Japanese Journal of Radiology (2019) 37:264–273
https://doi.org/10.1007/s11604-018-0784-6

TECHNICAL NOTE

Can the spherical gold standards be used as an alternative to painted gold standards for the computerized detection of lesions using voxel-based classification?

Yukihiro Nomura[1] · **Naoto Hayashi**[1] · **Shouhei Hanaoka**[2] · **Tomomi Takenaga**[1] · **Mitsutaka Nemoto**[3] · **Soichiro Miki**[1] · **Takeharu Yoshikawa**[1] · **Osamu Abe**[2]

Received: 11 June 2018 / Accepted: 8 October 2018 / Published online: 20 October 2018

Abstract

Purpose For the development of computer-assisted detection (CAD) software using voxel-based classification, gold standards defined by pixel-by-pixel painting, called painted gold standards, are desirable. However, for radiologists who define gold standards, a simplified method of definition is desirable. One of the simplest methods of defining gold standards is a spherical region, called a spherical gold standard. In this study, we investigated whether spherical gold standards can be used as an alternative to painted gold standards for computerized detection using voxel-based classification.

Materials and methods The spherical gold standards were determined by the center of gravity and the maximum diameter. We compared two types of gold standard, painted gold standards and spherical gold standards, by two types of CAD software using voxel-based classification.

Results The time required to paint the area of one lesion was 4.7–6.5 times longer than the time required to define a spherical gold standard. For the same performance of the CAD software, the number of training cases required for the spherical gold standard was 1.6–7.6 times that for the painted gold standards.

Conclusion Spherical gold standards can be used as an alternative to painted gold standards for the computerized detection of lesions with simple shapes.

Keywords Computer-assisted detection · Gold standard · Voxel-based classification

Introduction

In the development of computer-assisted detection (CAD) software, a sufficiently large image dataset for supervised learning is indispensable. In the dataset used for the computerized detection of lesions, which indicates the locations of lesion candidates, the areas of lesions must be clearly defined as gold standards either by pixel-by-pixel painting, by setting a region of interest (ROI), or by inputting the location of the lesion center [1]. Recently, several research groups have reported lesion candidate extraction algorithms using voxel-based classification [2–7]. For these methods, it is desirable to define gold standards by pixel-by-pixel painting (called painted gold standards hereafter).

However, the definition of a painted gold standard requires a significant effort. Painting of the areas of lesions is a tough task for radiologists because it is time consuming and requires concentration. Hence, it is practically difficult for radiologists to be assigned this task for long periods of time.

To define the gold standard for more cases in a limited time, it is desirable to reduce the effort of gold standard definition by employing a simplified method. One of the simplest methods of defining gold standards is a spherical region, called a spherical gold standard. A spherical gold standard is determined by the location of a lesion center and the measured size. Although supervised learning with a

✉ Yukihiro Nomura
nomuray-tky@umin.ac.jp

[1] Department of Computational Diagnostic Radiology and Preventive Medicine, The University of Tokyo Hospital, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan

[2] Department of Radiology, The University of Tokyo Hospital, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan

[3] Faculty of Biology-Oriented Science and Technology, Kindai University, Nishimitani 930, Kinokawa, Wakayama 649-6493, Japan

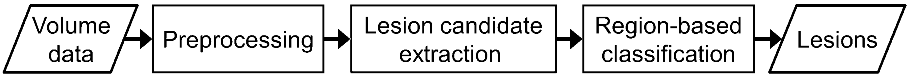

Fig. 1 Flowchart of our lesion detection algorithm

spherical gold standard requires a few more cases for training, if performance equivalent to that of supervised learning with painted gold standards can be obtained, spherical gold standards can be used as an alternative to painted gold standards.

In this study, we investigated whether spherical gold standards can be used as an alternative to painted gold standards for computerized detection using voxel-based classification.

Methods

CAD software

We targeted two types of CAD software for cerebral aneurysm detection in magnetic resonance angiography (MRA) images and lung nodule detection in chest computed tomography (CT) images. The basic detection algorithm used in this study consists of three steps: preprocessing, lesion candidate extraction using voxel-based classification, and region-based classification (Fig. 1). Details of the algorithm are described in "Appendix".

Datasets

We utilized a brain MRA dataset for cerebral aneurysm detection and a chest CT dataset for lung nodule detection collected from our institution. This study was approved by the ethical review board of our institution, and written informed consent to use the images for the study was obtained from all the subjects. Additionally, we utilized the publicly available Lung Image Database Consortium—Image Database Resource Initiative (LIDC–IDRI) dataset [8] as chest CT dataset B.

1. Brain MRA dataset

Four-hundred and fifty cases of 3D time-of-flight unenhanced MRA images with 573 cerebral aneurysms were used. These images were scanned at our institution with three 3-Tesla MR scanners (two Signa HDxt and one Discovery MR750, GE Healthcare, Waukesha, WI, USA). The acquisition parameters were as follows: field of view, 240 mm; matrix size, 512 × 512 pixels; pixel spacing, 0.469 mm; slice thickness, 1.2 mm; slice interval, 0.6 mm; repetition time, 22 or 25 ms; echo time, 2.7–3.3 ms; flip angle, 15°. Each case included at least one aneurysm of 2 mm or more in diameter, which was determined by consensual reading by two experienced radiologists. The diameter was manually measured by the radiologists at the initial reading. For each aneurysm, a board-certified radiologist (N.H., 29 years of experience in MRA interpretation) defined its area by pixel-by-pixel painting.

2. Chest CT dataset A

Four hundred cases of chest CT images with 505 lung nodules were used. These images were scanned at our institution with a GE LightSpeed CT scanner (GE Healthcare). The acquisition parameters were as follows: number of detector rows, 16; tube voltage, 120 kVp; tube current, 50–290 mA (automatic exposure control); noise index, 20.41; rotation time, 0.5 s; moving table speed, 70 mm/s; body filter, standard; reconstruction slice thickness and interval, 1.25 mm; field of view, 400 mm; matrix size, 512 × 512 pixels; pixel spacing, 0.781 mm. Each case included at least one nodule of 5 mm or more in diameter, which was determined by consensual reading by two experienced radiologists. The lung nodules of less than 5 mm were excluded from this study. The diameter was manually measured by the radiologists at the initial reading. For each nodule, a board-certified radiologist (N.H., 29 years of experience in chest CT interpretation; S.M., 11 years of experience in chest CT interpretation) defined its area by pixel-by-pixel painting.

3. Chest CT dataset B

We selected 419 chest CT volumes from the LIDC–IDRI database that satisfied two conditions: (1) the slice thickness and slice interval were less than or equal to 2.0 mm, (2) the CT scan contained at least one nodule of diameter 3–30 mm determined by at least three radiologists. In this dataset, there were 823 nodules, and three or four radiologists delineated the contour of each nodule. In the study, the painted gold standards were defined from contours provided by at least two radiologists.

Types of gold standard

We utilized two types of gold standard, painted gold standards and spherical gold standards (Fig. 2). Painted gold standards were defined in each dataset as mentioned above. The spherical gold standards were automatically determined from painted gold standards instead of manually input by a radiologist. Thus, the spherical gold standards were determined by the center of gravity and the maximum diameter,

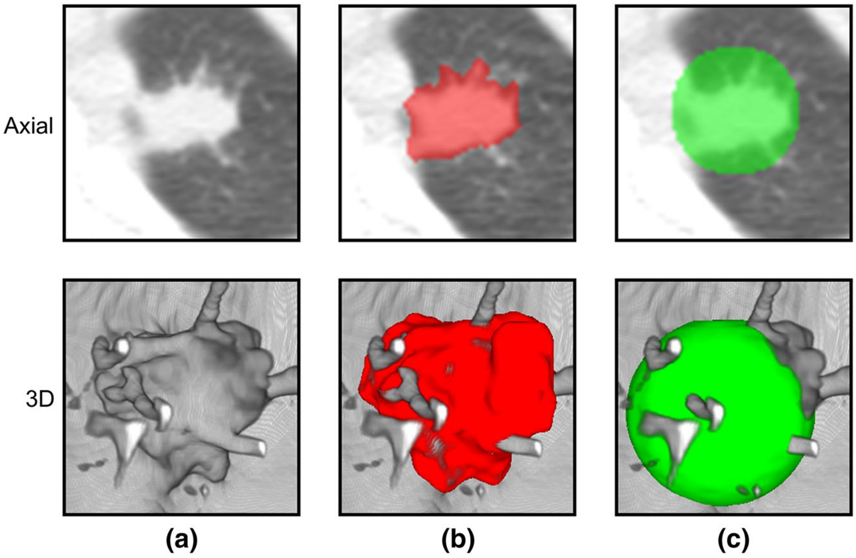

Fig. 2 Examples of gold standards: **a** original image, **b** painted gold standard (red), **c** spherical gold standard (green). The upper images show the axial section and the lower images show the 3D volume rendering

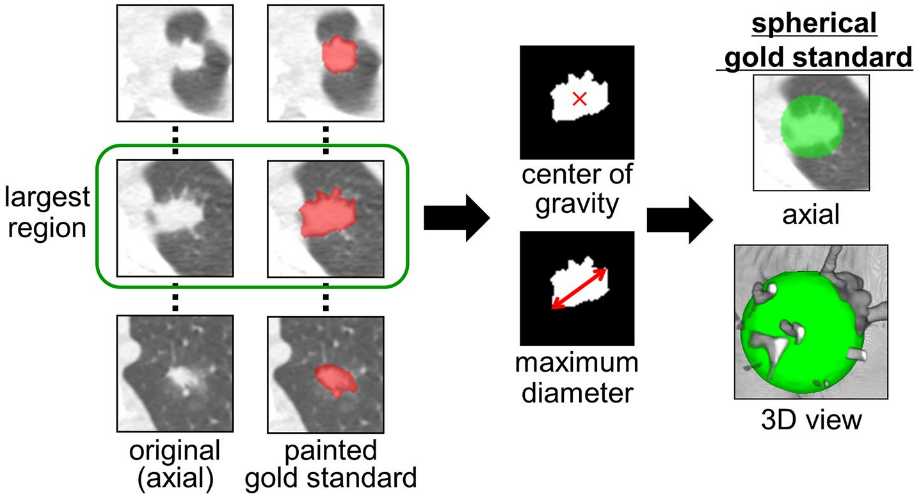

Fig. 3 Definition of spherical gold standard. The spherical gold standard was determined by the center of gravity and the maximum diameter, which were measured on the largest painted gold standard in the axial section

which were measured on the largest painted gold standard in the axial section (Fig. 3).

Evaluation of radiologists' workload

We evaluated the workload of radiologists to define the gold standards. Two radiologists, namely Radiologist A (N.H., 29 years of experience in brain MRA and chest CT interpretation) and Radiologist B (T.Y., 23 years of experience in brain MRA and chest CT interpretation), defined gold standards using a web-based image database system for CAD development developed by our group (not published). In the database system, a volume-based viewing and painting component were implemented to define gold standards on multi-planar reconstruction (MPR) images. The two radiologists were proficient in the use of the system. We utilized 10 cases for cerebral aneurysm detection (Fig. 4) and 10 cases for lung nodule detection (Fig. 5). Each case included one lesion. We measured the time required to define gold standards by the following procedure:

1. (for spherical gold standards): the time required to obtain the center of gravity and the maximum diameter of the largest region in the axial section
2. (for painted gold standards): the time required to paint the gold standard.

Evaluation of CAD performance

We evaluated the performance of CAD software using two types of gold standard through a simulation-based study. This simulation was conducted using a supercomputer system in our institute, which consists of 420 computing nodes equipped with two Intel Xeon E5-2695v4 processors and a

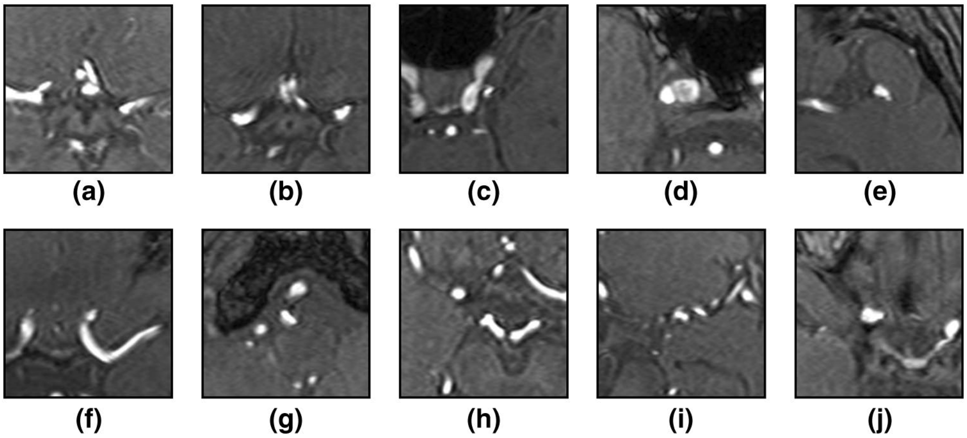

Fig. 4 Cerebral aneurysms used for evaluating workload of radiologists to define the gold standards: aneurisms of **a** 3 mm in anterior communicating artery, **b** 3 mm in right anterior cerebral artery, **c** 4 mm in left internal carotid artery (ICA), **d** 9 mm in right ICA, **e** 3 mm in left middle cerebral artery (MCA), **f** 3 mm in left ICA, **g** 4 mm in vertebral artery, **h** 2 mm in right posterior cerebral artery, **i** 2 mm in left MCA, **j** 2 mm in right ICA

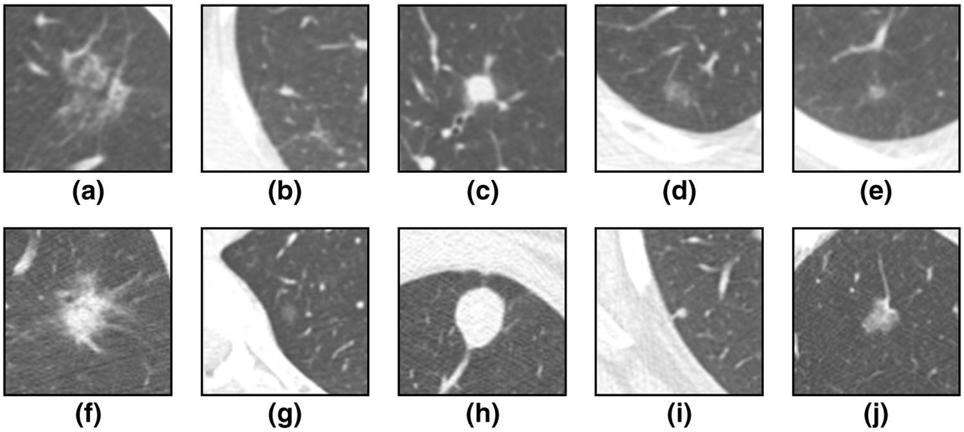

Fig. 5 Lung nodules used for evaluating workload of radiologists to define the gold standards: **a** 24 mm solid nodule, **b** 5 mm solid nodule, **c** 10 mm solid nodule, **d** 9 mm pure ground-glass nodule (pGGN), **e** 5 mm pGGN, **f** 27 mm solid nodule, **g** 8 mm pGGN, **h** 23 mm solid nodule, **i** 4 mm solid nodule, **j** 14 mm pGGN. The cases on the top row was selected from chest CT dataset A and those on the bottom row was selected from chest CT dataset B

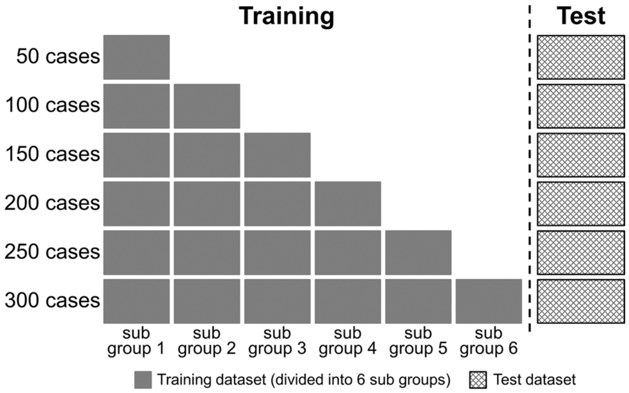

Fig. 6 Relationship between training data and test data for simulation

memory of 256 GB. The dataset was randomly divided with 300 cases used as training data and the remaining cases used as test data. The training data were divided into six subgroups (50 cases per sub group). Training and evaluation were performed by adding each sub-group of training data (Fig. 6). This procedure was repeated 50 times to reduce the sampling effect. In the performance evaluation, if the lesion candidate met the following condition, it was judged as a true positive (TP).

- Cerebral aneurysm: the distance between the center of gravity of the lesion candidate and the center of gravity of any aneurysm was less than 3.0 mm.
- Lung nodule: the center of gravity of the lesion candidate was contained within the gold standard.

The ANODE score [9], which defines the average sensitivity at predefined false-positive (FP) rates (1/8, 1/4, 1/2, 1, 2, 4, and 8 FPs/case) along a free-response receiver operating characteristic (FROC) curve, was employed as the evaluation criterion. To evaluate the differences in the gold standards, the TP ratio (TPR), the false-negative ratio (FNR), the FP ratio (FPR), and the dice coefficient (DC) were calculated as follows:

$$\mathrm{TPR} = \frac{\mathbf{P} \cap \mathbf{S}}{\mathbf{P}}, \quad (1)$$

$$\mathrm{FNR} = \frac{\mathbf{P} - \mathbf{P} \cap \mathbf{S}}{\mathbf{P}}, \quad (2)$$

Table 1 Time required to define each type of gold standard

Type of lesion	Type of gold standard	Radiologist A	Radiologist B	Total
Cerebral aneurysm	Painted	219.2 ± 125.2	201.7 ± 119.8	210.5 ± 119.6
	Spherical	59.8 ± 12.3	29.6 ± 2.8	44.7 ± 17.8
Lung nodule	Painted	378.0 ± 246.1	428.6 ± 317.9	403.3 ± 277.9
	Spherical	87.4 ± 35.5	37.5 ± 7.3	62.5 ± 35.7

Data are mean ± standard deviation (s)

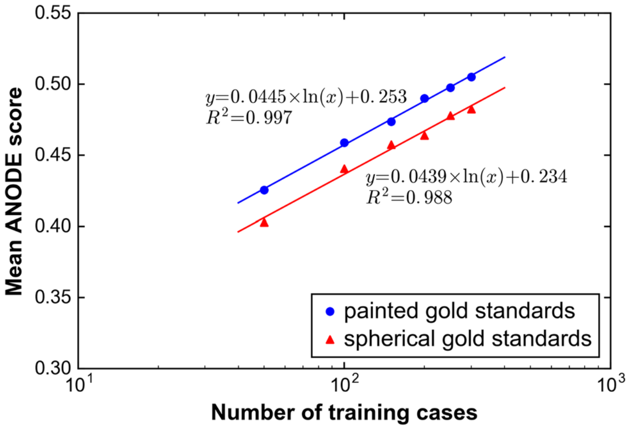

Fig. 7 Scatter plot showing relationship between number of training cases and performance of CAD software in cerebral aneurysm detection using brain MRA dataset. The lines were obtained by logarithmic fitting

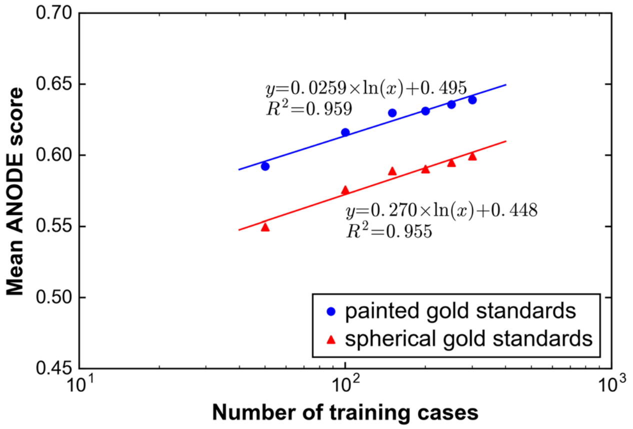

Fig. 8 Scatter plot showing relationship between number of training cases and performance of CAD software in lung nodule detection using chest CT dataset A. The lines were obtained by logarithmic fitting

$$\mathrm{FPR} = \frac{\mathbf{S} - \mathbf{P} \cap \mathbf{S}}{\mathbf{P}}, \tag{3}$$

$$\mathrm{DC} = \frac{2(\mathbf{P} \cap \mathbf{S})}{\mathbf{P} + \mathbf{S}}, \tag{4}$$

where **P** is the region of the painted gold standard and **S** is the region of the spherical gold standard. Larger values of TPR and DC and smaller values of FNR and FPR correspond to greater accuracy. The ratios were calculated within the voxels to be processed with the voxel-based classifier in the lesion candidate extraction algorithm.

Results

Table 1 shows the time required to define the gold standard. For the cerebral aneurysms, it took 210.5 ± 119.6 s to define the painted gold standard and 44.7 ± 17.8 s to define the spherical gold standard. For the lung nodules, it took 403.3 ± 277.9 s to define the painted gold standard and 62.5 ± 35.7 s to define the spherical gold standard.

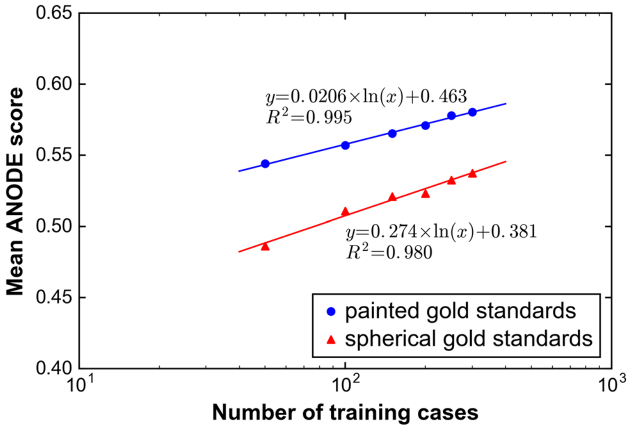

Fig. 9 Scatter plot showing relationship between number of training cases and performance of CAD software in lung nodule detection using chest CT dataset B. The lines were obtained by logarithmic fitting

Figure 7 shows scatter plots showing the relationship between the number of training cases and the performance of CAD software in cerebral aneurysm detection using the brain MRA dataset. Figures 8 and 9 show scatter plots showing the relationship between the number of training cases

Table 2 Accuracy of spherical gold standards for each type of dataset

Ratio	Brain MRA dataset	Chest CT dataset	
		Dataset A	Dataset B
TPR	0.827 ± 0.190	0.855 ± 0.119	0.909 ± 0.091
FNR	0.173 ± 0.190	0.145 ± 0.119	0.091 ± 0.091
FPR	0.268 ± 0.423	0.483 ± 0.493	0.478 ± 0.752
DC	0.795 ± 0.152	0.748 ± 0.113	0.789 ± 0.117

Data are mean ± standard deviation

TPR true-positive ratio, *FNR* false-negative ratio, *FPR* false-positive ratio, *DC* dice coefficient

and the performance of CAD software in lung nodule detection using chest CT datasets A and B, respectively. The mean ANODE score was improved by increasing the number of training cases regardless of the type of gold standard, dataset, and target lesion. From the results of logarithmic fitting, in the cerebral aneurysm detection, the performance of the painted gold standards with 50 training cases was almost equivalent to that of the spherical gold standards with 81 cases. In the lung nodule detection, the performance of the painted gold standards with 50 training cases was almost equivalent to that of the spherical gold standards with 243 cases (chest CT dataset A), and 378 cases (chest CT dataset B).

Table 2 shows the accuracy of the spherical gold standards for each dataset. TPR and DC were higher than 0.82 and 0.74, respectively, regardless of the dataset.

Discussion

We experimentally showed that the time required to paint the area of one lesion was about 4.7–6.5 times longer than the time required to define the spherical gold standard. The time required to paint the area of one lesion depends on the type and size of the lesion. In contrast, a spherical gold standard can be defined simply by inputting the center location of the lesion and measuring its diameter in the axial section, considerably reducing the effort required.

Regarding the performance of the computerized detection of lesions using voxel-based classification, the performance of the painted gold standards was superior to that of the spherical gold standard. In the cerebral aneurysm detection, the number of training cases required for the spherical gold standard was 1.6 times more than that required for the painted gold standards. This is because the spherical gold standards had similar shapes to the painted gold standards since cerebral aneurysms are almost spherical. By contrast, in the lung nodule detection, the number of training cases required for the spherical gold standard was 4.9–7.6 times more than that required for in the painted gold standards. This is because the shapes of lung nodules are more complicated than those of cerebral aneurysms, and the FP voxels were mainly found to be adjacent to anatomical structures such as vessels or bronchi.

As described above, the total workload to define the spherical gold standards is less than or equal to that to define the painted gold standards. Consequently, spherical gold standards can be used as an alternative to painted gold standards if the lesion shape is not complicated, such as for a cerebral aneurysm, or a large number of positive cases can be used for training.

The center and diameter of the spherical gold standards were, respectively, set as the center of gravity and the maximum diameter, which were measured on the largest painted gold standard in the axial section and are easy to measure by radiologists in a clinical environment. Using other size measurements, such as the short diameter or the average of short-axis and long-axis measurements, the performance of spherical gold standards may be improved by reducing the number of FP voxels. However, the effort required to input the gold standards is increased.

Semi-automatic segmentation algorithms [10–12] can also be used to define gold standards to reduce the efforts of radiologists. However, many of the algorithms are dedicated to a specific type of lesion. It is necessary to develop a semi-automatic segmentation algorithm for various types of lesions.

The limitations of this study are discussed as follows. First, we evaluated only one algorithm. The results may not be applicable to other algorithms or datasets. Further evaluation using other algorithms, which include lesion candidate extraction algorithms based on deep learning, and datasets is necessary. Second, the spherical gold standards were automatically determined from painted gold standards instead of manually inputting. Manually inputting by radiologists may cause input error affecting the performance of CAD. Third, we did not perform a non-inferior test or equivalence test for the purpose of this study. Fourth, we did not consider a cost of collecting a large number of image dataset.

Conclusion

We have investigated whether spherical gold standards can be used as an alternative to painted gold standards for computerized detection using voxel-based classification. According to our results, spherical gold standards can be used as an alternative to painted gold standards for the computerized detection of lesions with simple shapes.

Acknowledgements The Department of Computational Radiology and Preventive Medicine, The University of Tokyo Hospital, is sponsored by HIMEDIC Inc. and Siemens Healthcare K.K. This work was supported by Japan Society for the Promotion of Science (JSPS) KAKENHI Grant numbers 15K01325 and 18K12096.

Compliance with ethical standards

Conflict of interest The authors declare no conflicts of interest with regard to the present study.

Ethical statements This study was approved by the ethical review board of our institution, and written informed consent to use the images for the study was obtained from all the subjects.

Appendix: Algorithm of CAD software

Basic detection algorithm

In the preprocessing, input images are first resampled using trilinear interpolation to obtain the isotropic volume. The resampled voxel size is equal to the pixel size of the input image. After that, segmentation is carried out.

In the lesion candidate extraction, a classifier ensemble trained by AdaBoost [13] is employed to classify the voxels of the target region. All the weak classifiers within the classifier ensembles are decision stumps [14]. The 21 feature values used in the voxel-based classification are as follows:

- voxel value
- statistics of voxel value: four types (mean, standard deviation, skewness, and kurtosis) × three cubic region of interest (ROI) sizes (length = {3, 5, 9} [voxels])
- difference of Gaussian (DoG) (three pairs of σ, {1, 2}, {2, 4}, {4, 8} [voxels])
- shape index [15] ($\sigma = \{2, 4\}$ [voxels])
- multiscale dot and line enhancement filter [16]
- multiscale vessel enhancement filter [17] (the parameters α, β, and c are set to 0.5, 0.5, and 500, respectively).

The number of weak classifiers is set to 200 in the experiments. In the training of the classifiers, random under sampling is carried out to reduce the number of normal samples relative to the number of lesion samples. The voxels for which the output of the classifier is greater than or equal to 0 are extracted, and component analysis is performed. After that, small-region removal is performed.

In the region-based classification, a classifier ensemble trained by AdaBoost is employed to calculate the likelihoods of the lesion candidates based on 70 feature values of the candidates. All the weak classifiers within the classifier ensembles are decision stumps. The number of weak classifiers is set to 200 in the experiments. The feature values are as follows:

- volume (V) (mm^3) and surface area (mm^2)
- two surface features (surface exposure ratio R_{SE} and surface area to volume ratio R_{SV} in [18])
- contrast measures (Contrast1 and Contrast2 in [19])
- sphericity, ratio of V to the volume in the bounding box, and similarity to sphere [3]
- statistics of the distance between the center of the candidate and its boundary (minimum, maximum, mean, second moment, standard deviation, skewness, kurtosis, and minimum/maximum)
- statistics of voxel value, DoG, shape index, multiscale dot enhancement filter, multiscale line enhancement filter, and multiscale vessel enhancement filter[1] (minimum, maximum, mean, standard deviation, skewness, and kurtosis).

Cerebral aneurysm detection

In the preprocessing, the artery region is extracted by the algorithm in [20]. In the lesion candidate extraction, the voxels with shape index ($\sigma = 2.0$ [voxel]) ≥ 0.75 in the artery region are input to the voxel-based classifier. The range of σ for calculating the multiscale dot, line, or vessel enhancement filter is set to $\{1, 2^{1/2}, 2, 2^{3/2}, 4\}$ voxels. The threshold for the small-region removal is set to 1 (mm^3) in the experiments. In the region-based classification, the surface region for calculating R_{SE} is extracted from the extracted artery region.

Lung nodule detection

In the preprocessing, the lung volume **L** is obtained by the algorithm written in the later subsection. Using the extracted lung volume **L**, the binarized lung volume $\mathbf{L}_{\mathrm{bin}}$ is extracted as follows:

$$\mathbf{L}_{\mathrm{bin}} = \{p | I(\mathbf{x}) \geq -750, p \in \mathbf{L}\}, \tag{5}$$

where $I(\mathbf{x})$ is the CT value (HU) of voxel p and $\mathbf{x}$ is the three-dimensional (3D) coordinates of voxel p. In the lesion candidate extraction, the voxels in $\mathbf{L}_{\mathrm{bin}}$ are input to the voxel-based classifier. The range of σ for calculating the multiscale dot, line, or vessel enhancement filter is set to {1, 2, 4, 8} voxels. The threshold for the small-region removal is set to 8 (mm^3) in the experiments. In the region-based classification, the surface region for calculating R_{SE} is extracted from $\mathbf{L}_{\mathrm{bin}}$.

---

[1] The parameters of the DoG, shape index, dot enhancement filter, line enhancement filter, and vessel enhancement filter are the same as those in the voxel-based classification.

Algorithm of lung volume segmentation

The processing procedure of lung segmentation is described as follows.

(1) Extraction of the body trunk region $\mathbf{R}_{\text{body}}$.

(1-1) The CT volume data are resampled to 2.0 mm isotropic voxels.

(1-2) In each axial slice, the largest binarized area with a CT value of − 150 HU or higher is selected. After that, two-dimensional (2D) cavity deletion [21] is carried out.

(1-3) Morphological closing with a spherical kernel of 10 mm radius is carried out.

(1-4) After resampling to the original size, the result is defined as $\mathbf{R}_{\text{body}}$.

(2) Extraction of the bone region $\mathbf{R}_{\text{bone}}$.

(2-1) The CT volume data are resampled to 2.0 mm isotropic voxels.

(2-2) In $\mathbf{R}_{\text{body}}$, the largest binarized area with a CT value of 200 HU or higher is extracted. After that, morphological closing with a spherical kernel of 5 mm radius is carried out.

(2-3) After resampling to the original size, morphological dilation with a spherical kernel of 1.5 voxel radius is carried out. The result is defined as $\mathbf{R}_{\text{bone}}$.

(3) Extraction of thoracic slice range.

(3-1) The CT volume data are resampled to 2.0 mm isotropic voxels.

(3-2) In each axial slice of $\mathbf{R}_{\text{body}}$, the likelihood of thoracic slice $L_{\text{th}}(z)$ is calculated as follows:

$$L_{\text{th}}(z) = \frac{M_{\text{B}}(z)}{\overline{M_{\text{B}}}} R_{\text{air}}(z), \tag{6}$$

(3-3) where $M_{\text{B}}(z)$ is the second moment of $\mathbf{R}_{\text{bone}}$ in axial slice z, $\overline{M_{\text{B}}}$ is the mean of $M_{\text{B}}(z)$, and $R_{\text{air}}(z)$ is the ratio of the number of voxels with CT value ≤ -500 HU to the total number of voxels of $\mathbf{R}_{\text{body}}$ in the axial slice.

(3-4) The slice range above the mean of $L_{\text{th}}(z)$ including the maximum of $L_{\text{th}}(z)$ is defined as the thoracic slice range $\mathbf{Z}_{\text{th}} = [z_{\text{th\_min}}, z_{\text{th\_max}}]$.

(4) Estimation of border between left and right sides of body trunk.

(4-1) The CT volume data are resampled to 2.0 mm isotropic voxels.

(4-2) In each axial slice from the center quarter region of $\mathbf{Z}_{\text{th}}$, partial maximum intensity projection images of $\mathbf{R}_{\text{body}}$ and $\mathbf{R}_{\text{bone}}$, named $\mathbf{M}_{\text{body}}(z)$ and $\mathbf{M}_{\text{bone}}(z)$, respectively, are generated from five axial slices centered at z.

(4-3) The center of gravity of $\mathbf{M}_{\text{body}}(z)$ $(g_x(z), g_y(z))$ is calculated. After that, the number of voxels for which $\mathbf{M}_{\text{bone}}(z) = 1$ is counted every three degrees around $\mathbf{M}_{\text{body}}(z)$, and the maximum direction is defined as $\deg(z)$.

(4-4) The medians of $g_x(z)$ and $g_y(z)$ converted to the coordinates of the original CT volume are defined as G_x and G_y, respectively. The median of $\deg(z)$ is defined as D. The border between the left and right sides of the body trunk is set using these values.

(5) Extraction of pulmonary initial lung mask $\mathbf{R}_{\text{init}}$.

(5-1) In $\mathbf{R}_{\text{body}}$ within the thoracic slice range $\mathbf{Z}_{\text{th}}$, an initial threshold Th_{init} is calculated using Otsu's method [22]. After that, the binarized area with a CT value of Th_{init} or lower is extracted. These processes are performed on the left and right sides of body trunk, which are divided at the border determined in (4-4).

(5-2) In each axial slice from the range of $\mathbf{Z}_{\text{th}}$, connected component analysis is performed, and then up to two components with an area of 2500 mm^2 or more are extracted as seed areas.

(5-3) Region growing is carried out in $\mathbf{R}_{\text{body}}$, where the growing threshold is set to Th_{init}.

(5-4) Morphological dilation with a spherical kernel of 1.9 voxel radius is carried out. The result is defined as $\mathbf{R}_{\text{init}}$.

(6) Extraction of the bronchi region $\mathbf{R}_{\text{bronchi}}$.

(6-1) A gray-scale histogram of the voxels in $\mathbf{R}_{\text{init}}$ is generated, and the peak CT value of the histogram between the minimum CT value in $\mathbf{R}_{\text{body}}$ and − 650 HU, named T_{peak}, is obtained.

(6-2) In $\mathbf{R}_{\text{init}}$ within 10 cm of the upper end of the range of $\mathbf{Z}_{\text{th}}$, the binarized area with a CT value of Th_{peak} or lower is extracted, and then morphological opening with a spherical kernel of 2.5 mm radius and connected component analysis are carried out. Among the obtained connected components, the component with the shortest distance between the center of gravity of the component and (G_x, G_y) within 10 cm is extracted

(6-3) Morphological opening with a spherical kernel of 2.5 mm radius and connected component analysis are carried out, and then the largest component is extracted.

(6-4) The region obtained in (6-3) is set as a seed region, and region growing is repeated at most eight times with the following extension condition:

$$T_{\mathrm{bronchi}}(n) = \begin{cases} \max\left(I_{\min}, \mu_{\mathrm{seed}} - \sigma_{\mathrm{seed}}\right) & \text{if } n = 1 \\ T_{\mathrm{bronchi}}(1) + \frac{64}{2^{(n-2)}} & \text{if } 2 \le n \le 8 \end{cases}, \tag{7}$$

where $I_{\min}$ is the minimum CT value in $\mathbf{R}_{\mathrm{body}}$. μ_{seed} and σ_{seed} are the mean and standard deviation of the CT values in the seed region, respectively. If the volume of the nth region growing is less than twice the volume of the initial region growing, the result of the nth region growing is utilized. However, if this condition is not satisfied after the eighth region growing, the result of the initial region growing is used.

(6-5) Morphological dilation with a spherical kernel of 2.5 mm radius is carried out. The result is defined as $\mathbf{R}_{\mathrm{bronchi}}$.

(6-6) The region obtained by removing $\mathbf{R}_{\mathrm{bronchi}}$ from $\mathbf{R}_{\mathrm{init}}$ is defined as $\mathbf{R}_{\mathrm{coarse}}$.

(7) Lung separation.

Connected component analysis is performed in $\mathbf{R}_{\mathrm{coarse}}$, and the two largest connected components are extracted. If the number of voxels of the largest component exceeds 90% of the number of voxels of $\mathbf{R}_{\mathrm{coarse}}$, it is considered that both lung regions are connected, and the following processing is performed.

(7-1) In each axial slice, the Canny filter [23] is applied to detect the edges. The parameters of hysteresis thresholding are set to $\mathrm{Th}_{\min} = 250$ and $\mathrm{Th}_{\max} = 2500$. After that, morphological closing with a spherical kernel of 2.5 mm radius is carried out. The result is defined as $E(x, y, z)$.

(7-2) In each axial slice of $\mathbf{R}_{\mathrm{coarse}}$, the largest binarized area is extracted. If the number of voxels of the extracted area exceeds 90% of the number of voxels of $\mathbf{R}_{\mathrm{coarse}}$ in the same axial slice, it is considered that both lung regions are connected, and separation voxels, named $\mathrm{Sep}(x, y, z)$, are set as follows:

(a) Both lungs are separated in the axial slice at $z-1$:

$$\mathrm{Sep}(x, y, z) = \begin{cases} 1 & \text{if } R_{\mathrm{coarse}}(x, y, z) = 1 \text{ and } R_{\mathrm{init}}(x, y, z-1) = 0 \text{ and } E(x, y, z) = 1 \\ 0 & \text{otherwise} \end{cases}, \tag{8}$$

where $R_{\mathrm{coarse}}(x, y, z)$ is the voxel of $\mathbf{R}_{\mathrm{coarse}}$ at (x, y, z), and $R_{\mathrm{init}}(x, y, z)$ is the voxel of $\mathbf{R}_{\mathrm{init}}$ at (x, y, z).

(b) Both lungs are connected in the axial slice at $z-1$:

$$\mathrm{Sep}(x, y, z) = \begin{cases} 1 & \text{if } R_{\mathrm{coarse}}(x, y, z) = 1 \text{ and } \mathrm{Sep}'^{(x,y,z-1)} = 0 \text{ and } E(x, y, z) = 1 \\ 0 & \text{otherwise} \end{cases}, \tag{9}$$

$$\mathrm{Sep}'(x, y, z) = \begin{cases} 1 & \text{if } \sum_{i=x-1}^{x+1} \sum_{j=y-1}^{y+1} \mathrm{Sep}(i, j, z) \ge 1 \\ 0 & \text{otherwise} \end{cases}. \tag{10}$$

(7-4) The voxels with $\mathrm{Sep}(x, y, z) = 1$ are removed from $\mathbf{R}_{\mathrm{coarse}}$.

(7-5) Connected component analysis is performed on the result of (7-3). If the regions of both lungs are not separated, morphological erosion with a spherical kernel of 1.9 voxel radius is repeated until both lungs are separated (up to five times). The result, called $\mathbf{R}_{\mathrm{separated}}$, is represented by ternarized volume data (0: background, 1: label of the right lung, 2: label of the left lung).

(7-6) Gray-scale morphological reconstruction by recursive dilation [24] with a spherical kernel of 1.9 voxel radius is applied to $\mathbf{R}_{\mathrm{separated}}$. $\mathbf{R}_{\mathrm{coarse}}$ is utilized as the mask of the reconstruction. The result is used in the next step.

(8) Resegmentation of left and right lungs.

(8-1) Closing with a spherical kernel of 10 mm radius is applied to the masks for the left and right lungs to include lung nodules and pulmonary vessels. After that, 3D cavity deletion [21] is carried out.

(8-2) After obtaining the union before and after the processing of (8-1), intersection voxels of $\mathbf{R}_{\mathrm{bronchi}}$ and $\mathbf{R}_{\mathrm{bone}}$ are removed. The result is defined as $\mathbf{L}$.

References

1. Masutani Y, Nemoto M, Nomura Y, Hayashi N. Clinical machine learning in action: CAD system design, development, tuning, and long-term experience. In: Suzuki K, editor. Machine learning in

computer-aided diagnosis: medical imaging intelligence and analysis. Hershey: IGI Global; 2012. p. 159–76.

2. Suzuki K. A supervised 'lesion-enhancement' filter by use of a massive-training artificial neural network (MTANN) in computer-aided diagnosis (CAD). Phys Med Biol. 2009;54:S31–45.
3. Nomura Y, Masutani Y, Miki S, et al. Performance improvement in computerized detection of cerebral aneurysms by retraining classifier using feedback data collected in routine reading environment. J Biomed Graph Comput. 2014;4(4):12–21.
4. Dou Q, Chen H, Yu L, et al. Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks. IEEE Trans Med Imaging. 2016;35:1182–95.
5. Debats OA, Meijs M, Litjens GJ, Huisman HJ. Automated multistructure atlas-assisted detection of lymph nodes using pelvic MR lymphography in prostate cancer patients. Med Phys. 2016;43:3132–42.
6. Hamidian S, Sahiner B, Petrick N, Pezeshk A. 3D convolutional neural network for automatic detection of lung nodules in chest CT. Proc SPIE (Med Imaging). 2017;10134:1013409.
7. Nemoto M, Hayashi N, Hanaoka S, Nomura Y, Miki S, Yoshikawa T. Feasibility study of a generalized framework for developing computer-aided detection systems-a new paradigm. J Digit Imaging. 2017;30:629–39.
8. Armato SG 3rd, McLennan G, Bidaut L, et al. The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. Med Phys. 2011;38:915–31.
9. van Ginneken B, Armato SG 3rd, de Hoop B, et al. Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: the ANODE09 study. Med Image Anal. 2010;14:707–22.
10. Kuhnigk JM, Dicken V, Bornemann L, et al. Morphological segmentation and partial volume analysis for volumetry of solid pulmonary lesions in thoracic CT scans. IEEE Trans Med Imaging. 2006;25(4):417–34.
11. Diciotti S, Picozzi G, Falchini M, Mascalchi M, Villari N, Valli G. 3-D segmentation algorithm of small lung nodules in spiral CT images. IEEE Trans Inf Technol Biomed. 2008;12(1):7–19.
12. Lassen BC, Jacobs C, Kuhnigk JM, van Ginneken B, van Rikxoort EM. Robust semi-automatic segmentation of pulmonary subsolid nodules in chest computed tomography scans. Phys Med Biol. 2015;60(3):1307–23.
13. Schapire RE, Freund Y, Bartlett P, Lee WS. Boosting the margin: a new explanation for the effectiveness of voting methods. Ann Stat. 1998;26:1651–86.
14. Iba W, Langley P. Induction of one-level decision trees. In: Proceedings of international conference on machine learning (ICML 1992). 1992. pp. 233–240.
15. Dorai C, Jain AK. COSMOS-a representation scheme for 3D free-form objects. IEEE Trans Pattern Anal Mach Intell. 1997;19:1115–30.
16. Li Q, Sone S, Doi K. Selective enhancement filters for nodules, vessels, and airway walls in two- and three-dimensional CT scans. Med Phys. 2003;30:2040–51.
17. Frangi AF, Niessen WJ, Vincken KL, Viergever MA. Multiscale vessel enhancement filtering. MICCAI'98 LNCS. 1998;1496:130–137.
18. Nomura Y, Nemoto M, Masutani Y, et al. Reduction of false positives at vessel bifurcations in computerized detection of lung nodules. J Biomed Graph Comput. 2014;4(3):36–46.
19. te Brake GM, Karssemeijer N, Hendriks JHCL. An automatic method to discriminate malignant masses from normal tissue in digital mammograms. Phys Med Biol. 2000;45:2843–57.
20. Hanaoka S, Nomura Y, Nemoto M, et al. HoTPiG: a novel geometrical feature for vessel morphometry and its application to cerebral aneurysm detection. MICCAI Part II LNCS. 2015;2015(9350):103–10.
21. Wan SY, Higgins WE. Symmetric region growing. IEEE Trans Image Process. 2003;12:1007–15.
22. Otsu N. A threshold selection method from gray-level histograms. IEEE Trans Syst Man Cybern. 1979;9:62–6.
23. Canny J. A computational approach to edge detection. IEEE Trans Pattern Anal Mach Intell. 1986;PAMI-8:679–98.
24. Gu L, Peters T. 3D segmentation of medical images using a fast multistage hybrid algorithm. Int J Comput Assist Radiol Surg. 2006;1:23–31.