

Massive-training artificial neural network (MTANN) for reduction of false positives in computer-aided detection of polyps: Suppression of rectal tubes

Kenji Suzuki^{a)}

Department of Radiology, The University of Chicago, 5841 South Maryland Avenue, Chicago, Illinois 60637

Hiroyuki Yoshida and Janne Näppi

Department of Radiology, Massachusetts General Hospital and Harvard Medical School, 75 Blossom Court, Suite 220, Boston, Massachusetts 02114

Abraham H. Dachman

Department of Radiology, The University of Chicago, 5841 South Maryland Avenue, Chicago, Illinois 60637

(Received 4 April 2006; revised 27 July 2006; accepted for publication 4 August 2006; published 25 September 2006)

One of the limitations of the current computer-aided detection (CAD) of polyps in CT colonography (CTC) is a relatively large number of false-positive (FP) detections. Rectal tubes (RTs) are one of the typical sources of FPs because a portion of a RT, especially a portion of a bulbous tip, often exhibits a cap-like shape that closely mimics the appearance of a small polyp. Radiologists can easily recognize and dismiss RT-induced FPs; thus, they may lose their confidence in CAD as an effective tool if the CAD scheme generates such “obvious” FPs due to RTs consistently. In addition, RT-induced FPs may distract radiologists from less common true positives in the rectum. Therefore, removal RT-induced FPs as well as other types of FPs is desirable while maintaining a high sensitivity in the detection of polyps. We developed a three-dimensional (3D) massive-training artificial neural network (MTANN) for distinction between polyps and RTs in 3D CTC volumetric data. The 3D MTANN is a supervised volume-processing technique which is trained with input CTC volumes and the corresponding “teaching” volumes. The teaching volume for a polyp contains a 3D Gaussian distribution, and that for a RT contains zeros for enhancement of polyps and suppression of RTs, respectively. For distinction between polyps and nonpolyps including RTs, a 3D scoring method based on a 3D Gaussian weighting function is applied to the output of the trained 3D MTANN. Our database consisted of CTC examinations of 73 patients, scanned in both supine and prone positions (146 CTC data sets in total), with optical colonoscopy as a reference standard for the presence of polyps. Fifteen patients had 28 polyps, 15 of which were 5–9 mm and 13 were 10–25 mm in size. These CTC cases were subjected to our previously reported CAD scheme that included centerline-based segmentation of the colon, shape-based detection of polyps, and reduction of FPs by use of a Bayesian neural network based on geometric and texture features. Application of this CAD scheme yielded 96.4% (27/28) by-polyp sensitivity with 3.1 (224/73) FPs per patient, among which 20 FPs were caused by RTs. To eliminate the FPs due to RTs and possibly other normal structures, we trained a 3D MTANN with ten representative polyps and ten RTs, and applied the trained 3D MTANN to the above CAD true- and false-positive detections. In the output volumes of the 3D MTANN, polyps were represented by distributions of bright voxels, whereas RTs and other normal structures partly similar to RTs appeared as darker voxels, indicating the ability of the 3D MTANN to suppress RTs as well as other normal structures effectively. Application of the 3D MTANN to the CAD detections showed that the 3D MTANN eliminated all RT-induced 20 FPs, as well as 53 FPs due to other causes, without removal of any true positives. Overall, the 3D MTANN was able to reduce the FP rate of the CAD scheme from 3.1 to 2.1 FPs per patient (33% reduction), while the original by-polyp sensitivity of 96.4% was maintained. © 2006 American Association of Physicists in Medicine. [DOI: 10.1118/1.2349839]

Key words: virtual colonoscopy, colon cancer, CT colonography, computer-aided detection, pattern recognition, artificial neural network

I. INTRODUCTION

CT colonography (CTC), also known as virtual colonoscopy, is a technique for detecting colorectal neoplasms by use of a CT scan of the colon, and it is a promising technique for

providing mass screening for colorectal carcinoma.¹ However, the diagnostic performance of CTC in detecting polyps, precursors of colorectal cancer, remains uncertain, with a propensity for perceptual errors and substantial variations

among readers across different studies.² Computer-aided detection (CAD) of polyps has the potential to overcome these difficulties with CTC.^{3,4} CAD for CTC typically refers to a computerized scheme for detection of polyps in the CTC images and displays the locations of suspicious polyps for radiologists' review. CAD also has the potential to improve radiologists' diagnostic performance in the detection of polyps.^{3,4}

Several investigators have developed automated or semi-automated CAD schemes for detection of polyps in CTC, and have conducted clinical trials to demonstrate the performance of their CAD schemes.⁵⁻¹² Although the performance of the current CAD schemes demonstrated promise, some limitations remain; one of the major limitations is a relatively large number of false-positive (FP) detections. A large number of FPs could complicate the clinical application of the current CAD schemes because these FPs are likely to confound the radiologist's image-interpretation task and thus may lower their accuracy.

Haustral folds, residual stool, extra-colonic structures such as small bowel and stomach, rectal tubes (RTs), and the ileocecal valve have been reported as common sources of FPs generated by CAD schemes.⁴ Among them, the occurrence of RT-induced FPs is relatively low;³ however, RTs remain a source of FPs in CAD because a portion of a RT, especially a portion of a bulbous tip, often exhibits a cap-like shape that closely mimics the appearance of a small polyp. Moreover, the tip of a RT sometimes touches the colonic wall; this makes the differentiation of the RT from soft-tissue structures difficult, and thus it increases the chance that the tip of the RT is erroneously detected as a polyp. The performance of a CAD scheme involves a trade-off between sensitivity and specificity, which can be designed by a threshold on the output of a classifier. It is important to remove as many FPs, including RTs, as possible, while the sensitivity of the CAD scheme is maintained.

On the other hand, the interface between the air and the surface of a RT is clear; thus, radiologists can easily recognize and dismiss FPs due to RTs. Therefore, radiologists may lose their confidence in CAD as an effective tool if the CAD scheme generates such "obvious" FPs due to RTs consistently. Removal of CAD FPs due to RTs is thus desirable not only for improving the overall performance of CAD, but also for increasing the confidence of radiologists in CAD.

Although various methods characterizing FPs have been developed for reduction of their number,¹³⁻²¹ only one of these methods¹⁹ was specifically designed for RTs. We thus develop a three-dimensional massive-training artificial neural network (3D MTANN) to this reduction task and evaluate its performance based on clinical CTC cases.

In the field of image processing, ANN-based supervised nonlinear image-processing techniques, known as "neural filters,"²² and "neural edge enhancers,"²³ have been investigated for the reduction of quantum noise in coronary angiograms²⁴ and upper gastric radiographs²² and for the supervised detection of left ventricular contours traced by cardiologists in ventriculograms.²⁵ By extending the neural filter and the neural edge enhancer, two-dimensional (2D)

MTANNs²⁶ have been developed to accommodate the task of distinguishing a specific opacity from other opacities in medical images. 2D MTANNs have been applied for reduction of FPs in the computerized detection of lung nodules in low-dose CT^{26,27} and chest radiography,²⁸ for distinction between benign and malignant lung nodules in CT,²⁹ and for suppression of ribs in chest radiographs.³⁰ Thus, a MTANN for processing 3D volume data has not been developed yet.

II. MATERIALS AND METHODS

A. Architecture of a 3D massive-training artificial neural network (3D MTANN)

To process 3D volume data in CTC, we developed a 3D MTANN by extending the structure of the 2D MTANN. The architecture and training of a 3D MTANN are shown in Fig. 1. The 3D MTANN consists of a linear-output multilayer ANN model,³¹ which is capable of operating on voxel data directly. The 3D MTANN is trained with input CTC volumes and the corresponding "teaching" volumes for enhancement of polyps and suppression of RTs. The pixel size within a CT section is generally different from the reconstruction interval between sections, and the reconstruction interval is often different at different institutions and under different imaging protocols. To reduce such variations in CTC data, original CTC data are converted to isotropic volume data. The voxel values of the isotropic volumes are linearly scaled such that -1000 Hounsfield unit (HU) corresponds to 0 and 1000 HU corresponds to 1 (values below 0 and above 1 are allowed). The input to the 3D MTANN is the voxel values in a sub-volume, V_S , extracted from an input isotropic volume. The output, $O(x, y, z)$, of the 3D MTANN is a continuous scalar value, which corresponds to the center voxel in the sub-volume, and is represented by

$$O(x, y, z) = NN(\mathbf{I}_{x,y,z}), \quad (1)$$

where

$$\mathbf{I}_{x,y,z} = \{I(x-i, y-j, z-k) | i, j, k \in V_S\} \quad (2)$$

is the input vector to the 3D MTANN, x, y , and z are the coordinate indices, $NN(\cdot)$ is the output of the linear-output ANN model, and $I(x, y, z)$ is the normalized voxel value of the input isotropic volume.

We use a linear-output ANN model that employs a linear function instead of a sigmoid function as the activation function of the output unit, because the output and convergence characteristics of an ANN are improved significantly with a linear function when it is applied to the problem of continuous mapping of values in image processing, as follows:^{23,31} A conventional ANN hardly provides values near zero and one because of the characteristics of a sigmoid function, whereas the linear-output multilayer ANN outputs values linearly. Theoretically, the training for teaching values near zero and one converges more slowly than do other values with a conventional ANN, whereas these values are trained evenly with a linear-output multilayer ANN model. This affects the convergence characteristics and the output characteristics of ANN models. Therefore, the linear-output multilayer ANN

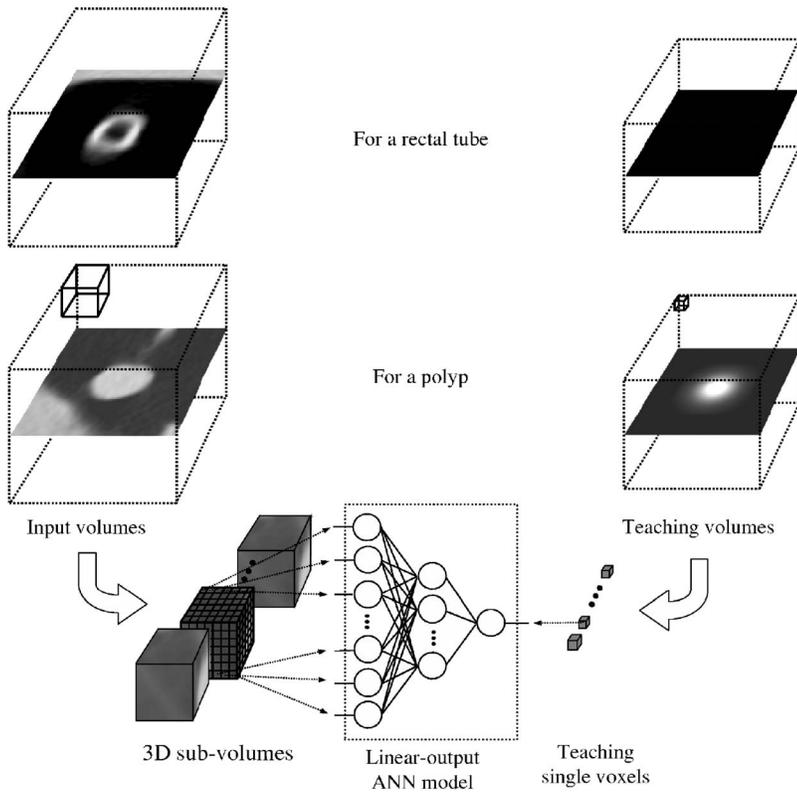


FIG. 1. Architecture and training of a 3D MTANN consisting of a linear-output multilayer ANN model and a massive-subvolumes training scheme. The input CTC volume including a polyp or a RT is divided voxel by voxel into a large number of overlapping 3D subvolumes. All voxel values in each of the subvolumes are entered as input to the 3D MTANN, whereas a voxel value at each voxel from the teaching volume is used as the teaching value.

would be suitable for image/volume processing, where the teaching values may be continuous values ranging from zero to one, whereas the conventional ANN is suitable for a classification task where the teaching values are assigned to classes (see Ref. 23 for theoretical considerations). The entire output volume is obtained by scanning of an input CTC volume with the 3D MTANN in 3D space. It should be noted that the input CTC volume needs to contain the entire polyp candidate (a polyp or a RT).

The training of a 2D MTANN for chest radiography,^{28,30} and thick-slice thoracic CT^{26,27,29} takes a substantially long time, e.g., 30 h on a PC-based workstation (CPU: Pentium IV, 1.7 GHz);²⁶ consequently, efficient architecture and training are a necessity for a 3D MTANN. Since the average shape of polyps approximates a sphere, the shape of the sub-volume input to the MTANN can be spherical rather than cubic. The volume of a sphere with radius r is $\pi/6$ times the volume of a cube with an edge length of $2r$ (i.e., the sphere

inscribed in a cube). By use of a spherical subvolume rather than a cubic, the computational cost can be reduced to 52% ($\pi/6$) (i.e., by 48%). Therefore, we employed the digital quasisphere shown in Fig. 2 as the input subvolume for the 3D MTANN. The number of voxels in this input subvolume (Fig. 2) is 171, or about 50% (171/343) of the number of voxels in a cubic subvolume in which the sphere is inscribed.

The number of hidden units may be selected by use of a method for designing the structure of an ANN.^{32,33} The method is a sensitivity-based pruning method, i.e., the sensitivity to the training error is calculated when a certain unit is removed experimentally, and the unit with the smallest training error is removed. Removing the redundant hidden units and retraining for recovering the potential loss due to the removal are performed repeatedly, resulting in a reduced structure where redundant units are removed. As a result, the number of hidden units is determined optimally.

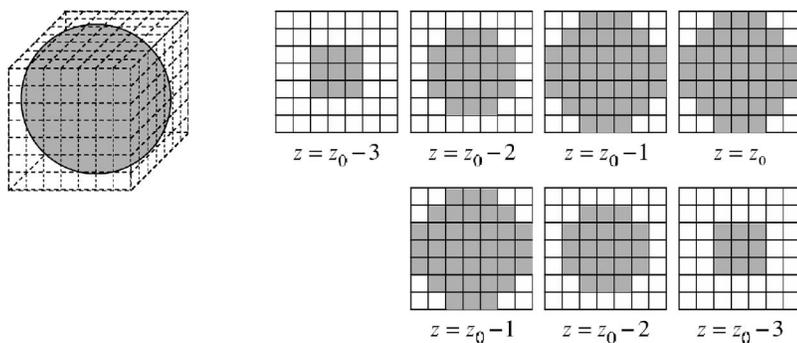


FIG. 2. The spherical-input subvolume of a 3D MTANN and the slice-by-slice representation of the digital quasisphere in a $7 \times 7 \times 7$ voxel cube. Each 2D matrix represents an x - y plane at a certain z position in the input 3D subvolume, where z_0 represents the central slice of the subvolume. A gray square in each matrix indicates the input voxel to the linear-output ANN in the 3D MTANN, and a white square indicates a non-used voxel.

B. Training of a 3D MTANN

For enhancement of polyps and suppression of RTs in CTC volumes, the teaching volume contains a 3D distribution of values that represent the “likelihood of being a polyp.” We used a 3D Gaussian distribution with standard deviation σ_T , the peak of which is located at the center of the polyp, as a teaching volume for a polyp and a volume that contains all zeros for a RT, represented by

$$T(x,y,z) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma_T} \exp\left\{-\frac{(x^2+y^2+z^2)}{2\sigma_T^2}\right\} & \text{if an actual polyp} \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

The 3D MTANN involves training with a large number of subvolume-voxel pairs; we call it a massive-subvolumes training scheme. A training volume V_T extracted from the input CTC volume is divided voxel by voxel into a large number of overlapping subvolumes. Single voxels are extracted from the corresponding teaching volume as teaching values. The 3D MTANN is massively trained by use of each of a large number of the input subvolumes together with each of the corresponding teaching single voxels. A training set of pairs of a subvolume and a teaching voxel is represented by

$$\{\mathbf{I}(x,y,z), T(x,y,z) | x,y,z \in V_T\} = \{(\mathbf{I}_1, T_1), (\mathbf{I}_2, T_2), \dots, (\mathbf{I}_p, T_p), \dots, (\mathbf{I}_{N_T}, T_{N_T})\}, \quad (4)$$

where V_T is a training volume, p is a voxel number in V_T , T_p is a teaching value in the teaching volume that corresponds to the center voxel in \mathbf{I}_p , and N_T is the number of voxels in V_T . The error to be minimized by training is given by

$$E = \frac{1}{P} \sum_c \sum_{x,y,z \in V_T} \{T_c(x,y,z) - O_c(x,y,z)\}^2, \quad (5)$$

where c is a training case number, and P is the number of total training voxels in V_T 's. The 3D MTANN is trained by a linear-output backpropagation (BP) algorithm,^{23,31} which was derived for the linear-output ANN model by use of the same method used for deriving the original BP algorithm³⁴ (see Refs. 23 and 31 for details and the property of the linear-output BP algorithm). After training, the 3D MTANN is expected to output the highest value when a polyp is located at the center of the subvolume of the 3D MTANN, a lower value as the distance from the subvolume center increases, and zero when the input subvolume contains a non-polyp.

C. 3D scoring method for classification of polyps and RTs

For distinction between polyps and RTs, we developed a 3D scoring method based on the output volume of the trained 3D MTANN. A score for a given polyp candidate from the 3D MTANN is defined as

$$S = \sum_{x,y,z \in V_E} f_G(\sigma; x,y,z) \times O(x,y,z), \quad (6)$$

where

$$f_G(\sigma; x,y,z) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x^2+y^2+z^2)}{2\sigma^2}\right\} \quad (7)$$

is a 3D Gaussian weighting function with standard deviation σ with its center corresponding to the center of the volume for evaluation, V_E ; V_E is the volume for evaluation that is sufficient to cover a polyp or a RT; and $O(x,y,z)$ is the output of the trained 3D MTANN. The use of the 3D Gaussian weighting function allows us to combine the individual voxel-based responses (outputs) of a trained 3D MTANN as a single score. The score obtained by the above-presented equations represents the weighted sum of the estimates for the likelihood that the volume (polyp candidate) contains an actual polyp near the center, i.e., a higher score would indicate a polyp, and a lower score would indicate a RT. The concept of this scoring is similar to that of a matched filter. We use the same 3D Gaussian weighting function as is used in the polyp teaching volumes. Thresholding is performed on the scores to distinguish between polyps and RTs.

It is difficult to distinguish a small distribution for a small polyp in the output volume from a small distribution due to noise; this difficulty can lower the ability of the 3D MTANN to differentiate polyps from RTs. To force the 3D MTANN to output a standard-sized (regular-sized) distribution for different-sized polyps, the same-sized Gaussian distribution is used in the teaching volumes. After training in this manner, the 3D MTANN is expected to output relatively regular-sized distributions for different-sized polyps, e.g., a relatively large output distribution for a small polyp and a relatively small output distribution for a large polyp. This property of the regular-sized output distributions is expected to increase the scores for small polyps and to improve the overall performance of a 3D MTANN.

D. Simulation experiments

To investigate the fundamental characteristics of the trained 3D MTANN, we carried out experiments with simulated CTC volumes that contained computer-simulated polyps and RTs. A polyp was modeled as a sphere with diameter d , and a RT was modeled as a hollow cylinder with diameter d_T , length l , and wall thickness t_W , as shown in Fig. 3. We employed these simple phantom models, because we aimed at examining only the fundamental characteristics of the trained 3D MTANN. The simulated CTC volumes with polyps and RTs of three different sizes (d : 6, 15, and 25 mm; d_T : 10, 13, and 16 mm) are illustrated in the top image in Fig. 4. The CT values for the simulated polyps and RTs were set to 60 and 180, respectively, based on measurements of actual polyps and RTs in the clinical CTC volumes. The length l was 70 mm; the wall thickness t_W was 2 mm. We applied the 3D MTANN trained with actual polyps and RTs (see Sec. III A) to the phantom CTC volumes, and applied the 3D scoring method to the resulting output volumes for distinc-

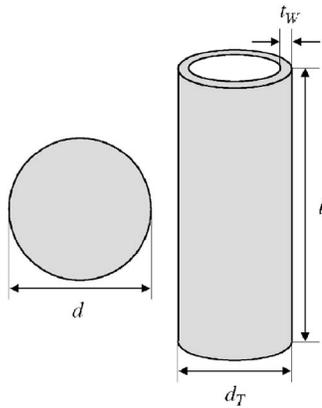


FIG. 3. Schematic illustration of a polyp phantom (a sphere) and a RT phantom (a hollow cylinder). These simple phantom models are employed for examining only the fundamental characteristics of the trained 3D MTANN.

tion between polyps and RTs. The standard deviation of the 3D Gaussian weighting function was the same as that of the 3D Gaussian distribution in the polyp teaching volume.

E. CTC database and the performance of our previously reported CAD scheme

We retrospectively collected CTC cases acquired at The University of Chicago Hospitals. CTC examinations were performed on 73 patients whose colons were prepared by standard pre-colonoscopy cleansing and were insufflated with room air or carbon dioxide. Each patient was scanned in both supine and prone positions with either a single- or a multi-detector CT scanner (HiSpeed CTi or LightSpeed QX/i, GE Medical Systems, Milwaukee, WI) with collimations between 2.5 and 5.0 mm, reconstruction intervals of 1.25–5.0 mm, and tube currents of 60–120 mA with 120 kVp. Thus, our database contained 146 CTC data sets. Each reconstructed CT section had a matrix size of 512×512 pixels, with an in-plane pixel size of 0.5–0.7 mm. All patients also underwent optical colonoscopy on the same day as the CTC. Radiologists established the locations of polyps in the CTC data sets by use of the colonoscopy reports, pathology reports, and multiplanar reformatted views of the CTC on a viewing workstation (GE Advantage Windows Workstation v.4.2, GE Medical Systems, Milwaukee, WI). In this study, we used 5 mm as the lower limit on the size of polyps, which is considered to be clinically significant.³⁵ Fifteen patients had 28 colonoscopy-confirmed polyps, 15 of which were 5–9 mm in diameter, and 13 were 10–25 mm. We also created a training RT database by manual extraction of volumes containing RTs from “normal” (nonpolyp) CTC cases.

The CTC cases were subjected to our previously reported CAD scheme,^{36–38} which included centerline-based extraction of the colon,³⁹ shape-based detection of polyps,^{37,38} and reduction of FPs by use of a Bayesian neural network⁴⁰ based on geometric and texture features.^{41,42} We evaluated CAD detections on supine and prone CTC volumes independently. This CAD scheme yielded a 96.4% (27/28 polyps)

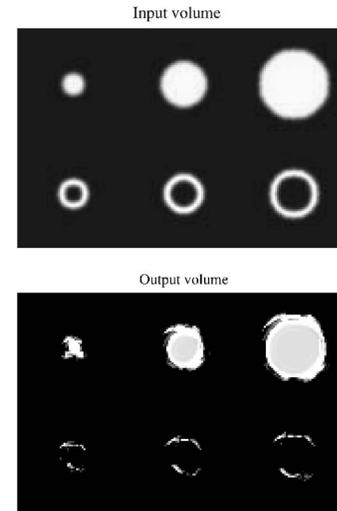


FIG. 4. Illustrations of polyp and RT phantoms of three different sizes and the corresponding output volumes of the 3D MTANN trained with actual polyps and RTs. In the output volumes, the simulated polyps are represented by bright voxels, whereas the simulated RTs appear as darker voxels.

by-polyp sensitivity with 3.1 (224/73) FPs per patient. Among these FPs, 20 FPs were caused by RTs. Forty-eight true-positive polyp detections in both supine and prone CTC volumes constituted 27 polyps. These CAD detections were used for training and evaluation of the 3D MTANN.

III. RESULTS

A. Training of a 3D MTANN

We manually selected 10 representative polyp volumes (10 actual polyps) from the 48 true-positive volumes (containing 27 actual polyps) in our CTC database and ten RTs from the training RT database (which was not used for testing) as the training cases for a 3D MTANN, as shown in Fig. 5. The selection was made by one of the authors (K.S.) based on the visual appearance of polyps and RTs in terms of size, shape, and contrast, so that these cases represent the database (ideally the population). For example, when there were two polyps similar to each other in terms of size and contrast, we selected only one of them because the other would be redundant. Ten cases were used for each category (polyp or RT), because the performance of a 3D MTANN was highest when the number of training cases was 20 (10 polyps and 10 RTs) in our experiment (see Sec. IV). A three-layer structure was employed for the 3D MTANN, because it has been proved theoretically that any continuous mapping can be realized approximately by a three-layer ANN.⁴³ The size of the training volume and the standard deviation of the 3D Gaussian distribution in the teaching volume were $13 \times 13 \times 13$ voxels and 4.5 voxels, respectively, which were empirically determined based on our previous studies.^{26,27,44} The number of hidden units was selected to be 25 by use of a method for designing the structure of an ANN.^{32,33} With the above-noted parameters, training of the 3D MTANN was performed 500 000 times and converged to a mean absolute error between teaching and output values of 0.073. To check the

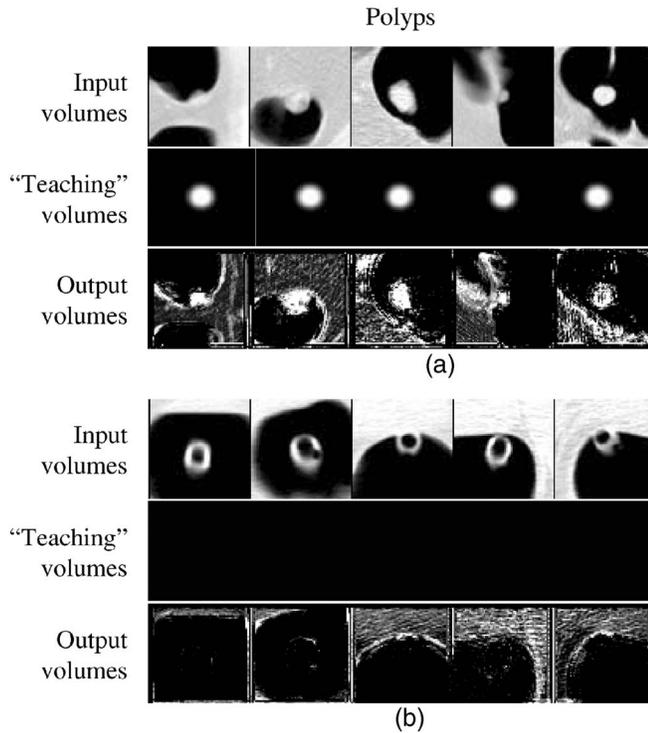


FIG. 5. Illustrations of training (a) polyps and (b) RTs. The central axial slices of the volumes are shown. Teaching volumes for polyps contain 3D Gaussian distributions, whereas those for RTs are zeros, i.e., they are completely dark. In the output volumes of the trained 3D MTANN, polyps are represented by bright voxels, whereas RTs are almost dark.

completion of the training, we applied the trained 3D MTANN to the training cases. In the output volumes shown in Fig. 5, training polyps are represented by distributions of bright voxels, whereas training RTs are almost dark, as expected. It should be noted that the output distribution for a small polyp (the farthest left image) is stronger and larger than the original polyp in the input volume, demonstrating the ability of the 3D MTANN to enhance small polyps.

B. Simulation experiments

We applied the 3D MTANN trained with actual polyps and RTs to the phantom volumes. The output volumes corresponding to the simulated polyps demonstrate bright voxels, whereas those corresponding to the simulated RTs appear mostly dark, but with some brighter segments on the upper and lower edges (Fig. 4). The scores for various-sized simulated polyps together with those for simulated RTs are shown in Fig. 6. The performance of the 3D MTANN decreased for simulated polyps less than 10 mm. The curve for polyps exceeds the maximum score for RTs at a polyp size of 4.5 mm, indicating that simulated polyps larger than 4.5 mm could be distinguished from RT phantoms by use of the 3D MTANN.

We investigated the effect of off-centering of the detected polyps on the performance of a 3D MTANN. Off-centering can always happen for irregular polyps or when an initial detection scheme fails to segment polyps. We used a simulated polyp 10 mm in diameter. The result indicates that the scores decrease when the position of the simulated polyp is

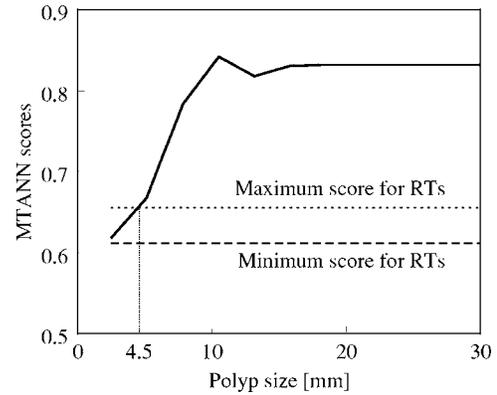


FIG. 6. Effects of simulated polyp sizes on the scores provided by the trained 3D MTANN. Also shown for reference are the minimum and maximum scores provided for the simulated RTs. Based on the scores, polyps larger than 4.5 mm can be distinguished from RTs by the 3D MTANN.

off-centered, as shown in Fig. 7. A detected polyp whose size is 10 mm can be off-centered by a range between 5 and 5 mm. The scores for the simulated polyps off-centered from -5 through 5 mm were above the maximum score for RTs. Therefore, the 3D MTANN was robust against off-centering in the distinction between polyps and RTs.

C. Experiments with actual polyps and RTs

We applied the trained 3D MTANN to 48 true-positive volumes (27 actual polyps) and 20 actual RTs (FPs) generated by our previously reported CAD scheme for detection of polyps. The output volumes for nontraining cases are shown in Fig. 8. The centers of the input volumes corresponded to the locations of polyp detections (containing true positives and FPs) provided by the CAD scheme; thus, this experiment included the effect of actual off-centering of polyp candidates produced by the initial detection scheme. Various polyps, including a flat lesion [fourth image from the left in the third row in Fig. 8(a)] and small sessile polyps [the farthest right images in the first and third rows in Fig. 8(a)], are represented in the output by distributions of bright voxels, whereas RTs appear as darker voxels, indicating the ability of

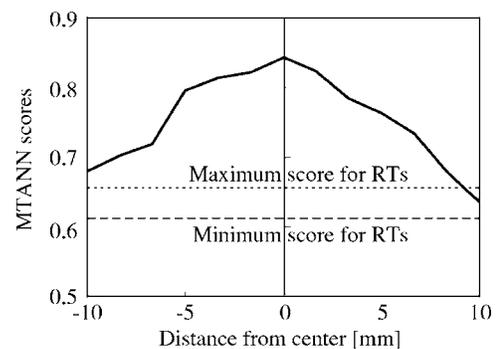


FIG. 7. Effect of off-centering of simulated polyps (10 mm in diameter) on the performance of the trained 3D MTANN. The scores for simulated polyps off-centered from -5 to 5 mm are above the maximum score for RTs, indicating the robustness of the 3D MTANN against off-centering in the distinction between polyps and RTs.

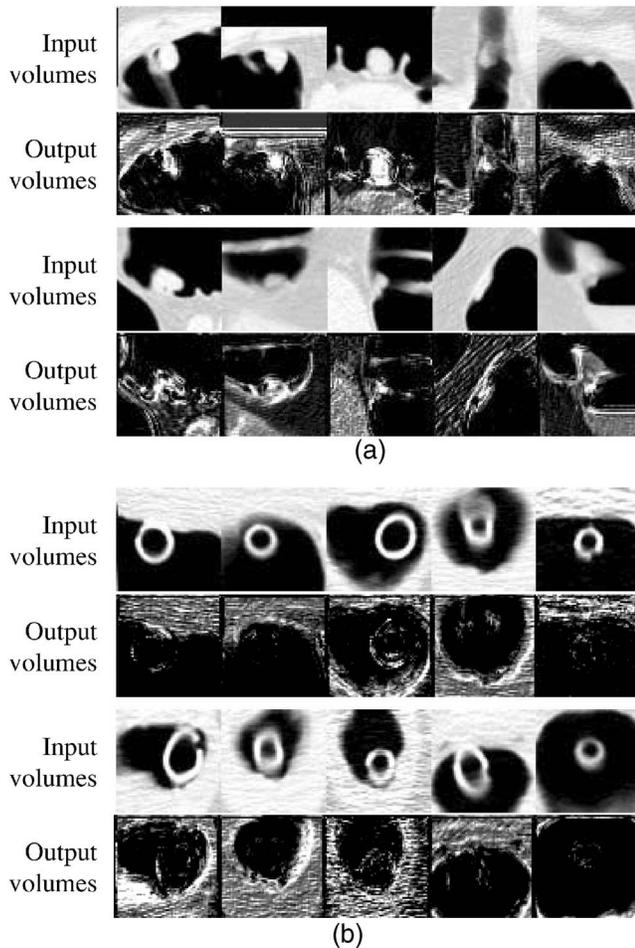


Fig. 8. Illustrations of (a) various nontraining polyps and the corresponding output volumes of the trained 3D MTANN and (b) nontraining RTs and the corresponding output volumes. In the output volumes, polyps appear as distributions of bright voxels, whereas RTs appear as dark voxels.

the 3D MTANN to enhance actual polyps and suppress actual RTs. We applied the 3D scoring method to the output volumes for polyps and RTs. The same standard deviation of the 3D Gaussian weighting function as that for the 3D Gaussian distribution in the polyp teaching volume was used. Distributions of scores for 38 nontraining true-positive volumes (10 training polyp volumes were excluded from the 48 true-positive volumes) and 20 actual RTs in an independent test are shown in Fig. 9. No polyp score overlaps with a RT score, indicating that the 3D MTANN achieved a specificity of 100%; in other words, the 3D MTANN was able to eliminate all RTs without removal of any true positives.

D. Evaluation of the performance for false-positive reduction

We applied the trained 3D MTANN to all 224 nontraining actual nonpolyps (FPs) produced by our original CAD scheme. The distributions of the scores are shown in Fig. 10. Although the two distributions overlap, a substantial fraction of nonpolyps can be eliminated by use of the 3D MTANN. We evaluated the overall performance of the 3D MTANN for

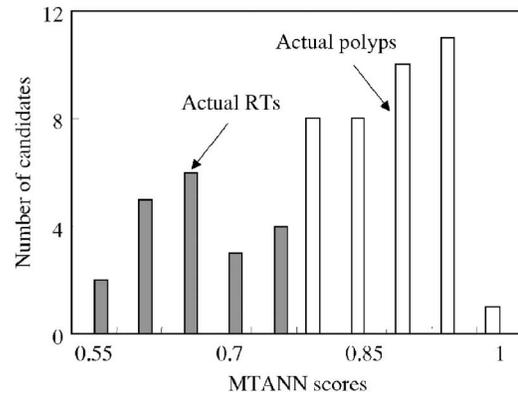


Fig. 9. Distributions of the independent test scores of 38 nontraining true-positive volumes and 20 nontraining actual RTs that were detected as FPs by our previously reported CAD scheme for detection of polyps.

FP reduction by use of free-response receiver-operating-characteristic (FROC) analysis.⁴⁵ The FROC curve of the trained 3D MTANN is shown in Fig. 11. The FROC curve indicates that the 3D MTANN was able to eliminate 33% (73/224) of all nonpolyps (FPs) without removal of any of 48 true-positive volumes, i.e., a 96.4% (27/28) overall by-polyp sensitivity was achieved at an FP rate of 2.1 (151/73) per patient.

Because the 48 true-positive volumes included the 10 training polyps, this evaluation could be biased. In order to reduce this bias, we excluded the 10 training polyps from the evaluation of the 3D MTANN. The FROC curve for only nontraining cases is shown in Fig. 11. The FROC curves indicate that the performance of the 3D MTANN for nontraining cases only is the same as that for the entire database when the number of FPs per patient is greater than 2, e.g., classification sensitivity of 100% (38/38) was achieved at a FP rate of 2.1 (151/73) per patient under both conditions. Among 73 FPs removed by the 3D MTANN, 20 were caused by RTs. The remaining 53 FPs included low-contrast small pieces of stool, tips of folds, and thin folds. The small stool and folds were removed probably because the MTANN tended to provide relatively low scores for small objects, as

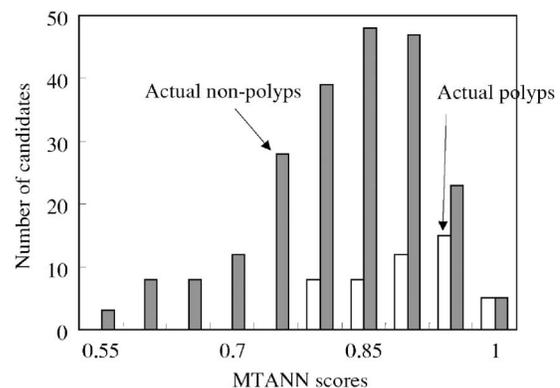


Fig. 10. Distributions of the scores for 48 true-positive volumes, containing 27 actual polyps, and 224 nonpolyps (FPs) generated by our previously reported CAD scheme for detection of polyps.

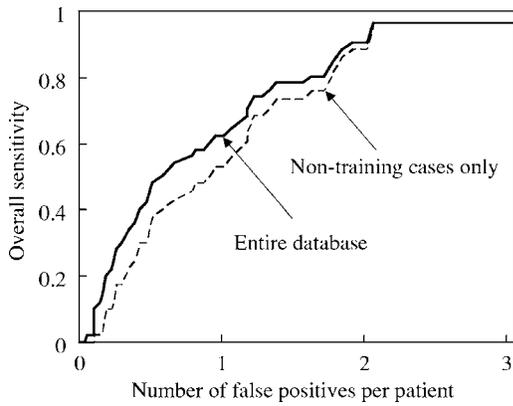


FIG. 11. The solid curve is a FROC curve that shows the overall performance of the trained 3D MTANN when it was applied to the entire database of 27 polyps (48 true-positive volumes) and 224 nonpolyps (FPs), and the dotted curve shows the performance of the 3D MTANN when it was tested on the nontraining-case-only database from which the 10 training polyps were excluded. These FROC curves indicate that the 3D MTANN yielded a reduction of 33% (73/224) of nonpolyps (FPs) without removal of any true positives, i.e., it achieved 100% (48/48) classification sensitivity.

shown in the phantom experiment illustrated in Fig. 6. Thin folds were removed probably because such a fold resembles a part of a RT. Typical examples of FPs that were not removed by the 3D MTANN are shown in Fig. 12; they include a bulbous-shaped fold, stool with bubbles, high-contrast stool, and the ileocecal valve.

IV. DISCUSSION

A limitation of this study is the use of a limited number CTC cases with polyps. Use of a larger database will provide more reliable evaluation results on the performance of the 3D MTANN. However, it should be noted that, although the 3D MTANN was trained with only 10 polyps, the performance for 27 polyps including the 10 polyps and that for only nontraining 17 polyps were very similar, especially at higher sensitivity levels; this is a good indication of the robustness of the 3D MTANN. This observation of the generalization ability of the 3D MTANN is consistent with that of MTANNs in our previous study.⁴⁴ Therefore, we expect that the performance of the 3D MTANN for a large database would potentially be comparable to that obtained in this study.

We investigated the effect of the intra-observer variation in selecting training cases on the performance of a 3D

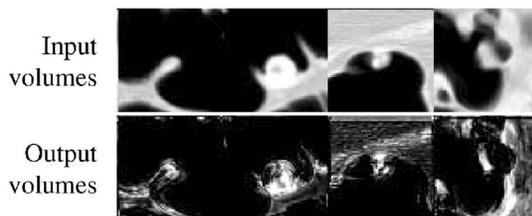


FIG. 12. Typical examples of sources of the remaining FPs that were not removed by the 3D MTANN. From left to right, a bulbous-shaped fold, high-contrast stool, stool with bubbles, and the ileocecal valve.

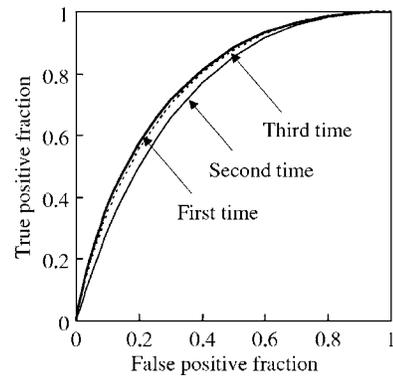


FIG. 13. Effect of the intra-observer variation in selecting training cases on the performance of a 3D MTANN. Three training sets were selected by an observer with at least a one-month interval (80% of cases overlapped between the sets selected at the first and second time; 50% percent between the sets selected at the first and third time). ROC curves indicate that the performances of the 3D MTANNs trained with three different training sets were slightly different (A_z values of 0.78, 0.75, and 0.77 for the first, second, and third times, respectively). The difference between A_z values with the sets selected at the first and second times was statistically significant, but those in other combinations were not.

MTANN, because the performance would depend on the manual selection of training cases. The same person who selected the training cases (the first time) described in Sec. III selected two more sets of training cases (the second and third times) in the same way as described in Sec. III A. To reduce a bias, each of the three selections was made at least one month apart. As a result, 80% of cases overlapped between the sets selected at the first and second times; 50% between the sets selected at the first and third times; and 40% between the sets selected at the second and third times. The performance of the trained 3D MTANNs with the three different sets was evaluated by receiver-operating-characteristic (ROC) analysis.^{46,47} The performance of the three 3D MTANNs for nontraining cases was slightly different, as shown in Fig. 13. The A_z values (area under the ROC curve)⁴⁸ for the 3D MTANNs trained with the sets selected the first, second, and third times were 0.78, 0.75, and 0.77, respectively. The difference between the A_z values with the sets selected at the first and second times was statistically significant (two-tailed p -value < 0.05),⁴⁹ but those for other combinations were not.

We conducted an observer study to investigate the difference in the performance of 3D MTANNs when different observers selected training cases. Two observers (medical imaging scientists who had no experience with CT colonography) participated in the observer study. They (observers B and C) were asked to select training cases from our databases. The cases selected by observer A (K.S.) overlapped with 35% and 60% of the cases selected by observers B and C, respectively. The performances of 3D MTANNs trained with the cases selected by the three different observers for nontraining cases are slightly different, as shown in Fig. 14. A_z values of 0.78, 0.74, and 0.73 were obtained for observers A, B, and C, respectively. The differences between A_z values in any combinations were not statistically significant.

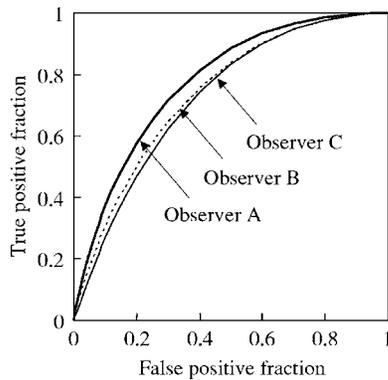


FIG. 14. Effect of the inter-observer variation in selecting training cases on the performance of a 3D MTANN. ROC curves indicate that the performance of the 3D MTANNs trained with the cases selected by three different observers was slightly different (A_z values of 0.78, 0.74, and 0.73 for observers A, B, and C, respectively). The differences among A_z values were not statistically significant.

cant (two-tailed p -value >0.05).⁴⁹ Thus, the performance of the 3D MTANN depended on the observer who selected the training cases, although it did not have a statistically significant difference.

We investigated the effect of random selection of training cases on the performance of a 3D MTANN, which may be considered as the worst-case scenario because random selection is a “blind” sampling without any knowledge. Five sets of training cases were selected randomly from our databases. The performance of the 3D MTANNs trained with five different training sets for nontraining cases is shown in Fig. 15. The performance (A_z values) ranged from 0.58 to 0.73 across the five different 3D MTANNs. The average performance (A_z value of 0.75) of the 3D MTANNs trained with the sets selected by the three observers was different by 0.07 from that (A_z value of 0.67) of the 3D MTANNs trained with

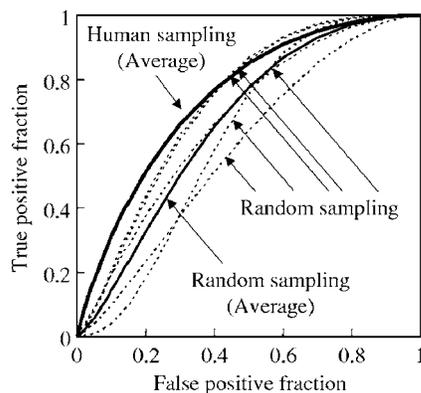


FIG. 15. Effect of random sampling of training cases on the performance of a 3D MTANN. Five ROC curves drawn by dashed lines represent the performance of the 3D MTANNs trained with five different training sets sampled randomly. The performance ranged from an A_z value of 0.58 to 0.73 (average of 0.67 with a standard deviation of 0.063) across the five different 3D MTANNs. The average performance (A_z value of 0.75) of the 3D MTANNs trained with three training sets selected by different observers was higher than that of the 3D MTANNs trained with randomly selected cases by 0.07 at a statistically significant level.

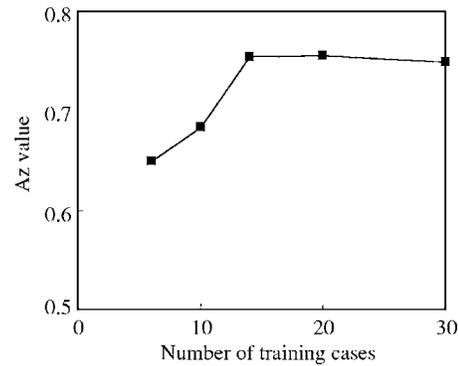


FIG. 16. Effect of change in the number of training cases on the performance of a 3D MTANN. There is little increase in the A_z value when the number of training cases is greater than 14 (7 polyps and 7 RTs). The A_z value was highest when 20 training cases were used.

randomly selected cases at a statistically significant level (two-tailed p -value <0.05) by Student’s t -test. This result was consistent with our previous study⁵⁰ of the distinction between nodules and non-nodules in thoracic CT with a MTANN: the performance (A_z value) of the MTANN was lower by 0.06–0.09 when training cases were selected randomly, compared to a human selection (the difference between the performance of the MTANNs based on random selection and human selection was statistically significant).

Imperfect selection of training cases degrades the performance of a 3D MTANN, as indicated by the three experiments mentioned earlier. The performance could drop by A_z values of 0.04–0.05 when different observers selected training cases. The worst-case scenario for selection would be random selection without any knowledge. The performance could drop by A_z values of 0.02–0.17 compared to the human selection. These results suggest that training cases should be selected by an informed human observer.

We investigated the effect of the number of training cases on the performance of a 3D MTANN. Training sets containing different numbers of cases were selected from our databases so that a set with a smaller number of training cases was a subset of a set with a larger number of training cases. The performance of 3D MTANNs trained with different numbers of training cases (from 6 to 30) is shown in Fig. 16. To reduce a bias, we excluded the 30 training cases from the evaluation. There was little increase in the A_z value when the number of training cases was greater than 14 (7 polyps and 7 RTs). The difference between an A_z value obtained with 6 training cases and that obtained with 20 training cases was statistically significant (two-tailed p -value <0.05),⁴⁹ but those for other combinations were not. The 3D MTANN trained with 20 cases had the highest A_z value. This was the reason for using 20 training cases. This tendency was consistent with that in our previous studies with MTANNs in the distinction between nodules and non-nodules in thoracic CT.^{26–29,44,50}

Iordanescu *et al.* developed an image-segmentation-based approach for reduction of FPs due to RTs,¹⁹ which finds the hole of the RT by matched filtering, extrapolates the axis,

and segments the RT by a conditional morphological dilation method. The FP-reduction technique developed in this study was able to track and label 72% of RTs successfully, and to reduce FPs by 9.2%. In constant, 3D MTANN uses a voxel-based classification approach for the same reduction task. The 3D MTANN was able to remove 100% of RTs, which corresponds to 8.9% (20/224) of all FPs. Therefore, the performance of the two methods is comparable.

A potential advantage of the 3D MTANN might be the reduction of FPs other than RTs; the 3D MTANN was able to remove 24% (53/224) of other types of FPs. A reason for this removal might be that some parts of RTs could be similar to parts of other types of FPs. It should be noted that, in spite of a large number of studies, accurate segmentation is still difficult for complicated patterns; thus, incorrect segmentation can occur for those patterns; for example, a region-growing technique may fail to segment a RT because of its holes or blockages inside the RT. The segmentation-based approach may also fail with RTs that are not compatible with the assumed model of the RT, such as changes in the diameter, shape, wall thickness of the RT, and twisting of a flexible RT, whereas the 3D MTANN approach does not require segmentation, but only the location of a potential RT. Further investigations are needed for determining the advantages and disadvantages of these two approaches.

V. CONCLUSION

We developed a 3D MTANN for reduction of FPs due to RTs in a CAD scheme for detection of polyps in CTC. With the 3D MTANN, we were able to eliminate the FPs due to RTs without removal of any true positives. Thus, our 3D MTANN could be useful for improving the specificity of a CAD scheme.

ACKNOWLEDGMENTS

The authors are grateful to S. G. Armato III, Ph.D., M. L. Giger, Ph.D., K. Doi, Ph.D., S. Kasai, Ph.D., and C. Muramatsu, for their valuable suggestions, Charles E. Metz, Ph.D., for the use of the ROCKIT software, and to E. F. Lanzl for improving the manuscript. This work was supported by Cancer Research Foundation Young Investigator Award.

^{a)}Electronic mail: suzuki@uchicago.edu

¹M. Macari and E. J. Bini, "CT colonography: Where have we been and where are we going?" *Radiology* **237**(3), 819–833 (2005).

²J. G. Fletcher, F. Booya, C. D. Johnson *et al.*, "CT colonography: Unraveling the twists and turns," *Curr. Opin. Gastroenterol.* **21**(1), 90–98 (2005).

³H. Yoshida and A. H. Dachman, "Computer-aided diagnosis for CT colonography," *Semin Ultrasound CT MR* **25**(5), 419–431 (2004).

⁴H. Yoshida and A. H. Dachman, "CAD techniques, challenges, and controversies in computed tomographic colonography," *Abdom. Imaging* **30**(1), 26–41 (2005).

⁵H. Yoshida, Y. Masutani, P. MacEneaney *et al.*, "Computerized detection of colonic polyps at CT colonography on the basis of volumetric features: pilot study," *Radiology* **222**(2), 327–336 (2002).

⁶H. Yoshida, J. Näppi, P. MacEneaney *et al.*, "Computer-aided diagnosis scheme for detection of polyps at CT colonography," *Radiographics* **22**(4), 963–979 (2002).

⁷H. Yoshida and J. Näppi, "Three-dimensional computer-aided diagnosis scheme for detection of colonic polyps," *IEEE Trans. Med. Imaging*

20(12), 1261–1274 (2001).

⁸R. M. Summers, C. D. Johnson, L. M. Pusanik *et al.*, "Automated polyp detection at CT colonography: Feasibility assessment in a human population," *Radiology* **219**(1), 51–59 (2001).

⁹D. S. Paik, C. F. Beaulieu, G. D. Rubin *et al.*, "Surface normal overlap: A computer-aided detection algorithm with application to colonic polyps and lung nodules in helical CT," *IEEE Trans. Med. Imaging* **23**(6), 661–675 (2004).

¹⁰G. Kiss, J. Van Cleynenbreugel, M. Thomeer *et al.*, "Computer-aided diagnosis in virtual colonography via combination of surface normal and sphere fitting methods," *Eur. Radiol.* **12**(1), 77–81 (2002).

¹¹J. Näppi, H. Frimmel, A. H. Dachman *et al.*, "A new high-performance CAD scheme for the detection of polyps in CT colonography," *Medical Imaging: Image Processing* **5370**, 839–848 (2004).

¹²R. M. Summers, J. Yao, P. J. Pickhardt *et al.*, "Computed tomographic virtual colonoscopy computer-aided polyp detection in a screening population," *Gastroenterology* **129**(6), 1832–1844 (2005).

¹³S. B. Gokturk, C. Tomasi, B. Acar *et al.*, "A statistical 3-D pattern processing method for computer-aided detection of polyps in CT colonography," *IEEE Trans. Med. Imaging* **20**(12), 1251–1260 (2001).

¹⁴J. Näppi and H. Yoshida, "Automated detection of polyps with CT colonography: Evaluation of volumetric features for reduction of false-positive findings," *Acad. Radiol.* **9**(4), 386–397 (2002).

¹⁵B. Acar, C. F. Beaulieu, S. B. Gokturk *et al.*, "Edge displacement field-based classification for improved detection of polyps in CT colonography," *IEEE Trans. Med. Imaging* **21**(12), 1461–1467 (2002).

¹⁶A. K. Jerebko, R. M. Summers, J. D. Malley *et al.*, "Computer-assisted detection of colonic polyps with CT colonography using neural networks and binary classification trees," *Med. Phys.* **30**(1), 52–60 (2003).

¹⁷A. K. Jerebko, J. D. Malley, M. Franaszek *et al.*, "Multiple neural network classification scheme for detection of colonic polyps in CT colonography data sets," *Acad. Radiol.* **10**(2), 154–160 (2003).

¹⁸J. Näppi, A. Okamura, H. Frimmel *et al.*, "Region-based supine-prone correspondence for the reduction of false-positive CAD polyp candidates in CT colonography," *Acad. Radiol.* **12**(6), 695–707 (2005).

¹⁹G. Iordanescu and R. M. Summers, "Reduction of false positives on the rectal tube in computer-aided detection for CT colonography," *Med. Phys.* **31**(10), 2855–2862 (2004).

²⁰R. M. Summers, J. Yao, and C. D. Johnson, "CT colonography with computer-aided detection: Automated recognition of ileocecal valve to reduce number of false-positive detections," *Radiology* **233**(1), 266–272 (2004).

²¹Z. Wang, Z. Liang, L. Li *et al.*, "Reduction of false positives by internal features for polyp detection in CT-based virtual colonoscopy," *Med. Phys.* **32**(12), 3602–3616 (2005).

²²K. Suzuki, I. Horiba, and N. Sugie, "Efficient approximation of neural filters for removing quantum noise from images," *IEEE Trans. Signal Process.* **50**(7), 1787–1799 (2002).

²³K. Suzuki, I. Horiba, and N. Sugie, "Neural edge enhancer for supervised edge enhancement from noisy images," *IEEE Trans. Pattern Anal. Mach. Intell.* **25**(12), 1582–1596 (2003).

²⁴K. Suzuki, I. Horiba, N. Sugie *et al.*, "Neural filter with selection of input features and its application to image quality improvement of medical image sequences," *IEICE Trans. Inf. Syst.* **E85-D**(10), 1710–1718 (2002).

²⁵K. Suzuki, I. Horiba, N. Sugie *et al.*, "Extraction of left ventricular contours from left ventriculograms by means of a neural edge detector," *IEEE Trans. Med. Imaging* **23**(3), 330–339 (2004).

²⁶K. Suzuki, S. G. Armato, F. Li *et al.*, "Massive training artificial neural network (MTANN) for reduction of false positives in computerized detection of lung nodules in low-dose CT," *Med. Phys.* **30**(7), 1602–1617 (2003).

²⁷H. Arimura, S. Katsuragawa, K. Suzuki *et al.*, "Computerized scheme for automated detection of lung nodules in low-dose computed tomography images for lung cancer screening," *Acad. Radiol.* **11**(6), 617–629 (2004).

²⁸K. Suzuki, J. Shiraishi, H. Abe *et al.*, "False-positive reduction in computer-aided diagnostic scheme for detecting nodules in chest radiographs by means of massive training artificial neural network," *Acad. Radiol.* **12**(2), 191–201 (2005).

²⁹K. Suzuki, F. Li, S. Sone *et al.*, "Computer-aided diagnostic scheme for distinction between benign and malignant nodules in thoracic low-dose CT by use of massive training artificial neural network," *IEEE Trans. Med. Imaging* **24**(9), 1138–1150 (2005).

- ³⁰K. Suzuki, H. Abe, H. MacMahon *et al.*, "Image-processing technique for suppressing ribs in chest radiographs by means of massive training artificial neural network (MTANN)," *IEEE Trans. Med. Imaging* **25**(4), 406–416 (2006).
- ³¹K. Suzuki, I. Horiba, K. Ikegaya *et al.*, "Recognition of coronary arterial stenosis using neural network on DSA system," *Syst. Comput. Japan* **26**(8), 66–74 (1995).
- ³²K. Suzuki, I. Horiba, and N. Sugie, "A simple neural network pruning algorithm with application to filter synthesis," *Neural Process. Lett.* **13**(1), 43–53 (2001).
- ³³K. Suzuki, "Determining the receptive field of a neural filter," *J. Neural Eng.* **1**(4), 228–237 (2004).
- ³⁴D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature (London)* **323**, 533–536 (1986).
- ³⁵C. D. Johnson and A. H. Dachman, "CT colonography: The next colon screening examination?" *Radiology* **216**(2), 331–341 (2000).
- ³⁶H. Yoshida, J. Nappi, P. MacEneaney *et al.*, "Computer-aided diagnosis scheme for detection of polyps at CT colonography," *Radiographics* **22**(4), 963–979 (2002).
- ³⁷H. Yoshida and J. Nappi, "Three-dimensional computer-aided diagnosis scheme for detection of colonic polyps," *IEEE Trans. Med. Imaging* **20**(12), 1261–1274 (2001).
- ³⁸H. Yoshida, Y. Masutani, P. MacEneaney *et al.*, "Computerized detection of colonic polyps at CT colonography on the basis of volumetric features: Pilot study," *Radiology* **222**(2), 327–336 (2002).
- ³⁹H. Frimmel, J. Nappi, and H. Yoshida, "Fast and robust computation of colon centerline in CT colonography," *Med. Phys.* **31**(11), 3046–3056 (2004).
- ⁴⁰M. A. Kupinski, D. C. Edwards, M. L. Giger *et al.*, "Ideal observer approximation using Bayesian classification neural networks," *IEEE Trans. Med. Imaging* **20**(9), 886–899 (2001).
- ⁴¹J. Nappi and H. Yoshida, "Automated detection of polyps with CT colonography: Evaluation of volumetric features for reduction of false-positive findings," *Acad. Radiol.* **9**(4), 386–397 (2002).
- ⁴²J. Nappi and H. Yoshida, "Feature-guided analysis for reduction of false positives in CAD of polyps for computed tomographic colonography," *Med. Phys.* **30**(7), 1592–1601 (2003).
- ⁴³K. Funahashi, "On the approximate realization of continuous mappings by neural networks," *Neural Networks* **2**, 183–192 (1989).
- ⁴⁴K. Suzuki and K. Doi, "How can a massive training artificial neural network (MTANN) be trained with a small number of cases in the distinction between nodules and vessels in thoracic CT?," *Acad. Radiol.* **12**(10), 1333–1341 (2005).
- ⁴⁵D. P. Chakraborty and L. H. Winter, "Free-response methodology: Alternate analysis and a new observer-performance experiment," *Radiology* **174**(3 Pt 1), 873–881 (1990).
- ⁴⁶C. E. Metz, "ROC methodology in radiologic imaging," *Invest. Radiol.* **21**(9), 720–733 (1986).
- ⁴⁷C. E. Metz, B. A. Herman, and J. H. Shen, "Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data," *Stat. Med.* **17**(9), 1033–1053 (1998).
- ⁴⁸J. A. Hanley and B. J. McNeil, "A method of comparing the areas under receiver operating characteristic curves derived from the same cases," *Radiology* **148**(3), 839–843 (1983).
- ⁴⁹C. E. Metz, B. A. Herman, and C. A. Roe, "Statistical comparison of two ROC-curve estimates obtained from partially-paired data sets," *Med. Decis Making* **18**(1), 110–121 (1998).
- ⁵⁰K. Suzuki, S. G. Armato, F. Li *et al.*, "Effect of a small number of training cases on the performance of massive training artificial neural network (MTANN) for reduction of false positives in computerized detection of lung nodules in low-dose CT," *Proc. SPIE* **5032**, 1355–1366 (2003).