Feasibility Study of a Generalized Framework for Developing Computer-Aided Detection Systems—a New Paradigm

Mitsutaka Nemoto¹ · Naoto Hayashi¹ · Shouhei Hanaoka² · Yukihiro Nomura¹ · Soichiro Miki¹ · Takeharu Yoshikawa¹

Published online: 12 April 2017
© Society for Imaging Informatics in Medicine 2017

Abstract We propose a generalized framework for developing computer-aided detection (CADe) systems whose characteristics depend only on those of the training dataset. The purpose of this study is to show the feasibility of the framework. Two different CAdE systems were experimentally developed by a prototype of the framework, but with different training datasets. The CAdE systems include four components; preprocessing, candidate area extraction, candidate detection, and candidate classification. Four pretrained algorithms with dedicated optimization/setting methods corresponding to the respective components were prepared in advance. The pretrained algorithms were sequentially trained in the order of processing of the components. In this study, two different datasets, brain MRA with cerebral aneurysms and chest CT with lung nodules, were collected to develop two different types of CAdE systems in the framework. The performances of the developed CAdE systems were evaluated by threefold cross-validation. The CAdE systems for detecting cerebral aneurysms in brain MRAs and for detecting lung nodules in chest CTs were successfully developed using the respective datasets. The framework was shown to be feasible by the successful development of the two different types of CAdE systems. The feasibility of this framework shows promise for a new paradigm in the development of CAdE

systems: development of CAdE systems without any lesion specific algorithm designing.

Keywords Computer-aided detection (CADe) system · Generalized CAdE framework · Automatic optimization · Machine learning method · CAdE training dataset

Background

Computer-aided detection (CADe) is a pattern recognition technology that identifies features of concern on a medical image and brings them to the attention of the radiologist [1]. Currently, the radiologist first reviews the images, then activates the CAdE system and re-evaluates the CAdE-marked areas of concern before issuing the final report. CAdE has been widely researched, and various types of CAdE system have been developed [2, 3]. Effective CAdE systems speed up the medical diagnostic process, reduce diagnostic errors, and improve quantitative evaluations [4].

CADe systems are difficult to develop because of their multiple requirements of specialized expertise in image processing, machine learning, medical physics, and radiology [4]. Moreover, most of the algorithms used in the components of CAdE systems must be manually designed by developers with these areas of expertise. Recently, some components of CAdE systems have been automatically optimized using machine learning methods [5, 6] and feature selection methods [7, 8]. Machine learning methods are computational methodologies by which computer algorithms can automatically learn complex relationships or patterns and devise accurate decision rules or models from training datasets [9–12]; machine learning is data-driven. For example, machine learning methods such as artificial neural networks [5] and support vector machines [6] are used to optimize the classifiers used for

✉ Mitsutaka Nemoto
mnemo-tky@umin.net

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

candidate classification in CAdE systems. Feature selection methods are data-driven optimization methods and are used to automatically select optimal features [13]. For example, various methods such as a stepwise selection method [7] and feature selection based on the genetic algorithm [8] are used to select the features for candidate classification, which are calculated in the feature extraction, from a feature bank including a large number of features prepared in advance by CAdE system developers. However, these data-driven methods are not yet extensively used in the development of CAdE systems. Since it is still necessary to manually design most of the specific algorithms in CAdE system development, the developers require specialized expertise in multiple research fields, making it very difficult for outsiders to enter the field.

Generally, a CAdE system can be divided into five components: preprocessing, segmentation, candidate detection, feature extraction, and candidate classification (Fig. 1). It works by sequentially executing the algorithms for each component. Although the components may be similar among CAdE systems, the algorithms of the components are usually completely different, depending on factors such as the detection target, the images to be processed, and the design policy of the algorithms. Commonly, a training dataset which consists of medical images and the area of the target lesions is collected in advance. Algorithms of all the components are manually designed based on the training dataset. Some

parameters used in the algorithms are optimized by trial and error with the training dataset. And the other parameters are optimized automatically by machine learning methods with the training dataset.

In this paper, we propose a generalized CAdE framework to develop various CAdE systems whose characteristics depend only on those of the training dataset. Pretrained algorithms, which are designed generally but without parameter optimization, are prepared for each component of the CAdE system. The pretrained algorithms are designed to be applicable to various lesion detection problems. Each pretrained algorithm is automatically optimized by a dedicated method based on machine learning and a training dataset. The framework was designed with the aim of automatically developing CAdE systems structured with the components of trained algorithms. If the framework is feasible, it is expected to lead to be a paradigm shift in the development of CAdE systems.

In this framework, developers are not required to manually design algorithms and are only required to collect the training dataset. Such a framework would enable outsiders, including clinicians, without programming skills to take part in the development of CAdE systems. It would provide opportunities for many clinicians to develop their own CAdE systems to meet their specific needs. While there are many algorithms available online, it is difficult for laypersons to integrate them to meet their needs. Therefore, a framework with all the necessary algorithms already prepared would be extremely valuable.

The purpose of this study is to show the feasibility of the generalized CAdE framework. To achieve this, two different CAdE systems were experimentally developed with the same components and the same pretrained algorithms, but with different training datasets.

Methods

Framework Design

Overview

In this study, CAdE systems are experimentally developed through a prototype generalized CAdE framework (Fig. 2). This prototype is a restricted version that detects only small and roughly spherical lesions in three-dimensional medical image data (volume data). A representative CAdE training dataset is first collected to develop a CAdE system. The dataset includes the target lesion area, which is the three-dimensional area of a target lesion defined by voxel-by-voxel labeling. The CAdE system includes four components: preprocessing, candidate area extraction, candidate detection, and candidate classification. In this framework, four pretrained algorithms with dedicated optimization/setting

Fig. 1 Flowchart of conventional CAdE system

Fig. 2 Prototype generalized CADe framework proposed in this study

methods corresponding to the respective components are prepared in advance. Since the optimization of the pretrained algorithm in a component requires the results of the previous component, the pretrained algorithms are sequentially trained in the order of processing of the components. In this section, we describe the details of each pretrained algorithm with the dedicated optimization/setting method.

Algorithms of the Components

1) Preprocessing

The first component creates isotropic volume data from input volume data to facilitate the subsequent processing. Trilinear interpolation is used to obtain the isotropic volume data, where the voxel size is $r_i \times r_i \times r_i$.

Setting Method r_i is automatically set on the basis of the size of the smallest lesion in the CADe training dataset. The bounding box is the minimum enclosing box for the smallest lesion. When the bounding box is $\mathbf{B}_{\min}$, r_i is set as follows:

$$r_i = \frac{1}{4} |\mathbf{B}_{\min}|^{\frac{1}{3}}, \quad (1)$$

where $|\mathbf{B}_{\min}|$ is the volume of $\mathbf{B}_{\min}$.

2) Candidate area extraction

The second component extracts the candidate area from the isotropic volume data. The candidate area is a limited three-dimensional region in which candidates are detected in the next component. The candidate area is extracted by a cascade [14] of voxel classifier ensembles that are automatically optimized by a boosting method and a training dataset, as described below. Each classifier ensemble consists of weak

classifiers whose classification performances are limited. The output value of the ensemble, which is used as a likelihood, is a weighted sum of the weak classifiers' outputs. All the weak classifiers within the classifier ensembles are decision stumps [15]. A decision stump is a simple classification model consisting of a one-level decision tree and makes a prediction by thresholding a single feature. In this component, the features used in all the weak classifiers are automatically selected from the voxel feature bank shown in Table 1 [16–18]. The feature selection results through the optimization process are affected by the characteristics of the training dataset. The voxel classifier ensembles within the cascade identify input voxels either as being inside of the target lesion (positive class) or as being outside of the target lesion (negative class). The voxels classified as belonging to the positive class by all classifier ensembles in the cascade are joined together to form the candidate area (Fig. 3). An example of a cascade processing result is shown in Fig. 4.

Optimization Method The voxel classifier ensembles and the number of classifiers in the cascade are automatically trained and optimized as follows. Cost-sensitive [19] AdaBoost [14, 20] is used to train all classifier ensembles within the cascade. Cost-sensitive learning is a classifier learning method with arbitrary weights added to training samples. For example, it is used to solve the imbalance between the training sample sizes of positive and negative classes. In the AdaBoost training, the features used in the weak classifiers (the decision stumps) are selected from the voxel feature bank shown in Table 1 [16–18]. The threshold value for the feature used in the decision stump is optimized through the AdaBoost training. Since the classifiers within the cascade are trained in numerical order, the voxel dataset used to train the t th ensemble comprises the voxels classified as members of the true class by the first $(t - 1)$ ensembles in the CADe training dataset. Each of the training voxel data is given a teacher label

Table 1 Contents of the voxel feature bank

Scaled voxel values ($n = 5$):

Original image and smoothed images scaled by uniform weighted cubic kernel (size = 1, 2, 4, 8, 16 voxels)

Morphological features ($n = 10$):

2 types (top-hat and bottom-hat filtering) $\times$ 5 spherical kernel sizes (radius = 1, 2, 4, 8, 16 voxels)

Difference of Gaussian features ($n = 5$):

5 pairs of Gaussian smoothers σ : (1, 2), (2, 4), (4, 8), (8, 16), (16, 32) voxels

Voxel value statistics ($n = 20$):

4 types (average, deviation, skewness, kurtosis) $\times$ 5 cubic ROI sizes (side length = 3, 5, 9, 17, 33 voxels)

3D Haar features [16] ($n = 95$):

19 block mask combinations $\times$ 5 cubic ROI sizes (side length = 3, 5, 9, 17, 33 voxels)

Haralick texture features [17] ($n = 55$):

11 types (angular second moment, inverse difference moment, contrast, variance, entropy, correlation, sum average, sum variance, sum entropy, difference variance, difference entropy) $\times$ 5 sets of 3D spatial offsets $(\delta_x, \delta_y, \delta_z)$: $\{(\pm d, 0, 0), (0, \pm d, 0), \text{ and } (0, 0, \pm d) | d = 1, 2, 4, 8, 16 \text{ voxels}\}$

Hessian matrix-derived features ($n = 45$):

9 types (3 eigenvalues, mean and Gaussian curvatures, shape index, curvedness, and principal curvatures k_1 and k_2) $\times$ 5 standard deviations of Gaussian kernel ($\sigma = 1, 2, 4, 8, 16$ voxels)

Local optimum scale-derived features [18] ($n = 8$):

Local optimum scale σ_{opt} (selected from 1, 2, 3, 4, and 5 voxels) + 7 types of feature: principal curvatures ($k_1, k_2, k_1/k_2$), magnitude of first derivative, S_{blob} , S_{bif} , and S_{line} calculated from Gaussian smoothed image by σ_{opt} -sized kernel

Standardized coordinate features ($n = 3$):

Standardized x , y , and z coordinates are used as the voxel features (see Appendix).

and a cost weight. The teacher label y_x of training voxel $\mathbf{x}$ is given by referring to the target area as follows:

$$y_x = \begin{cases} +1 & (\mathbf{x} \in \text{target area}) \\ -1 & (\text{otherwise}) \end{cases} \quad (2)$$

When the t th ensemble is undergoing training, the cost weight $w_{e(t),x}$ of voxel $\mathbf{x}$ is given as follows:

$$w_{e(t),x} = \begin{cases} \frac{\alpha_t}{n_{e(t)}^+} & (y_x = +1) \\ \frac{1}{n_{e(t)}^-} & (y_x = -1) \end{cases}, \quad (3)$$

where $n_{e(t)}^+$ and $n_{e(t)}^-$ are the numbers of “+1”-labeled voxels and “−1”-labeled voxels in the CADe training dataset for the t th ensemble, respectively. α_t is a parameter that controls the priority level of the positive voxels in the training. If $\alpha_t = 1.0$,

an ensemble is trained with equal weight on voxel sensitivity (ratio of correct classification for the positive voxels) and voxel specificity (ratio of correct classification for the negative voxels). If α_t is larger than 1.0, the training weight on the voxel sensitivity increases, if α_t is smaller than 1.0, the training weight on the voxel specificity increases. From among the many voxel classifier ensembles trained by adopting various parameters α_t , the classifier ensemble with the highest voxel classification specificity while achieving over 99.95% voxel sensitivity is selected as a member of the cascade. This sensitivity is calculated from the number of lesion voxels classified correctly and the total number of voxels within the volumetric areas of target lesions defined by experienced radiologists.

The number of ensembles in the cascade is optimized by performing a forward sequential search involving the repeated addition of an ensemble and evaluation of the lesion sensitivity. The addition of an ensemble to the cascade leads to not only the correct rejection of non-lesion voxels but also the

Fig. 3 Voxel classification cascade

Fig. 4 Example of the cascade processing result for a brain MRA. The arrow in the upper left image gives the position of a cerebral aneurysm. The bright areas in the binary images are aggregations of voxels

misclassification of lesion voxels. Thus, the sequential search is performed while maintaining a cascade classification that at least one voxel for each lesion in the training dataset is correctly classified. A later classifier ensemble of the cascade has a larger number of weak classifiers, as in [8]. Specifically, the numbers of weak classifiers included in the first nine ensembles are preset to 1, 10, 10, 25, 25, 50, 50, 100, and 100.

3) Candidate point detection

The third component detects candidate points of the target lesion from a candidate area. A candidate point is a local maximum of the lesion voxel likelihood. The lesion voxel likelihood, which is calculated at every voxel in the candidate area, is the output value of a voxel classifier ensemble. This ensemble is constructed from decision stumps based on voxel features but is different from the ensembles within the cascade and is trained to enhance voxels belonging to the target lesions. Figure 5 shows images giving an example of the calculated lesion voxel likelihood. The local maxima of the lesion voxel likelihood are selected as the candidate points.

Optimization Method Cost-sensitive AdaBoost is used to train the voxel classifier ensemble. The classifier ensemble consists of 200 decision stumps. Through the AdaBoost training method, voxel features used in the decision stumps are selected from a voxel feature bank different from the bank used in the optimization for candidate area extraction. The feature bank includes the features shown in Table 1 and the outputs of the voxel classifier ensembles used in the candidate area extraction. The voxel dataset used to train this voxel classifier ensemble is extracted from the rectangular sampling area $\mathbf{B}'$ for all lesions within the CADE training dataset. The rectangular sampling area $\mathbf{B}_\omega'$ for lesion ω is twice the size of

classified as belonging to the positive class by the first, second, and final cascade layers

the bounding box for the target area of ω ($\mathbf{B}_\omega$). Each of the training voxel data is given a teacher label (y_x) and a cost weight ($w_{d,x}$). The labeling rule for training voxels is the same as that in the training algorithm for candidate area extraction. The cost weight $w_{d,x}$ at training voxel $\mathbf{x}$ is given as follows:

$$w_{d,x} = \begin{cases} \frac{1}{n_d^+} & (y_x = +1) \\ \frac{1}{n_d^-} & (y_x = -1) \end{cases} \quad (4)$$

Here, n_d^+ is the number of “+1”-labeled voxels in the training voxel dataset and n_d^- is the number of “−1”-labeled training voxels in the training voxel dataset.

4) Candidate point classification

All candidate points are classified by a candidate classifier ensemble and are given lesion likelihoods. The lesion likelihood is the output of a candidate classifier ensemble. The ensemble is constructed from decision stumps and is different from the ensembles used in the candidate area extraction and candidate detection.

The weak classifiers constructing the candidate classifier ensemble are decision stumps based on pooling features, which are calculated by applying feature pooling operators [21] to voxel features. In this algorithm, three pooling operators are used, max-pooling $f_{\max}(\cdot)$, min-pooling $f_{\min}(\cdot)$, and average pooling $f_{\text{avg}}(\cdot)$, given as follows:

$$f_{\max}(v, \mathbf{Q}) = \max_{\mathbf{x} \in \mathbf{Q}} v_{\mathbf{x}}, \quad (5)$$

$$f_{\min}(v, \mathbf{Q}) = \min_{\mathbf{x} \in \mathbf{Q}} v_{\mathbf{x}}, \quad (6)$$

$$f_{\text{avg}}(v, \mathbf{Q}) = \text{average}_{\mathbf{x} \in \mathbf{Q}} v_{\mathbf{x}}. \quad (7)$$

Fig. 5 Example of calculation results for the lesion voxel likelihood showing an axial slice of a brain MRA with a cerebral aneurysm (a), cropped axial (left)/coronal (center)/sagittal (right) images centered on

the cerebral aneurysm (b), and an example of the calculation results for the lesion voxel likelihood corresponding to the cropped images (c). The arrows in (b) give the position of the same cerebral aneurysm

Here, v_x is the value of voxel feature v calculated at voxel x within a local region Q located at or around the detected candidate point.

Optimization Method Cost-sensitive AdaBoost is used to train the candidate classifier ensemble, which consists of 200 decision stumps. The dataset used to train the classifier ensemble includes candidate points detected from the CAde training dataset by executing the trained algorithms corresponding to preprocessing, candidate area extraction, and candidate detection. The candidate points inside/outside the target area are labeled as “+1”/“−1”. When the numbers of “+1”- and “−1”-labeled candidate points are n_c^+ and n_c^- , respectively, the label of candidate point p is y_p and its cost weight $w_{c,p}$ is

$$w_{c,p} = \begin{cases} \frac{1}{n_c^+} & (y_p = +1) \\ \frac{1}{n_c^-} & (y_p = -1) \end{cases}. \quad (8)$$

The feature bank used to train the candidate classifier ensemble includes the pooling features calculated on the basis of the voxel features shown in Table 1, the outputs of the ensembles used for candidate area extraction, and the output of the ensemble used for candidate detection. To calculate the pooling features in the feature bank, 64 different local regions are used, which have different sizes and positions relative to the candidate point. All the local regions are located within the region of interest, which comprises $8 \times 8 \times 8$ voxels centered on the candidate point.

Development and Evaluation of CAde System

Two different CAde training datasets were collected to develop two different types of CAde systems in the generalized CAde framework. These two CAde training datasets differ in terms of the image modality, scanning range, and target lesion. Here, two radiologists first independently interpreted the medical images, and then the final decision was determined with the agreement of the two radiologists. Both radiologists had over 10 years of CT- and MRA-reading experience. Both datasets include three-dimensional areas of all lesions that were manually defined by the radiologists by voxel-by-voxel painting.

Brain Magnetic Resonance Angiography (MRA) Volume Dataset with Cerebral Aneurysms

Three hundred sets of 3D time-of-flight unenhanced MRA images were used (151 males and 149 females). The average patient age was 59.8 years with a standard deviation (SD) of 11.4 years. These images were scanned at our institution with a GE Signa HDxt 3.0T or a GE Discovery MR750 3.0T magnetic resonance imaging scanner (GE Healthcare, Waukesha, WI, USA). The acquisition parameters were as follows: echo time, 2.7 or 2.9 ms; repetition time, 25 ms; flip angle, 15° ; field of view, 240 mm; slice thickness, 0.6 mm; slice interval, 1.2 mm; matrix size, 512×512 pixels. The original spatial resolution of all obtained images was $0.468 \times 0.468 \times 0.60 \text{ mm}^3/\text{voxel}$. Each set included at least one aneurysm of 2 mm or more in diameter. The diameter was manually measured by radiologists at the initial interpretation. The average diameter of the aneurysms is 3.1 mm with an SD of 1.4 mm. The average number of lesions included in a set was 1.11 with an SD of 0.33. Examples of images with aneurysms are shown in Fig. 6 (a).

Chest Computed Tomography (CT) Volume Dataset with Pulmonary Nodules One hundred and twenty-nine sets of unenhanced chest CT images were used (74 males and 55 females). The average patient age was 59.3 years with an SD of 10.6 years. 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; field of view, 400 mm; rotation time, 0.5 s; moving table speed, 70 mm/s; body filter, standard; matrix size, 512×512 pixels. The original spatial resolution of all images was $0.781 \times 0.781 \times 1.25 \text{ mm}^3/\text{voxel}$. Each set included at least one solid nodule or ground glass opacity (GGO) nodule of 5 mm or more in diameter. The diameter was manually measured by radiologists at the initial interpretation. The average diameter of the nodules is 7.8 mm with an SD of 3.8 mm. The average number of lesions included in a set was 1.45 with an SD of 0.83. Examples of images with nodules are shown in Fig. 6 (b).

The performances of the CAde systems developed through the generalized CAde framework were evaluated by threefold cross-validation. In threefold cross-validation, the dataset was divided into three subsets: two subsets for optimizing

pretrained algorithms and remaining one subset for evaluating the obtained CAde system. The validation was iterated three times by rotating the subset. Consequently, three CAde systems were obtained from each CAde training dataset in this study. In the performance evaluation of the developed CAde systems, candidate points located inside/outside any painted lesion area were judged from the lesion likelihood as either true positive (TP) or false positive (FP), respectively. Free-response receiver operating characteristic (FROC) curves and ANODE scores [22] were calculated, and the time required for the lesion detection for each input volume was measured. By changing the threshold value for the lesion likelihood, plots of the TP fraction (sensitivity) and FP per case (the average number of FPs for a case) for the FROC curves were obtained. The ANODE score defines the average sensitivity at predefined FP rates (1/8, 1/4, 1/2, 1, 2, 4, and 8 FPs per case) along a FROC curve. The length of time from inputting a CAde training dataset to outputting a CAde system was also measured.

C/C++ language was used to implement the algorithms and the automatic optimization/setting methods. The experiments were performed on a workstation with Intel Xeon X5680 (3.33 GHz $\times$ 2) CPUs, 72 GB RAM, and a Microsoft Windows 7 Professional SP1 operating system. This study

Fig. 6 Examples of images included in the brain MRA volume dataset (a) and examples of images included in the chest CT volume dataset (b). Arrows in the images gives the positions of lesions

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.

Results

CADe systems for detecting cerebral aneurysms in the brain MRA were successfully developed using the brain MRA volume dataset with cerebral aneurysms. The average time to develop the CADe systems was 28.7 h with an SD of 3.0 h. The cerebral aneurysm detection processes for all 300 sets were completed without problems. The ANODE score had an average of 0.239 with an SD of 0.062 (Fig. 7). The configurations of the CADe systems developed with the same type of dataset resembled each other, as inferred from the small SD for each item shown in Table 2. The average detection time per case was 106.2 s with an SD of 19.6 s.

The CADe systems for detecting lung nodules in the chest CT were also successfully developed using the chest CT volume dataset with pulmonary nodules. The average time to develop the CADe systems was 22.5 h with an SD of 0.4 h. The lung nodule detection process for all 129 sets was completed without problems. The ANODE score had an average of 0.218 with an SD of 0.058 (Fig. 8). The configurations of the developed CADe systems also resembled each other, as inferred from the small SD for each item shown in Table 2. The average processing time per case was 200.2 s with an SD of 22.5 s.

Fig. 7 FROC curves of the CADe systems for detecting cerebral aneurysms

Discussion

We experimentally confirmed that two different types of CADe system were developed successfully through the generalized CADe framework using two different training datasets. The CADe systems were developed automatically without the need for manual design, the characteristics of which depended only on the CADe training datasets. The two successfully developed CADe systems show that the generalized CADe framework is feasible.

To the best of our knowledge, this is the first report on the development of a whole CADe system without any manual design of the algorithms. Classically, the algorithms in the CADe systems were designed manually. In the generalized CADe framework, all the algorithms were automatically trained by optimization/setting methods using the collected CADe training dataset. None of the pretrained algorithms were specialized for a particular modality or type of lesion. Moreover, the machine learning methods used in the optimization methods included a function to automatically select optimal features from a large-scale feature bank since the optimal feature set was dependent on the characteristics of the detection/classification task. Theoretically, with an appropriate training dataset, a CADe system for detecting any type of lesion can be developed in this framework.

The results of this study show promise of a new paradigm for CADe system developments. It would enable clinicians without expertise in medical image analysis to develop CADe systems to use in their clinical site. Anyone with access to an appropriate dataset of medical images could develop various CADe systems required for their particular clinical situation. In addition, artificial intelligence researchers will be able to concentrate on the development of algorithms in the framework without considering the specific features of the disease to be detected. A generalized CADe framework would enable the development of CADe systems to be divided into the development of clinical applications and the development of the algorithms.

CADe system development in the generalized CADe framework requires less labor than previous CADe system developments with the manual design of algorithms. The manual design of algorithms is a trial-and-error procedure using a CADe training dataset, which is labor-intensive. In the generalized CADe framework, the manual design of algorithms is not required because all the pretrained algorithms are automatically trained by the optimization methods. Using the generalized CADe framework, the collection of the CADe training dataset is the only human task required to develop CADe systems, which was also necessary in previous CADe system developments. Adopting machine learning methods will greatly reduce the workload of CADe system developers.

A limitation of this study is the low performances of the developed CADe systems, which were inferior to those of

Table 2 Specifications and detection performance of the CADe systems for detecting cerebral aneurysms and lung nodules developed through the generalized CADe framework

	Parameter(s)	Performance/specification (mean $\pm$ standard deviation)	
		Cerebral aneurysm CADe	Lung nodule CADe
Preprocessing	Voxel size (mm)	0.676 ± 0.017	1.700 ± 0.047
Candidate area extraction	Number of cascade layers	7.67 ± 1.15	8.00 ± 1.00
	Ratio of voxel removal (%)	99.92 ± 0.04	99.15 ± 0.31
	Lesion-based sensitivity	0.986 ± 0.018	0.970 ± 0.028
Candidate detection	Number of candidates/cases	743.6 ± 333.3	7642.2 ± 1898.5
	Lesion detection sensitivity	0.871 ± 0.032	0.882 ± 0.049
Candidate classification	ANODE score	0.239 ± 0.062	0.218 ± 0.058

state-of-the-art CADe systems [18, 22, 23]. Low performance may be a price we need to pay for generalization of development. For the cerebral aneurysm CADe system, our system achieved a median sensitivity of 56.8% with ten FPs/case, while Yang et al. achieved a sensitivity of 80% with three FPs/case and 95% with nine FPs/case in a performance evaluation with about 287 MRA scans [23]. For the lung nodule CADe, our system achieved a mean ANODE score of 0.218, while the ANODE09 study [19] in 2009 achieved a median ANODE score of 0.272. However, it should be noted that the performances of CADe systems trained with different datasets cannot easily be compared since CADe systems evaluated with a small dataset often show superior performance. Arimura et al. developed a cerebral aneurysm CADe system with only 31 cases [24], which showed better performance (a sensitivity of 100% with 2.4 FPs/case) than that of Yang et al. [23].

However, the aim of this paper is not to maximize the performances of the developed CADe systems but to show the feasibility of the framework to develop various CADe

systems without manual design of the algorithms since this is a pilot study of the generalized CADe framework. There is ample room for improving this prototype framework. One weakness of this prototype is the inability to integrate local anatomical information into the CADe system. In most CADe systems with segmentation algorithms, the algorithms are uniquely designed for the specific detection of the target lesion, which is the mainstream design for integrating local anatomical information. While a versatile segmentation algorithm that can be applied to any anatomical structure has not yet been proposed [4], introducing general anatomical features such as the distance to an anatomical landmark [25–27] will improve the performance of CADe systems developed with this framework without application-specific design.

Another weakness of the prototype is that it was applied under restricted conditions; the target was limited to roughly spherical lesions with a small volume and the features used were limited in number. Extending the range of the pretrained algorithms will enable the detection of larger lesions, and expanding the range of applications of the pretrained algorithms will enable the detection of lesions from two-dimensional image data. Moreover, the features used were derived from a feature bank with a limited number of voxel features. Feature generation methods, such as deep learning methods, will be effective means of obtaining the optimal features for the target lesions without any manual design [28].

Conclusion

We proposed a generalized CADe framework for developing CADe systems without any manual design for various types of target detection. We experimentally confirmed that two different types of CADe system were developed successfully through the generalized CADe framework using two different training datasets. All the algorithms used in the developed CADe systems were trained automatically by optimization/setting methods instead of being designed manually. The

Fig. 8 FROC curves of the CADe systems for detecting lung nodules

characteristics of the CADe systems depended only on the CADe training datasets. The feasibility of this framework shows promise of a new paradigm for the development of CADe systems. In this framework, developers are not required to manually design algorithms and are only required to collect the training dataset. Such a framework would enable clinicians without programming skills to take part in the development of CADe systems. This would provide opportunities for many clinicians to develop their own CADe systems to meet their specific needs. A generalized CADe framework with all the necessary algorithms already prepared would be extremely valuable.

Acknowledgements The Department of Computational Radiology and Preventive Medicine, The University of Tokyo Hospital is sponsored by HIMEDIC Inc., Siemens Japan K.K., and GE Healthcare Japan.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Ethical Standards The study described in this manuscript was approved by the Research Ethics Board of the University of Tokyo Hospital. Informed consent was obtained from all individual participants included in the study.

Appendix: Calculation of the standard coordinate features

The standard coordinate features were the standardized x , y , and z coordinates. To calculate the features, n_a atlas volumes were randomly chosen from the CADe training dataset and were registered to an input volume in advance. In this study, five atlas volumes and a volume registration method based on Markov random fields [29, 30] were used. The standard coordinate $\mathbf{c}_p$ of voxel $\mathbf{p}$ in the input volume is given as follows:

$$\mathbf{c}_p = \frac{1}{n_a} \sum_{i=1}^{n_a} \mathbf{r}_p^i. \quad (9)$$

Here, $\mathbf{r}_p^i$, which is the coordinate in the i th registered atlas volume, corresponds to $\mathbf{p}$ in the input volume.

References

- Castellino RA: Computer aided detection (CAD): an overview. *Cancer Imaging* 5:17–19, 2005
- Giger ML, Chan H, Boone J: Anniversary paper: history and status of CAD and quantitative image analysis: the role of medical physics and AAPM. *Medical Physics* 35:5799–5820, 2008
- Doi K: Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Computerized Medical Imaging and Graphics* 31:198–211, 2007
- van Ginneken B, Schaefer-Prokop CM, Prokop M: Computer-aided diagnosis: how to move from the laboratory to the clinic. *Radiology* 261:719–732, 2011
- Chan HP, Lo SCB, Sahiner B, Lam KL, Helvie MA: Computer-aided detection of mammographic microcalcifications: pattern recognition with an artificial neural network. *Medical Physics* 22: 1555–1567, 1995
- Gokturk SB, Tomasi C, Acar B, Beaulieu CF, Paik DS, Jeffrey RB Jr, Yee J, Napel S: A statistical 3-D pattern processing method for computer-aided detection of polyps in CT colonography. *IEEE Trans Med Imaging* 20:1251–1260, 2001
- Nemoto M, Shimizu A, Hagiwara Y, Kobatake H, Nawano S: Improvement of tumor detection performance in mammograms by feature selection from a large number of features and proposal of fast feature selection method. *Syst Comput Jpn* 37:56–68, 2006
- Miller MT, Jerebko AK, Malley JD, Summers RM: Feature selection for computer-aided polyp detection using genetic algorithms. *Proceedings of SPIE (Medical Imaging)* 5031:102–110, 2003
- Wang S, Summers RM: Machine learning and radiology. *Med Image Anal* 16:933–951, 2012
- Bishop CM: *Pattern recognition and machine learning*, Springer, New York, 2006
- Duda RO, Hart PE, Stork DG: *Pattern classification*, 2nd edition, John Wiley & Sons, New York, 2000
- Mitchell TM: *Machine learning*, International Edition, McGraw-Hill Education, New York, 1997
- Sahiner B, Chan H, Petrick N, Wagner RF, Hadjiiski L: Feature selection and classifier performance in computer-aided diagnosis: the effect of finite sample size. *Med Phys* 27:1509–1522, 2000
- Viola P, Jones M: Rapid object detection using a boosted cascade of simple features. *Proc IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001)* 1:1511–1518, 2001
- Iba W, Langley P: Induction of one-level decision trees. *Proc International Conference on Machine Learning (ICML 1992)*: 233–240, 1992
- Tu Z, Bai X: Auto-context and its application to high-level vision tasks and 3D brain image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32:1744–1757, 2010
- Haralick RM: Statistical and structural approaches to texture. *Proc IEEE* 67:786–804, 1979
- Nomura Y, Masutani Y, Miki S, Nemoto M, Hanaoka S, Yoshikawa T, Hayashi N, Ohtomo K: Performance improvement in computerized detection of cerebral aneurysms by retraining classifier using feedback data collected in routine reading environment. *Journal of Biomedical Graphics and Computing* 4:12–21, 2014
- Sun Y, Kamel MS, Wong AK, Wang Y: Cost-sensitive boosting for classification of imbalanced data. *Pattern Recognition* 40:3358–3378, 2007
- Tu Z, Zhou XS, Barbu A, Bogoni L, Comaniciu D: Probabilistic 3D polyp detection in CT images: the role of sample alignment. *Proc IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2006
- Boureau Y, Ponce J, LeCun Y: A theoretical analysis of feature pooling in visual recognition. *Proc International Conference on Machine Learning (ICML)*:111–118, 2010
- van Ginneken B, Armato SG, de Hoop B, van Amelsvoort-van de, Vorst, Saskia, Duindam T, Niemeijer M, Murphy K, Schilham A, Retico A, Fantacci ME: Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: the ANODE09 study. *Medical Image Analysis* 14:707–722, 2010
- Yang X, Blezek DJ, Cheng LT, Ryan WJ, Kallmes DF, Erickson BJ: Computer aided detection of intracranial aneurysms in MR angiography. *J Digital Imaging* 24:86–95, 2011

24. Arimura H, Li Q, Korogi Y, Hirai T, Abe H, Yamashita Y, Katsuragawa S, Ikeda R, Doi K: Automated computerized scheme for detection of unruptured intracranial aneurysms in three-dimensional magnetic resonance angiography. *Acad Radiol* 11: 1093–1104 2004
25. Rohr K: Landmark-based image analysis: using geometric and intensity models. Springer Netherlands, Utrecht, 2001
26. Nemoto M, Masutani Y, Hanaoka S, Nomura Y, Yoshikawa T, Hayashi N, Yoshioka N, Ohtomo K: A unified framework for concurrent detection of anatomical landmarks for medical image understanding. *Proc SPIE (Medical Imaging)* 7962:79323E, 2011
27. Hanaoka S, Shimizu A, Nemoto M, Nomura Y, Miki S, Yoshikawa T, Hayashi N, Ohtomo K, Masutani Y: Automatic detection of over 100 anatomical landmarks in medical CT images: a framework with independent detectors and combinatorial optimization. *Med Image Anal* 35:192–214, 2016
28. Nemoto M, Hayashi N, Hanaoka S, Nomura Y, Miki S, Yoshikawa T, Ohtomo K: A primitive study of voxel feature generation by multiple stacked denoising autoencoders for detecting cerebral aneurysms on MRA. *Proc SPIE (Medical Imaging)* 9785:97852S, 2016
29. Glocker B, Sotiras A, Komodakis N, Paragios N: Deformable medical image registration: setting the state of the art with discrete methods. *Annu Rev Biomed Eng* 13:219–244, 2011
30. Komodakis N, Tziritas G, Paragios N: Fast, approximately optimal solutions for single and dynamic MRFs. *Proc IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*: 1–8, 2007