# Computer-aided diagnosis of lung nodule using gradient tree boosting and Bayesian optimization

**Mizuho Nishio**[1,2]*, **Mitsuo Nishizawa**[3], **Osamu Sugiyama**[2], **Ryosuke Kojima**[4],
**Masahiro Yakami**[1,2], **Tomohiro Kuroda**[5], **Kaori Togashi**[1]

**1** Department of Diagnostic Imaging and Nuclear Medicine, Kyoto University Graduate School of Medicine, Shogoin, Sakyo-ku, Kyoto, Kyoto, Japan, **2** Preemptive Medicine and Lifestyle Disease Research Center, Kyoto University Hospital, Shogoin, Sakyo-ku, Kyoto, Kyoto, Japan, **3** Department of Radiology, Osaka Medical College, Takatsuki, Osaka, Japan, **4** Department of Biomedical Data Intelligence, Kyoto University Graduate School of Medicine, Sakyo-ku, Kyoto, Kyoto, Japan, **5** Division of Medical Information Technology and Administrative Plannnig, Kyoto University Hospital, Shogoin, Sakyo-ku, Kyoto, Kyoto, Japan

* nmizuho@kuhp.kyoto-u.ac.jp, jurader@yahoo.co.jp

## Abstract

We aimed to evaluate a computer-aided diagnosis (CADx) system for lung nodule classification focussing on (i) usefulness of the conventional CADx system (hand-crafted imaging feature + machine learning algorithm), (ii) comparison between support vector machine (SVM) and gradient tree boosting (XGBoost) as machine learning algorithms, and (iii) effectiveness of parameter optimization using Bayesian optimization and random search. Data on 99 lung nodules (62 lung cancers and 37 benign lung nodules) were included from public databases of CT images. A variant of the local binary pattern was used for calculating a feature vector. SVM or XGBoost was trained using the feature vector and its corresponding label. Tree Parzen Estimator (TPE) was used as Bayesian optimization for parameters of SVM and XGBoost. Random search was done for comparison with TPE. Leave-one-out cross-validation was used for optimizing and evaluating the performance of our CADx system. Performance was evaluated using area under the curve (AUC) of receiver operating characteristic analysis. AUC was calculated 10 times, and its average was obtained. The best averaged AUC of SVM and XGBoost was 0.850 and 0.896, respectively; both were obtained using TPE. XGBoost was generally superior to SVM. Optimal parameters for achieving high AUC were obtained with fewer numbers of trials when using TPE, compared with random search. Bayesian optimization of SVM and XGBoost parameters was more efficient than random search. Based on observer study, AUC values of two board-certified radiologists were 0.898 and 0.822. The results show that diagnostic accuracy of our CADx system was comparable to that of radiologists with respect to classifying lung nodules.

## Introduction

Lung cancer is the leading cause of cancer deaths in the United States [1] because it is frequently diagnosed at an advanced stage, and this prevents effective treatment. Results from the

National Lung Screening Trial (NLST) show that compared with chest X-ray screening, lung cancer screening with low-dose CT significantly reduced lung cancer mortality among heavy smokers by detecting lung cancers at an early stage [2,3]. However, false positives in low-dose CT screening can be problematic and can result in unnecessary follow-up CT, positron emission tomography, or invasive procedures. In NLST, 96.4% of the positive results in the low-dose CT group were false positives [2,3].

Computer-aided diagnosis (CAD) has the potential of optimizing radiologists' workloads. CAD can assist radiologists in detection (CADe) and differentiation (CADx) of lung nodules [4–23]. For example, CADx is useful for assisting radiologists in differentiating between benign and malignant lung nodules [6], and it is expected that CADx is useful for reducing false positives in lung cancer screening with low-dose CT.

Gradient tree boosting is superior to off-the-shelf classifiers such as random forest or support vector machine (SVM) [24,25]. Because performance of CADx is affected by machine learning algorithms, gradient tree boosting may improve the performance of CADx. However, to the best of our knowledge, no other study has investigated the usefulness of gradient tree boosting in CADx of lung nodules. In our study, we used XGBoost as an implementation of gradient tree boosting [25] and applied it to CADx system of lung nodules.

It is necessary to optimize parameters of machine learning algorithms to ensure good performance. Grid search has been frequently used for this purpose [26]. However, when the number of parameters is increased, grid search is not feasible because of its computational cost. As an alternative to grid search, random search and Bayesian optimization were used for parameter optimization [27,28]. Because XGBoost has many parameters, random search and Bayesian optimization are suitable for parameter optimization.

The purpose of the current study was to develop and evaluate the CADx system, focusing on (i) usefulness of the conventional CADx system (hand-crafted imaging feature + machine learning algorithm), (ii) comparison between SVM and XGBoost as machine learning algorithms, and (iii) effectiveness of parameter optimization using random search and Bayesian optimization. Herein, a variant of the local binary pattern (LBP) [11,18,29–31] was used as the hand-crafted imaging feature for calculating a feature vector that was fed into the machine learning algorithms.

## Methods

This study used anonymized data from a public database. Regulations of Japan did not require institutional review board approval for use of a public database.

### CT images

Our CADx system was tested using chest CT images obtained from The Cancer Imaging Archive (TCIA). TCIA is an open database of medical images, mainly consisting of CT, MRI, and nuclear medicine images that are stored as anonymized DICOM data. We used two sets of chest CT images from TCIA; one set from the LUNGx Challenge and one from the NSCLC Radiogenomics [20,21,32–35]. The LUNGx Challenge provided 60 test sets of chest CT images with 10 calibration sets. The 60 test sets included 73 lung nodules; a list of these nodules is available on the LUNGx Challenge website [34]. Among the 73 nodules from the LUNGx Challenge, 36 were lung cancers and 37 were benign. In NSCLC Radiogenomics, each of 26 sets of chest CT images included lung cancer. By combining data from LUNGx Challenge with those from NSCLC Radiogenomics, a total of 99 lung nodules (62 lung cancers and 37 benign nodules) were used for the development and evaluation of our CADx system.

## Image preprocessing

First, CT images were loaded, and their voxel sizes were converted to $1 \times 1 \times 1$ mm. Next, the center was determined for each of the 99 nodules. Coordinates of the center of the lung nodules were provided via spreadsheet in the LUNGx Challenge and utilized here. Conversely, no such information was available for NSCLC Radiogenomics. Therefore, the center of the lung nodule was visually validated by two board-certified radiologists (M.N. and M.N.). A $64 \times 64 \times 64$ 3D bounding box was set for each nodule, and CT images inside the bounding box were cropped. The cropped 3D CT images were analyzed as the input to our CADx system. Areas of the CT images outside the bounding box were not assessed. As shown, manual segmentation of lung nodule was not needed in this preprocessing (only the center of lung nodule was necessary).

## Calculation of a feature vector

The local binary pattern on three orthogonal planes (LBP-TOP) was used for calculating a feature vector [11,18,29,30,31]. Naïve implementation of 2D LBP was represented as follows:

$$LBP(x, R, P) = \sum_{i=0}^{P-1} 2^i \times s(d_i)$$

$$d_i = I(n(x, R, i)) - I(x), \tag{I}$$

where $x$ is the center pixel where LBP is calculated; $P$ is the number of samples; $n(x, R, i)$ is the $i$th neighbor pixel around the center pixel $x$ and the distance between the center pixel $x$ and the neighbor pixel is $R$; $I(u)$ is the CT density of pixel $u$ and $s(v)$ is an indicator function, where $s(v)$ is 1 if $v \geq 0$ and 0 otherwise. We used a uniform pattern and rotation invariant type instead of naïve implementation as naïve implementation cannot handle large $P$ values because they make feature vectors too long. Both uniform pattern and rotation invariant type can enhance the robustness of LBP as a feature vector. To utilize LBP in 3D images, LBP-TOP was used in this study. In LBP-TOP, 2D LBP was calculated on the XY, XZ and YZ planes and the texture information on other 3D planes was ignored. Then, the results of 2D LBP on the XY, XZ and YZ planes were converted into 1D histograms, which were concatenated. In this method, rotation invariance of LBP was retained only in the rotation of XY, XZ and YZ planes. To use LBP-TOP as feature vectors of CADx of lung nodules, 3D cropped CT images were evaluated with uniform pattern and rotation invariant type of LBP-TOP, and 1D feature vectors were calculated.

## Machine learning algorithm

Our CADx system was built using SVM or XGBoost [24,25]. Implementations of SVM and XGBoost were freely available. SVM or XGBoost were trained using the feature vector obtained by LBP-TOP and its corresponding label. SVM is a widely used machine learning algorithm, and we used SVM with kernel trick (radial basis function) in this study. XGBoost builds an efficient classifier using gradient tree boosting. Gradient tree boosting is invariant to scaling of a feature vector, and it can determine higher-order interaction between a feature vector. The usefulness of XGBoost has been validated in a number of machine learning and data mining challenges (Please refer to the Introduction section of [25]). Gradient tree boosting is trained in an additive manner. At each time step $t$, it grows a tree to minimize the residual of the current model. Formally, the objective function of XGBoost can be described as follows:

$$L^t = \sum_{i=1}^{n} l(y_i, y_i^{t-1} + f_t(x_i)) + \Omega(f_t), \tag{II}$$

where $x_i$ and $y_i$ are the feature vector and its label at the $i^{th}$ instance, $n$ is the number of training data, $y_i^{t-1}$ is the prediction of the $i^{th}$ instance at the $t-1^{th}$ iteration, $f_t$ is a new tree that classifies the $i^{th}$ instance using $x_i$, $l$ denotes a loss function that measures the difference between the label and the prediction at the last step plus the new tree output, and $\Omega$ is the regularization term that penalizes the complexity of the new tree.

## Parameters

The following parameter space was used for parameter optimization.

- For SVM, $C$ and $\gamma$ were used for controlling SVM with a radial basis function kernel [26]; C is a parameter for balancing classification error and regularization, and $\gamma$ is a free parameter for bandwidth of radial basis function kernel. The range of $C$ and $\gamma$ were as follows: $C$, $1.0 \times 10^{-5}$–$1.0 \times 10^{5}$ and $\gamma$, $1.0 \times 10^{-5}$–$1.0 \times 10^{5}$.

- For XGBoost, parameters and their range were as follows: eta, 0.2–0.6; max_depth, 1–13; min_child_weight, 1–10; gamma, 0–1. The concise explanation of XGBoost parameters are as follows: eta for step size shrinkage used in updating a tree [25], max_depth for maximum depth of a tree, min_child_weight for minimum sum of instance weight needed in a child (If the tree partition step results in a leaf node with the sum of instance weight less than min_-child_weight, the process of tree building will stop further partitioning), gamma for minimum loss reduction required to make a further partition on a leaf node of the tree. The detail of these parameters can be available elsewhere [36].

- LBP-TOP has two parameters ($R$ and $P$). The values of $R$ and $P$ were as follows: $R = 7, 8$ and $P = 40, 48$.

## Parameter optimization

The parameter space was defined in the previous subsection. Here we denoted the parameters as $\boldsymbol{\theta}$. When using machine learning algorithm $A$ (SVM or XGBoost) and the parameter $\boldsymbol{\theta}$, we trained $A$ using training data and validated its performance using validation data. We used $L$ $(A, \boldsymbol{\theta}, D_{train}, D_{valid})$ to denote the validation loss that $A$ achieved on validation data $D_{valid}$ when $A$ was trained on $\boldsymbol{\theta}$ and $D_{train}$. The parameter optimization problem under $K$-fold cross-validation was then to minimize the black box function:

$$f(\boldsymbol{\theta}) = \frac{1}{K}\sum_{i=1}^{K} L(A, \boldsymbol{\theta}, D_{train}^{i}, D_{valid}^{i}),$$  (III)

where $D_{train}^{i}$ and $D_{valid}^{i}$ were training data and validation data of the $i$-th fold of $K$-fold cross-validation, respectively. Bayesian optimization was used for optimizing this black box function $f(\boldsymbol{\theta})$ and for searching for the optimal parameter $\boldsymbol{\theta}$. Tree Parzen Estimator (TPE) was utilized for solving this problem [27]. Random search was used to compare the performance of TPE. Number of trials for TPE or random search was as follows: 10, 100, 200, and 1000.

## Software and outline of CADx

Code of our CADx system and binary data of lung nodules are available as S1 File of Supporting information. For our CADx system, python (version 2.7, https://www.python.org/), and the following packages were used: scikit-image (version 0.18.1, http://scikit-image.org/), sciki-learn (version 0.18.1, http://scikit-learn.org/), xgboost (version 0.6, http://xgboost.readthedocs.io/en/latest/), and hyperopt (version 0.1, http://hyperopt.github.io/hyperopt/). An outline of our CADx system is shown in Fig 1.
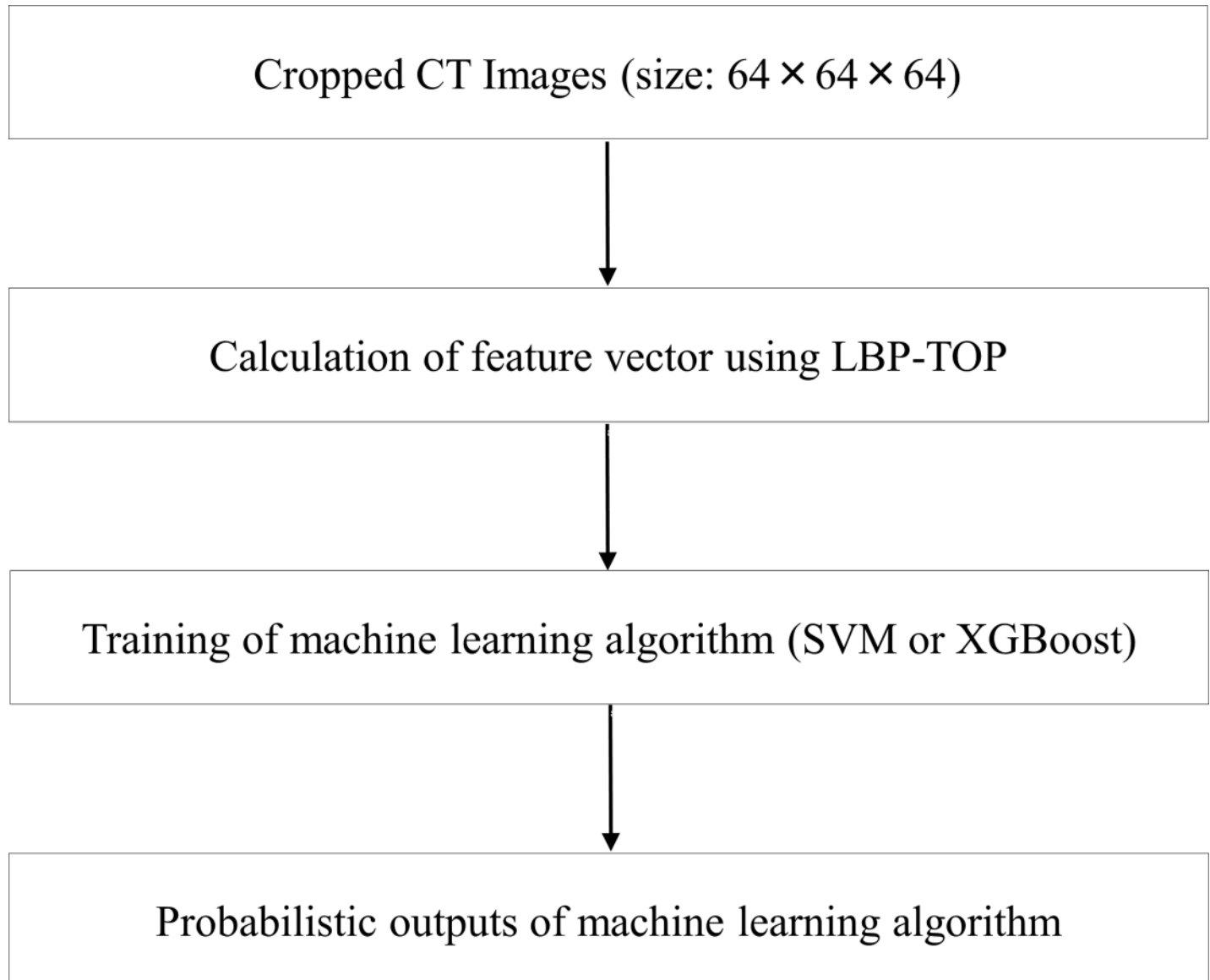
**Fig 1. Outline of our CADx system.** Abbreviations: CADx, computer-aided diagnosis; LBP-TOP, local binary pattern on three orthogonal planes; SVM, support vector machine.

### Observer study

Two board-certified radiologists (M.N. and M.Y.) were included in observer study for assessing the 99 lung nodules. They evaluated CT images of lung nodules with lung window condition (window width = 1500 HU and window level = −600 HU), and could change the window condition if necessary. They rated their suspicion of malignancy with 10-point scale (1 = definitely benign lung nodule; 10 = definitely lung cancer).

### Statistical analysis

Leave-one-out cross-validation was used for optimizing and evaluating the performance of our CADx system. Validation loss under leave-one-out cross-validation was used for

**Table 1. Results of CADx when using SVM and parameter optimization.**

| Algorithm | Number of trial | Validation loss | AUC | Accuracy |
|---|---|---|---|---|
| Random | 10 | 0.528 | 0.792 | 0.734 |
| Random | 100 | 0.481 | 0.832 | 0.780 |
| Random | 200 | 0.460 | 0.848 | 0.794 |
| Random | 1000 | 0.451 | 0.849 | 0.789 |
| TPE | 10 | 0.515 | 0.797 | 0.724 |
| TPE | 100 | 0.461 | 0.847 | 0.802 |
| TPE | 200 | 0.458 | 0.846 | 0.792 |
| TPE | 1000 | 0.453 | 0.850 | 0.797 |

Abbreviation: computer-aided diagnosis, CADx; support vector machine, SVM; Tree Parzen Estimator, TPE; area under the curve, AUC.

parameter optimization. After parameter optimization, probabilistic outputs of our CADx system with optimal parameters were analyzed using accuracy and area under the curve (AUC) of receiver operating characteristic (ROC) analysis. Classification results of our CADx system were output as probabilities of lung cancer to calculate AUC. For each number of trial, AUC and accuracy were calculated 10 times, and their averages were obtained. For observer study, AUC and accuracy of the two board-certified radiologists were also calculated for comparison between our CADx system and radiologists.

## Results

The averaged validation loss, AUC, and accuracy of our CADx system are shown in Tables [1] and [2] and Figs [2]–[4]. [S1 Table] of Supporting information shows raw results of validation loss, AUC, and accuracy of our CADx system for each setting. Tables [1] and [2] show the averages of the raw results listed in [S1 Table] of Supporting information. Comparing the results depicted in Tables [1] and [2], XGBoost was generally superior to SVM. According to [Table 1], the best averaged AUC of SVM was 0.850 when using TPE and number of trials = 1000. [Table 2] shows that the best averaged AUC of XGBoost was 0.896 when using TPE and number of trials = 1000. According to [S1 Table] of Supporting information, the best AUC and accuracy of SVM was 0.855 and 0.834, respectively, and the best AUC and accuracy of XGBoost was 0.903 and 0.859, respectively. In XGBoost, the averaged AUC of TPE was better than that of random search when the number of trials was 100, 200, or 1000. In SVM, the averaged AUC of TPE was better than that of the random search when the number of trials was 100. However, when the number

**Table 2. Results of CADx when using XGBoost and parameter optimization.**

| Algorithm | Number of trial | Validation loss | AUC | Accuracy |
|---|---|---|---|---|
| Random | 10 | 0.488 | 0.838 | 0.756 |
| Random | 100 | 0.451 | 0.864 | 0.771 |
| Random | 200 | 0.440 | 0.868 | 0.784 |
| Random | 1000 | 0.422 | 0.878 | 0.806 |
| TPE | 10 | 0.494 | 0.838 | 0.762 |
| TPE | 100 | 0.427 | 0.876 | 0.811 |
| TPE | 200 | 0.419 | 0.881 | 0.804 |
| TPE | 1000 | 0.394 | 0.896 | 0.820 |

Abbreviation: computer-aided diagnosis, CADx; support vector machine, SVM; Tree Parzen Estimator, TPE; area under the curve, AUC.
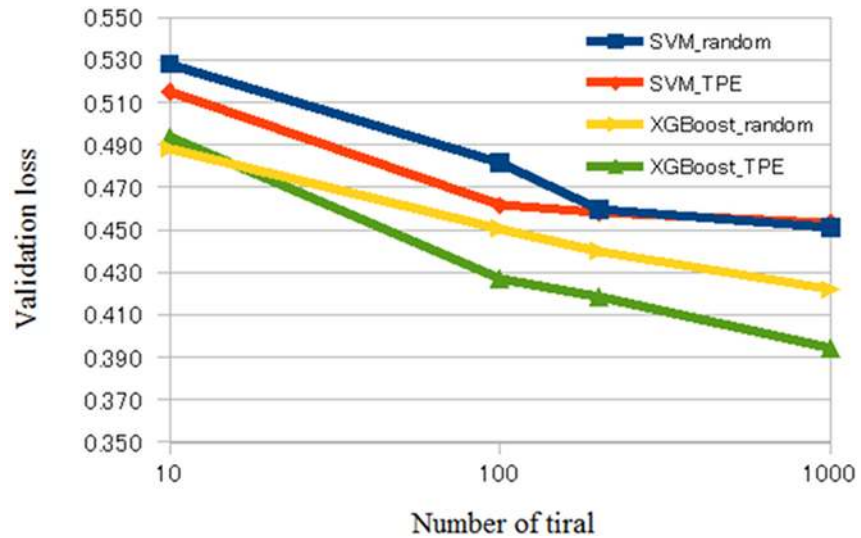
**Fig 2. Validation loss of CADx.** Abbreviations: CADx, computer-aided diagnosis; SVM, support vector machine.

https://doi.org/10.1371/journal.pone.0195875.g002

of trials was 10, the difference of the averaged AUC was minimal between random search and TPE in SVM and XGBoost. In addition, in SVM, the difference of averaged AUC was minimal between random search and TPE when the number of trials was 200 or 1000.

From the results of observer study, AUC and accuracy of the two board-certified radiologists were as follows: radiologist1, 0.898 and 0.838; radiologist2, 0.822 and 0.717. Fig 5 shows the corresponding ROC curves of the two radiologists.

## Discussion

In this study, we used two different sets of CT images for evaluating our CADx system; one set from the LUNGx Challenge and the other from the NSCLC Radiogenomics. Using XGBoost and TPE, the best averaged AUC under leave-one-out cross-validation was 0.896 (the best
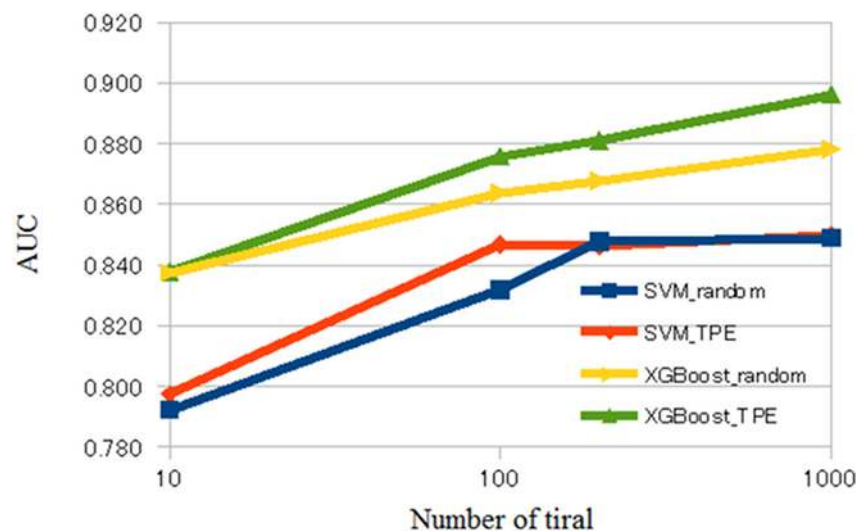


**Fig 3. AUC of CADx.** Abbreviations: CADx, computer-aided diagnosis; SVM, support vector machine; AUC, area under the curve.

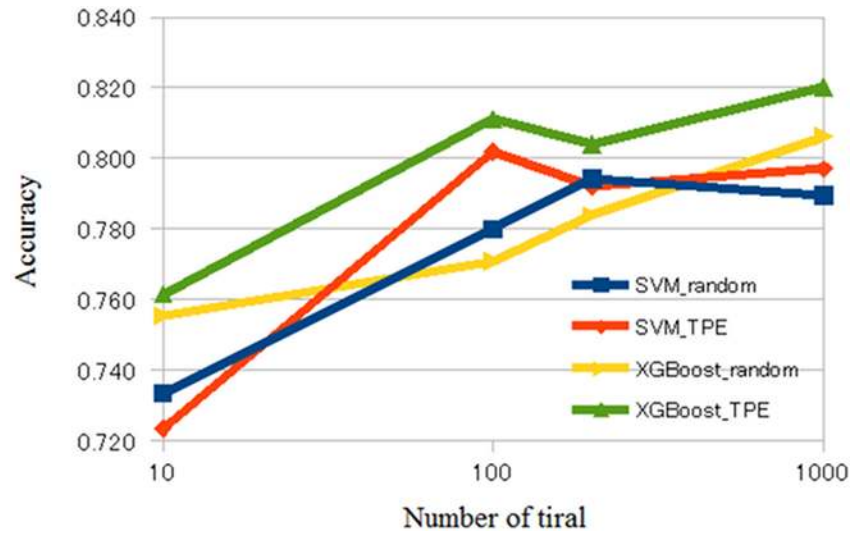https://doi.org/10.1371/journal.pone.0195875.g003

**Fig 4. Accuracy of CADx.** Abbreviations: CADx, computer-aided diagnosis; SVM, support vector machine.

AUC under leave-one-out cross-validation was 0.903). AUC values of the two board-certified radiologists were 0.898 and 0.822. These results of our CADx system show the following three main points; (i) the diagnostic accuracy of our conventional CADx system (hand-crafted imaging feature + machine learning algorithm) might be comparable to that of the radiologists; (ii) XGBoost was better than SVM; and (iii) parameter optimization with TPE was better than that with random search.

From the comparison between our CADx system and radiolgists, we speculated that the diagnostic accuracy of our CADx system was comparable to that of the radiologists with respect to classifying lung nodules; AUC values of both our CADx system and radiologists were nearly 0.9.
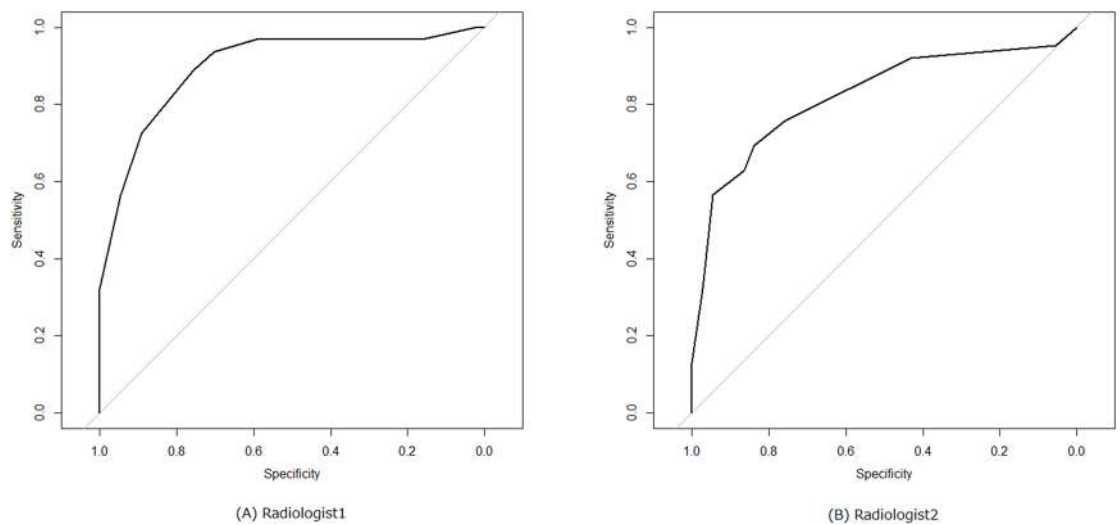


**Fig 5. ROC curves of two radiologists.** Note: (A) radiologist1 and (B) radiologist2. Abbreviations: ROC, receiver operating characteristic.

A few previous studies have utilized XGBoost for developing a clinical model. One study used XGBoost for classifying symptom severity based on text information in the form of psychiatrist notes [37]. Another study showed the usefulness of XGBoost for differentiation between subjects with epilepsy and healthy subjects using patients' cerebral activity assessed by functional MRI [38]. In conjunction with the results of these studies, we found that XGBoost was useful for developing an efficient and reliable clinical model. Although SVM was widely used as a machine learning algorithm in CADx, AUC of CADx using XGBoost was better than that using SVM in our study. A prime reason for the superiority of XGBoost to SVM is invariant to scaling of a feature vector. As well-known kernels for SVM, such as radial basis function and linear kernels, are scale dependent, the output value of SVM is affected by scaling of a feature vector. In addition, because results of LBP-TOP can be viewed as a type of categorical data, it is speculated that tree-based classifiers, such as XGBoost, are more suitable for feature vectors obtained by LBP-TOP than SVM.

Previous studies have shown that Bayesian optimization was useful in several domains of clinical application [39–42]. The results of the current study are compatible with those of the previous studies. Figs 2–4 show that, in general, TPE is better than random search for optimizing parameters in SVM and XGBoost. However, when the number of trials was 10, the difference in performance between TPE and random search was minimal. This result suggests that the small number of trials (10) hindered parameter optimization of SVM and XGBoost. When the number of trials was 200 or 1000 in SVM, the difference in performance between random search and TPE was also minimal. Because parameter space of SVM was narrower than that of XGBoost in the current study, we surmised that both random search and TPE could almost fully optimize parameters and the difference in performance may be minimal.

We used the conventional CADx system because the number of lung nodule was less than 100. Although results of recent studies suggest that deep learning is superior to conventional machine learning algorithms, deep learning requires a large number of training data [43]. Therefore, we focused on the conventional CADx system in the current study. Generally speaking, it is more difficult to collect training data for medical image analysis than for other fields of image analysis. When the number of training data is limited, our methodology may be more useful than deep learning.

Because we used the established software (LBP, XGBoost, and TPE), our CADx system was technically simple. However, our results show that diagnostic accuracy of our system might be comparable to that of the radiologists. The previous studies required nodule segmentation, calculation of many types of image features, or radiological findings for successful differentiation of lung nodules [8, 17]. The simplicity was the main advantage of our CADx system.

The previous study shows that AUC value of CADx system was more than 0.8 by using the 73 lung nodules of LUNGx Challenge and SVM with a linear kernel [11]. Because of differences in CT images, quality of labels, and kernel type of SVM, it is difficult to precisely compare the diagnostic accuracy of the CADx system between the current study and the previous study. However, the diagnostic accuracy of our CADx system using SVM might be comparable to that of the previous study.

There were several limitations to our study. First, the number of lung nodules was relatively small. Hence, our CADx system might overfit the dataset of the current study. We speculated that the possibility of overfitting was not so high because this dataset consisted of two different sets of CT images and the conditions and parameters of the images were variable (i.e. variability in use of contrast material and thickness of CT images). However, this speculation might be optimistic as we cannot deny a possibility that our CADx system overfitted the dataset of the current study. Future studies should be conducted using a large number of lung nodules to prevent overfitting and evaluate the generalizability of our CADx system. Second, this study

focused on the investigation of technical usefulness of XGBoost and Bayesian optimization from the viewpoint of CADx of lung nodules, and we ignored the clinical usefulness of our CADx system. Because the results of our study showed that the diagnostic ability of our CADx system may be comparable to that of radiologists, we expect that our CADx system will be useful for classifying lung nodules in a practical clinical setting. Third, the parameter space was relatively limited in this study. The parameters of our study were divided into two types: the parameter of machine learning algorithms (i.e. $C$ for SVM and eta for XGBoost) and the parameter of feature vectors ($R$ and $P$ of LBP). Because the results of parameter optimization were not stable when the parameter space of feature vectors was wide, we restricted the parameter space of feature vectors in our study. Last, we did not compare our CADx system with CADx using deep learning. We plan to develop a CADx system with deep learning and will use TPE for parameter optimization of deep learning in a future study.

In conclusion, XGBoost was better than SVM for classifying lung nodules. For optimizing parameters of both SVM and XGBoost, Bayesian optimization was more efficient than random search. Although our results were preliminary, the diagnostic accuracy of our CADx system may be comparable to that of radiologists for classifying lung nodules.

## Supporting information

**S1 Table. Raw results of parameter optimization.** S1 Table shows the raw results of validation loss, AUC, accuracy of our CADx system for each setting. Tables 1 and 2 show the averages of the raw results.
(DOCX)

**S1 File. Code of our CADx system and binary data of lung nodules.** S1 File includes Python script of our CADx system and binary data of lung nodules stored as NPY.
(ZIP)

## Acknowledgments

## Author Contributions

**Conceptualization:** Mizuho Nishio, Masahiro Yakami.

**Data curation:** Mizuho Nishio, Mitsuo Nishizawa, Masahiro Yakami.

**Formal analysis:** Mizuho Nishio.

**Funding acquisition:** Mizuho Nishio.

**Investigation:** Mizuho Nishio.

**Methodology:** Mizuho Nishio, Osamu Sugiyama, Ryosuke Kojima.

**Project administration:** Mizuho Nishio, Tomohiro Kuroda.

**Resources:** Mizuho Nishio.

**Software:** Mizuho Nishio.

**Supervision:** Tomohiro Kuroda, Kaori Togashi.

**Validation:** Mizuho Nishio.

**Visualization:** Mizuho Nishio.

**Writing – original draft:** Mizuho Nishio.

**Writing – review & editing:** Mizuho Nishio, Mitsuo Nishizawa, Osamu Sugiyama, Ryosuke Kojima, Masahiro Yakami, Tomohiro Kuroda, Kaori Togashi.

# References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. *CA Cancer J Clin*. 2016; 66(1):7–30. https://doi.org/10.3322/caac.21332 PMID: 26742998

2. The National Lung Screening Trial Research Team. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. *N Engl J Med*. 2011; 365(5):395–409. https://doi.org/10.1056/NEJMoa1102873 PMID: 21714641

3. Aberle DR, Abtin F, Brown K. Computed tomography screening for lung cancer: has it finally arrived? Implications of the national lung screening trial. *J Clin Oncol*. 2013; 31(8):1002–1008. https://doi.org/10.1200/JCO.2012.43.3110 PMID: 23401434

4. Suzuki K. A review of computer-aided diagnosis in thoracic and colonic imaging. *Quant Imaging Med Surg*. 2012; 2(3):163–176. https://doi.org/10.3978/j.issn.2223-4292.2012.09.02 PMID: 23256078

5. Lee SLA, Kouzani AZ, Hu EJ. Automated detection of lung nodules in computed tomography images: a review. *Mach Vis Appl*. 2012; 23(1):151–163. https://doi.org/10.1007/s00138-010-0271-2

6. El-Baz A, Beache GM, Gimel'farb G, Suzuki K, Okada K, Elnakib A, et al. Computer-aided diagnosis systems for lung cancer: challenges and methodologies. *Int J Biomed Imaging*. 2013; 2013:942353. https://doi.org/10.1155/2013/942353 PMID: 23431282

7. lin D, Vasilakos AV, Tang Y, Yao Y. Neural networks for computer-aided diagnosis in medicine: A review. *Neurocomputing*. 2016; 216:700–708. https://doi.org/10.1016/j.neucom.2016.08.039

8. Kawagishi M, Chen B, Furukawa D, Sekiguchi H, Sakai K, Kubo T, et al. A study of computer-aided diagnosis for pulmonary nodule: comparison between classification accuracies using calculated image features and imaging findings annotated by radiologists. *Int J Comput Assist Radiol Surg*. 2017; 12(5):767–776. https://doi.org/10.1007/s11548-017-1554-0 PMID: 28285338

9. Benzakoun J, Bommart S, Coste J, Chassagnon G, Lederlin M, Boussouar S, et al. Computer-aided diagnosis (CAD) of subsolid nodules: Evaluation of a commercial CAD system. *Eur J Radiol*. 2016; 85(10):1728–1734. https://doi.org/10.1016/j.ejrad.2016.07.011 PMID: 27666609

10. de Carvalho Filho AO, Silva AC, Cardoso de Paiva A, Nunes RA, Gattass M. Computer-Aided Diagnosis of Lung Nodules in Computed Tomography by Using Phylogenetic Diversity, Genetic Algorithm, and SVM. *J Digit Imaging*. May 2017:1–11.

11. Nishio M, Nagashima C. Computer-aided Diagnosis for Lung Cancer: Usefulness of Nodule Heterogeneity. *Acad Radiol*. 2017; 24(3):328–336. https://doi.org/10.1016/j.acra.2016.11.007 PMID: 28110797

12. Cao P, Liu X, Yang J, Zhao D, Li W, Huang M, et al. A multi-kernel based framework for heterogeneous feature selection and over-sampling for computer-aided detection of pulmonary nodules. *Pattern Recognit*. 2017; 64:327–346. https://doi.org/10.1016/j.patcog.2016.11.007

13. Doi K. Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. *Comput Med Imaging Graph*. 2007; 31(4–5):198–211. https://doi.org/10.1016/j.compmedimag.2007.02.002 PMID: 17349778

14. Matsumoto S, Kundel HL, Gee JC, Gefter WB, Hatabu H. Pulmonary nodule detection in CT images with quantized convergence index filter. *Med Image Anal*. 2006; 10(3):343–352. https://doi.org/10.1016/j.media.2005.07.001 PMID: 16542867

15. Suzuki K. Computer-Aided Detection of Lung Cancer. In: *Image-Based Computer-Assisted Radiation Therapy*. Singapore: Springer Singapore; 2017:9–40. https://doi.org/10.1007/978-981-10-2945-5_2

16. Mahersia H, Zaroug M, Gabralla L. Lung Cancer Detection on CT Scan Images: A Review on the Analysis Techniques. *Int J Adv Res Artif Intell*. 2015; 4(4):38–45. www.ijarai.thesai.org.

17. Way T, Chan HP, Hadjiiski L, Sahiner B, Chughtai A, Song TK, et al. Computer-Aided Diagnosis of Lung Nodules on CT Scans: ROC Study of Its Effect on Radiologists' Performance. *Acad Radiol*. 2010; 17(3):323–332. https://doi.org/10.1016/j.acra.2009.10.016 PMID: 20152726

18. Arai K, Herdiyeni Y, Okumura H. Comparison of 2D and 3D Local Binary Pattern in Lung Cancer Diagnosis. *Int J Adv Comput Sci Appl*. 2012; 3(4):89–95. https://doi.org/10.14569/IJACSA.2012.030416

19. Chen H, Xu Y, Ma Y, Ma B. Neural Network Ensemble-Based Computer-Aided Diagnosis for Differentiation of Lung Nodules on CT Images. *Acad Radiol*. 2010; 17(5):595–602. https://doi.org/10.1016/j.acra.2009.12.009 PMID: 20167513

**20.** Armato SG 3rd, Hadjiiski L, Tourassi GD, Drukker K, Giger ML, Li F, et al. LUNGx Challenge for computerized lung nodule classification: reflections and lessons learned. *J Med imaging (Bellingham, Wash)*. 2015; 2(2):20103. https://doi.org/10.1117/1.JMI.2.2.020103 PMID: 26158094

**21.** Armato SG 3rd, Drukker K, Li F, Hadjiiski L, Tourassi GD, Engelmann RM, et al. LUNGx Challenge for computerized lung nodule classification. *J Med Imaging*. 2016; 3(4):44506. https://doi.org/10.1117/1.JMI.3.4.044506 PMID: 28018939

**22.** Ciompi F, Chung K, van Riel SJ, Setio AAA, Gerke PK, Jacobs C, et al. Towards automatic pulmonary nodule management in lung cancer screening with deep learning. *Sci Rep*. 2017; 7:46479. https://doi.org/10.1038/srep46479 PMID: 28422152

**23.** Liang M, Tang W, Xu DM, Jirapatnakul AC, Reeves AP, Henschke CI, et al. Low-Dose CT Screening for Lung Cancer: Computer-aided Detection of Missed Lung Cancers. *Radiology*. 2016; 281(1):279–288. https://doi.org/10.1148/radiol.2016150063 PMID: 27019363

**24.** Chang C-C, Lin C-J. LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol*. 2011; 2(3):1–27. https://doi.org/10.1145/1961189.1961199

**25.** Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. *Proc 22nd ACM SIGKDD Int Conf Knowl Discov Data Min—KDD '16*. 2016:785–794.

**26.** Hsu CW, Chang CC, Lin CJ. A practical guide to support vector classification. Available from http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf. 2003.

**27.** Bergstra J, Bardenet R, Bengio Y, Kégl B. Algorithms for Hyper-Parameter Optimization. 25th Annual Conference on Neural Information Processing Systems 2011, December 12–15, 2011, Granada, Spain. Neural Information Processing Systems; 2012. http://dl.acm.org/citation.cfm?id=2986743.

**28.** Bergstra J, Bengio Y. Random Search for Hyper-Parameter Optimization. *J Mach Learn Res*. 2012; 13 (Feb):281–305.

**29.** Ojala T, Pietikainen M, Harwood D. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognit*. 1996; 29(1):51–59. https://doi.org/10.1016/0031-3203(95)00067-4

**30.** Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans Pattern Anal Mach Intell*. 2002; 24(7):971–987. https://doi.org/10.1109/TPAMI.2002.1017623

**31.** Zhao G, Pietikainen M. Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions. *IEEE Trans Pattern Anal Mach Intell*. 2007; 29(6):915–928. https://doi.org/10.1109/TPAMI.2007.1110 PMID: 17431293

**32.** Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, et al. The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. *J Digit Imaging*. 2013; 26(6):1045–1057. https://doi.org/10.1007/s10278-013-9622-7 PMID: 23884657

**33.** Gevaert O, Xu J, Hoang CD, Leung AN, Xu Y, Quon A, et al. Non–Small Cell Lung Cancer: Identifying Prognostic Imaging Biomarkers by Leveraging Public Gene Expression Microarray Data—Methods and Preliminary Results. *Radiology*. 2012; 264(2):387–396. https://doi.org/10.1148/radiol.12111607 PMID: 22723499

**34.** Armato SG III, Hadjiiski L, Tourassi GD, Drukker K, Giger ML, Li F, et al. SPIE-AAPM-NCI Lung Nodule Classification Challenge Dataset. The Cancer Imaging Archive. http://doi.org/10.7937/K9/TCIA.2015.UZLSU3FL. Published 2015. Accessed June 27, 2017.

**35.** Napel Sandy, & Plevritis Sylvia K. NSCLC Radiogenomics: Initial Stanford Study of 26 Cases. The Cancer Imaging Archive. http://doi.org/10.7937/K9/TCIA.2014.X7ONY6B1. Published 2014. Accessed June 27, 2017.

**36.** XGBoost Parameters https://github.com/dmlc/xgboost/blob/master/doc/parameter.md. Accessed February 8, 2018.

**37.** Liu Y, Gu Y, Nguyen JC, Li H, Zhang J, Gao Y, et al. Symptom severity classification with gradient tree boosting. *J Biomed Inform*. 2017. https://doi.org/10.1016/j.jbi.2017.05.015 PMID: 28545836

**38.** Torlay L, Perrone-Bertolotti M, Thomas E, Baciu M. Machine learning—XGBoost analysis of language networks to classify patients with epilepsy. *Brain Informatics*. April 2017:1–11. https://doi.org/10.1007/s40708-017-0065-7 PMID: 28434153

**39.** Wei B, Peng Q, Chen X, Zhao J. Bayesian optimization algorithm-based methods searching for risk/protective factors. *Chinese Sci Bull*. 2013; 58(23):2828–2835. https://doi.org/10.1007/s11434-012-5475-6

**40.** Lorenz R, Monti RP, Violante IR, Faisal AA, Anagnostopoulos C, Leech R, et al. Stopping criteria for boosting automatic experimental design using real-time fMRI with Bayesian optimization. November 2015. http://arxiv.org/abs/1511.07827. Accessed June 29, 2017.

**41.** Lorenz R, Monti RP, Hampshire A, Koush Y, Anagnostopoulos C, Faisal AA, et al. Towards tailoring non-invasive brain stimulation using real-time fMRI and Bayesian optimization. In: *2016 International*

Workshop on Pattern Recognition in Neuroimaging (PRNI). IEEE; 2016:1–4. https://doi.org/10.1109/PRNI.2016.7552338

42. Schubert T, Eggensperger K, Gkogkidis A, Hutter F, Ball T, Burgard W. Automatic bone parameter estimation for skeleton tracking in optical motion capture. In: 2016 IEEE International Conference on Robotics and Automation (ICRA). IEEE; 2016:5548–5554. https://doi.org/10.1109/ICRA.2016.7487771

43. Smith LN. Best Practices for Applying Deep Learning to Novel Applications. arXiv:1704.01568