature selection could retain relevant features without losing any critical information.

The images in Figure 1 illustrate improved classification results after switching from Lasso-based feature selection to Elastic Net-based selection. In the temporal subtraction image (c) generated from the current image (a) and the previous image (b), many artifacts remained when classification was performed using Lasso-based feature selection. However, after switching to Elastic Net feature selection, the region was correctly classified as lung cancer. This suggests that Elastic Net feature selection improved accuracy by preserving useful features and minimizing the influence of normal tissue features, even in noise-rich temporal subtraction images.

#### 5. Conclusion

In this paper, we proposed a method for extracting and classifying lesions from chest CT images to reduce radiologists' workload and improve lesion detection accuracy. The method involves the following steps. First, initial candidate regions were identified using temporal subtraction on current and previous chest CT images. Next, radiomics features were extracted from these candidate regions, and feature selection was performed using Elastic Net to retain only significant features. Finally, the selected features were used as explanatory variables, and supervised learning was applied to classify unknown data as either lesion or normal tissue.

The proposed method was evaluated on 231 candidate lesion areas, with an AUC of 0.835, accuracy of 79.24%, TPR of 81.69%, and FPR of 21.41% when using LightGBM with Elastic Net-based feature selection.


Future work will focus on expanding the dataset, identifying optimal feature selection methods, and implementing ensemble learning with XGBoost and LightGBM for classification.

#### References

1. F. Shariaty and M. Mousavi, "Application of CAD systems for the automatic detection of lung nodules", *Informatics in Medicine Unlocked*, Vol. 20, 2020, pp. 391-397.
2. Y. Yoshino et al., "Automatic classification of lung nodules on MDCT images with the temporal subtraction technique", *International Journal of Computer Assisted Radiology and Surgery*, Vol. 12, 2017, pp. 1789-1798.
3. M. E. Mayerhoefer et al., "Introduction to Radiomics", *Journal of Nuclear Medicine*, Vol. 61, No. 4, 2020, pp. 488-495.
4. J. J. M. van Griethuysen et al., "Computational radiomics system to decode the radiographic phenotype", *Cancer Research*, Vol. 77, 2017, pp. 104-107.
5. J. O. Ogutu et al., "Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions", *BMC Proceedings*, Vol. 6, No. 10, 2012.
6. J. Morgado et al., "Machine learning and feature selection methods for EGFR mutation status prediction in lung cancer" *Applied Sciences*, Vol. 11, 2021.

7. H. Zou and T. Hastie, "Regularization and variable selection via the elastic net" *Statistical Methodology*, Vol. 67, 2005, pp. 301-320.
8. V. Fonti and E. Belister, "Feature selection using lasso", *VU Amsterdam Res. Paper Business Anal*, Vol. 30, 2017, pp. 1-25.
9. M. Togacar et al., "Application of breast cancer diagnosis based on a combination of convolutional neural networks, ridge regression and liner discriminant analysis using invasive breast cancer images processed with autoencoders", *Medical Hypotheses*, Vol. 135, 2020, pp. 135-145.
10. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system", *Machine Learning*, Vol. 3, 2016, pp. 785-794.
11. S. Li and X. Zhang, "Research on orthopedic auxiliary classification and prediction model based on XGBoost algorithm", *Neural Computing and Applications*, Vol. 32, 2020, pp. 1971-1979.
12. J. Cai et al., "An assembly-level neutronic calculation method based on LightGBM algorithm", *Annals of Nuclear Energy*, Vol. 150, 2021, pp. 871-881.
13. Y. Wang and T. Wang, "Application of improved LightGBM model in blood glucose prediction", *Applied Sciences*, Vol. 10, No. 9, 2020, pp. 3227-3243.
14. N. Baba et al., "A detection method for nodular shadows from temporal subtraction images using a machine learning technique incorporated radiomics features", *Internatinal Conference on Control, Automation and Systems*, 2023, pp. 871-881.

Takatoshi Aoki, Prof. M.D., Ph.D



He is a professor of Radiology at University of Occupational and Environmental Health, Japan, and the vice president of Japanese Society of Thoracic Radiology. His clinical and research interests include the imaging modalities of lung cancer, respiratory function.

---

---

### Authors Introduction

Ms. Natsuho Baba



She received her Bachelor's degree in Engineering in 2023 from the Faculty of Engineering, Kyushu Institute of technology in Japan. She 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 in the

Department of Mechanical and Control Engineering at Kyushu Institute of Technology. His research interests are focused on image processing, medical application of image analysis, etc.

Takashi Terasawa, M.D., Ph.D



He received his M.D. degree from University of Occupational and Environmental Health, Japan in 2011 and Ph.D. in medicine in 2020. He is a radiologist at the University of Occupational and Environmental Health, Japan.