Mixture of expert 3D massive-training ANNs for reduction of multiple types of false positives in CAD for detection of polyps in CT colonography

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, 25 New Chardon Street, Suite 400C, Boston, Massachusetts 02114

Samuel G. Armato III and Abraham H. Dachman
Department of Radiology, The University of Chicago, 5841 South Maryland Avenue, Chicago, Illinois 60637

One of the major challenges in computer-aided detection (CAD) of polyps in CT colonography (CTC) is the reduction of false-positive detections (FPs) without a concomitant reduction in sensitivity. A large number of FPs is likely to confound the radiologist's task of image interpretation, lower the radiologist's efficiency, and cause radiologists to lose their confidence in CAD as a useful tool. Major sources of FPs generated by CAD schemes include haustral folds, residual stool, rectal tubes, the ileocecal valve, and extra-colonic structures such as the small bowel and stomach. Our purpose in this study was to develop a method for the removal of various types of FPs in CAD of polyps while maintaining a high sensitivity. To achieve this, we developed a "mixture of expert" three-dimensional (3D) massive-training artificial neural networks (MTANNs) consisting of four 3D MTANNs that were designed to differentiate between polyps and four categories of FPs: (1) rectal tubes, (2) stool with bubbles, (3) colonic walls with haustral folds, and (4) solid stool. Each expert 3D MTANN was trained with examples from a specific non-polyp category along with typical polyps. The four expert 3D MTANNs were combined with a mixing artificial neural network (ANN) such that different types of FPs could be removed. Our database consisted of 146 CTC datasets obtained from 73 patients whose colons were prepared by standard pre-colonoscopy cleansing. Each patient was scanned in both supine and prone positions. Radiologists established the locations of polyps through the use of optical-colonoscopy reports. Fifteen patients had 28 polyps, 15 of which were 5–9 mm and 13 were 10–25 mm in size. The CTC cases were subjected to our previously reported CAD method consisting of centerline-based extraction of the colon, shape-based detection of polyp candidates, and a Bayesian-ANN-based classification of polyps. The original CAD method yielded 96.4% (27/28) by-polyp sensitivity with an average of 3.1 (224/73) FPs per patient. The mixture of expert 3D MTANNs removed 63% (142/224) of the FPs without the loss of any true positive; thus, the FP rate of our CAD scheme was improved to 1.1 (82/73) FPs per patient while the original sensitivity was maintained. By use of the mixture of expert 3D MTANNs, the specificity of a CAD scheme for detection of polyps in CTC was substantially improved while a high sensitivity was maintained. © *2008 American Association of Physicists in Medicine*. [DOI: 10.1118/1.2829870]

Key words: virtual colonoscopy, colon cancer, mixture of expert networks, CT colonography, computer-aided diagnosis, false-positive reduction, artificial neural network

I. INTRODUCTION

Colorectal cancer is the second leading cause of cancer deaths in the United States.[1] Evidence suggests that early detection and removal of polyps (i.e., precursors of colorectal cancer) can reduce the incidence of colorectal cancer.[2,3] Computers tomography (CT) colonography (CTC), also known as virtual colonoscopy, is a technique for detecting colorectal neoplasms by use of a CT scan of the colon.[4] The diagnostic performance of CTC in detecting polyps, however, remains uncertain due to a propensity for perceptual errors.[5] Computer-aided detection (CAD) of polyps has been investigated to overcome these difficulties with CTC.[6,7] CAD has the potential to improve radiologists' diagnostic performance in the detection of polyps.[6,7]

Several investigators have developed automated or semi-automated CAD schemes for the detection of polyps in CTC.[8–15] Although current CAD schemes could be useful for the detection of polyps, some limitations remain. One of the major limitations with current CAD schemes is a relatively large number of false-positive detections (FPs), which could adversely affect the clinical application of CAD for colorectal cancer screening. A large number of FPs is likely to confound the radiologist's task of image interpretation and thus

lower radiologist efficiency. In addition, radiologists may lose their confidence in CAD as a useful tool. Therefore, it is important to reduce the number of FPs as much as possible while maintaining a high sensitivity.

Major sources of FPs generated by CAD schemes include haustral folds, residual stool, rectal tubes, the ileocecal valve, and extra-colonic structures such as the small bowel and stomach.[17] Although rectal tubes are relatively obvious FPs, radiologists may lose their confidence in CAD as an effective tool if the CAD scheme generates such obvious FPs. There-fore, removal of rectal-tube-induced FPs is desirable. To ad-dress this issue, we previously reported a three-dimensional (3D) massive-training artificial neural network (MTANN) for distinction between polyps and rectal tubes in 3D CTC volu-metric data.[16] The 3D MTANN eliminated all rectal-tube-induced FPs without removal of any true positives.

Various methods have been developed for the reduction of FPs. Gokturk *et al.*[17] developed a CAD scheme based on statistical pattern recognition, and they applied a 3D pattern-processing method for the reduction of FPs. Näppi *et al.*[18] developed a method for FP reduction based on volumetric features and another method[19] based on supine-prone corre-spondence. Acar *et al.*[20] used edge-displacement fields to model the changes in consecutive cross-sectional views of CTC data and quadratic discriminant analysis for FP reduc-tion. Jerebko *et al.*[21] used a standard ANN to classify polyp candidates in their CAD scheme and improved the perfor-mance by incorporating a committee of ANNs[22] and a com-mittee of support vector machines.[23] Iordanescu and Summers[24] developed an image-segmentation-based ap-proach for the reduction of FPs due to rectal tubes. Summers *et al.*[25] developed and analyzed a method for the reduction of FPs caused by the ileocecal valve. Wang *et al.*[26] developed a FP reduction method based on internal features of polyps.

Our previously reported CAD scheme[27,28] employs an FP reduction method based on a Bayesian ANN[29] with geomet-ric and texture features.[18,28] Our purpose in this study was to develop a "mixture of expert" 3D MTANNs for further re-duction of FPs in a polyp-detection CAD scheme while maintaining high sensitivity.

II. MATERIALS AND METHODS

II.A. CTC database

CTC examinations were performed on 73 patients at The University of Chicago Medical Center. The patients' colons were prepared by standard pre-colonoscopy cleansing with administration of cathartics following a water diet or low-fiber diet, and they were insufflated with room air or carbon dioxide. Each patient was scanned in both supine and prone positions. Our database thus contained 146 CTC datasets. The CT scans were performed with either a single- or a multi-detector-row CT scanner (HiSpeed CTi or LightSpeed QX/i, GE Medical Systems, Milwaukee, WI). The CT scan-ning parameters included collimations between 2.5 and 5.0 mm, reconstruction intervals of 1.0–5.0 mm [1.0 mm ($n=2$, 1% of the CTC datasets), 1.25 mm ($n=3$, 2%), 1.5 mm ($n=59$, 41%), 2.5 mm ($n=79$, 54%), and 5.0 mm

($n=3$, 2%)], and tube currents of 60–120 mA with 120 kVp. Each reconstructed CT section had a matrix size of 512 ×512 pixels, with an in-plane pixel size of 0.5–0.7 mm. The CT sections were interpolated to isotropic resolution by use of linear interpolation in the transverse direction. All patients underwent "reference-standard" optical colonos-copy. Radiologists established the locations of polyps in the CTC datasets by use of the colonoscopy and pathology re-ports, as well as multiplanar re-formatted views of the CTC on a viewing workstation (GE Advantage Windows Worksta-tion v.4.2, GE Medical Systems, Milwaukee, WI). In this study, we used 5 mm as the threshold for clinically signifi-cant polyps.[30] Fifteen patients had 28 polyps, 15 of which were 5–9 mm in diameter, and 13 were 10–25 mm. There was no polyp that was submerged in fluid. Fluid was mini-mized by use of a saline cathartic preparation as the standard preparation, not a colon gavage. We also created a training database of non-polyps by manual extraction of volumes containing non-polyps from 27 "normal" (non-polyp) CTC cases.

II.B. Architecture of a "mixture of expert" 3D massive-training artificial neural networks (3D MTANNs)

By extension of "neural filters"[31,32] and "neural edge enhancers,"[33,34] which are ANN based, supervised nonlinear image-processing techniques, two-dimensional (2D) MTANNs[35] have been developed to accommodate the task of distinguishing a specific opacity from other opacities in medical images. Two-dimensional (2D) MTANNs have been applied for the reduction of FPs in the computerized detec-tion of lung nodules in low-dose CT[35,36] and chest radiography,[37] for the distinction between benign and malig-nant lung nodules in CT,[38] and for the suppression of ribs in chest radiographs.[39] To process 3D volumetric CTC data, we developed a 3D MTANN[16] by extending the structure of the 2D MTANN.

A single MTANN cannot reduce multiple types of FP sources effectively, because the capability of a single MTANN is limited.[35] To reduce various types of FPs, we extended the capability of a single 3D MTANN and devel-oped a mixture of expert 3D MTANNs. The architecture of a mixture of expert 3D MTANNs is shown in Fig. 1. A mixture of expert 3D MTANNs consists of several 3D MTANNs that are arranged in parallel. Each expert 3D MTANN is trained independently by use of a specific type of non-polyp and a common set of actual polyps. Each expert 3D MTANN acts as an expert for distinguishing polyps from a specific type of non-polyp, e.g., 3D MTANN No. 1 is trained to distinguish polyps from rectal tubes; 3D MTANN No. 2 is trained to distinguish polyps from stool with bubbles; and so on.

Each expert 3D MTANN consists of a linear-output multilayer ANN model,[33] which is capable of operating on voxel data directly. The expert 3D MTANN is trained with input CTC volumes and the corresponding "teaching" vol-umes for enhancement of polyps and suppression of a spe-cific type of non-polyp. The pixel size within a CT section is
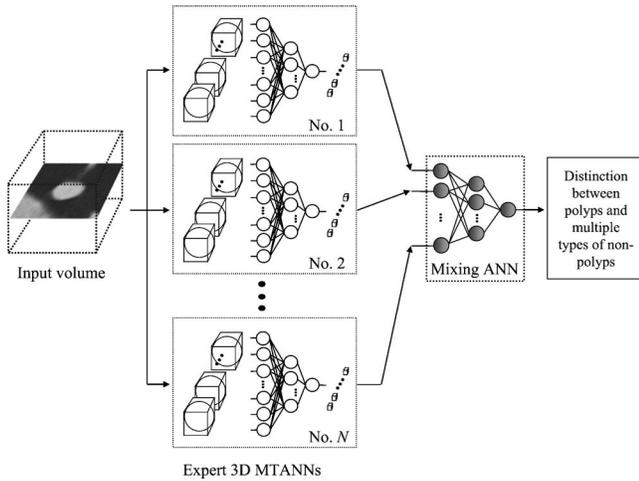
FIG. 1. Architecture of a "mixture of expert" 3D MTANNs for distinguishing polyps from various types of FPs. Each 3D MTANN acts as an expert for distinction between polyps and a specific type of non-polyp for which the expert 3D MTANN is trained, for example, expert 3D MTANN No. 1 is trained to distinguish polyps from rectal tubes, expert 3D MTANN No. 2 is designed for distinction between polyps and stool with bubbles, and so on. The outputs of individual expert 3D MTANNs are combined with a mixing ANN so that the mixture of expert 3D MTANNs can remove various types of non-polyps.

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 the CTC data, original CTC data are converted to isotropic volume data by use of a linear interpolation technique. The voxel values of the isotropic volumes are linearly scaled such that—1000 Hounsfield units (HU) corresponds to 0 and 1000 HU corresponds to 1 (values below 0 and above 1 are allowed). The input to the expert 3D MTANN consists of voxel values in a subvolume, V_S, extracted from an input volume. The output of the nth expert 3D MTANN is a continuous scalar value, which is associated with the center voxel in the subvolume, and is represented by

$$O_n(x,y,z) = NN_n\{I(x-i, y-j, z-k) | (i,j,k) \in V_S\}, \quad (1)$$

where x, y, and z are the coordinate indices, $NN_n(\cdot)$ is the output of the nth linear-output ANN model, and $I(x,y,z)$ is a voxel value of the input volume.

The training of a 2D MTANN for chest radiography[37,39] and thoracic CT[35,36,38] takes a substantially long time; consequently, efficient architecture and training are a necessity for a 3D MTANN. Because the average shape of polyps approximates a sphere and the operation for processing polyps should be isotropic, the shape of the subvolume input to an expert 3D MTANN should be spherical rather than cubic. By use of a spherical subvolume rather than a cubic one, the computational cost can be reduced to 52% ($\pi/6$) (i.e., by 48%). We employed a digital quasi-sphere as the input subvolume for an expert 3D MTANN.[16] The number of hidden units may be selected by use of a method for designing the structure of an ANN.[40,41] This method is a sensitivity-based pruning method, i.e., the sensitivity of the unit removal 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 re-training to recover the potential loss due to the removal are performed repeatedly, resulting in a reduced structure from which redundant units are removed.

II.C. Training method of expert 3D MTANNs

For enhancement of polyps and suppression of non-polyps in CTC volumes, the teaching volume contains a 3D Gaussian distribution with standard deviation σ_T. This distribution represents the "likelihood of being a polyp" for a polyp and zero for a non-polyp:

$$T(x,y,z)$$
$$= \begin{cases} \dfrac{1}{\sqrt{2\pi}\sigma_T} \exp\left\{ -\dfrac{(x^2+y^2+z^2)}{2\sigma_T^2} \right\} & \textit{for a polyp} \\ 0 & \textit{otherwise}. \end{cases}$$
$$(2)$$

A 3D Gaussian distribution is used to approximate an average shape of polyps. The expert 3D MTANN involves training with a large number of subvolume-voxel pairs; we call it a massive-subvolumes training scheme. To enrich the training samples, 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 expert 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; hence the term "massive-training ANN." The error to be minimized by training of the nth expert 3D MTANN is given by

$$E_n = \frac{1}{P_n} \sum_c \sum_{(x,y,z) \in V_{T_n}} \{T_{n,c}(x,y,z) - O_{n,c}(x,y,z)\}^2, \quad (3)$$

where c is a training case number, $O_{n,c}$ is the output of the nth expert MTANN for the cth case, $T_{n,c}$ is the teaching value for the nth expert MTANN for the cth case, and P_n is the number of total training voxels in the training volume for the nth expert 3D MTANN, V_{T_n}. The expert 3D MTANN is trained by a linear-output back-propagation (BP) algorithm.[33] After training, the expert 3D MTANN is expected to output the highest value when a polyp is located at the center of the subvolume of the expert 3D MTANN, a lower value as the distance from the subvolume center increases, and zero when the input subvolume contains a non-polyp.

II.D. 3D scoring method for combining output voxels

For combining output voxels from the trained expert 3D MTANNs, we developed a 3D scoring method, as shown in Fig. 2. A score for a given polyp candidate from the nth expert 3D MTANN is defined as
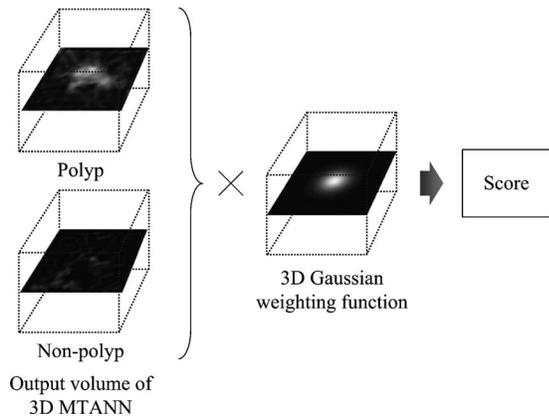
FIG. 2. Schematic illustration of a scoring method for combining output voxels from a trained expert 3D MTANN to obtain a case-based score.

$$S_n = \sum_{(x,y,z) \in V_E} f_G(\sigma_n; x, y, z) \times O_n(x, y, z), \qquad (4)$$

where

$$f_G(\sigma_n; x, y, z) = \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left\{ -\frac{(x^2 + y^2 + z^2)}{2\sigma_n^2} \right\} \qquad (5)$$

is a 3D Gaussian weighting function with standard deviation σ_n, and with its center corresponding to the center of the volume for evaluation, V_E; and $O_n(x, y, z)$ is the output volume of the nth trained expert 3D MTANN, where its center corresponds to the center of V_E. The use of the 3D Gaussian weighting function allows us to combine the responses (outputs) of a trained expert 3D MTANN as a 3D distribution. A 3D Gaussian function is used for scoring, because the output of a trained expert 3D MTANN is expected to be similar to the 3D Gaussian distribution used in the teaching volume. This score represents the weighted sum of the estimates for the likelihood that the volume (polyp candidate) contains a polyp near the center, i.e., a higher score would indicate a polyp, and a lower score would indicate a non-polyp.

II.E. Mixing ANN for combining expert 3D MTANNs

The scores from the expert 3D MTANNs are combined by use of a mixing ANN such that different types of non-polyps can be distinguished from polyps. The mixing ANN consists of a linear-output multilayer ANN model with a linear-output BP training algorithm[33] for processing of continuous output/teaching values; the activation functions of the units in the input, hidden, and output layers are an identity, a sigmoid, and a linear function, respectively. One unit is employed in the output layer for distinction between a polyp and a non-polyp. The scores of each expert 3D MTANN are used for each input unit in the mixing ANN; thus, the number of input units corresponds to the number of expert 3D MTANNs, N. The scores of each expert 3D MTANN act as the features for distinguishing polyps from a specific type of non-polyp for which the expert 3D MTANN is trained. The output of the mixing ANN for the cth polyp candidate is represented by

$$M_c = NN[\{S_{n,c}\} | 1 \le n \le N], \qquad (6)$$

where $NN(\cdot)$ is the output of the linear-output ANN model. The teaching values for polyps are assigned the value one, and those for non-polyps are zero. Training of the mixing ANN may be performed by use of a leave-one-lesion-out cross-validation scheme.[42] After training, the mixing ANN is expected to output a higher value for a polyp and a lower value for a non-polyp. Thus, the output can be considered to be a "likelihood of being a polyp." By thresholding the output, a distinction between polyps and non-polyps can be made. The balance between true-positive rate and FP rate is determined by the selected threshold value. If the scores of each expert 3D MTANN properly characterize the specific type of non-polyp for which the expert 3D MTANN is trained, the mixing ANN combining several expert 3D MTANNs will be able to distinguish polyps from various types of non-polyps.

III. RESULTS

III.A. Performance of our previously reported CAD scheme

We applied our previously reported CAD scheme[27,28] to the 73 CTC cases. The scheme included centerline-based extraction of the colon,[43] shape-based detection of polyps,[8,10] and initial reduction of FPs by use of a Bayesian ANN[29] based on geometric and texture features.[18,28] We evaluated supine and prone CTC volumes independently. This CAD scheme achieved a 96.4% (27/28 polyps) by-polyp sensitivity with an average of 3.1 (224/73) FPs per patient. Forty-eight true-positive polyp detections in both supine and prone CTC volumes represented 27 polyps. We combined our previously reported CAD scheme with the mixture of expert 3D MTANNs for further reduction of FPs, as shown in Fig. 3.

III.B. Training of expert 3D MTANNs

We manually selected ten representative polyp volumes (ten polyps) from the 48 true-positive volumes (containing 27 polyps) in our CTC database as the training polyp cases for expert 3D MTANNs. We classified CAD-generated FP sources into eight categories, i.e., rectal tubes, small bulbous folds, solid stool, stool with bubbles, colonic walls with haustral folds, elongated folds, strip-shaped folds, and the ileocecal valve. We manually selected ten non-polyps in each of the eight categories from the training non-polyp database (which was not used for testing). The above training volume selections were made by one of the authors (K.S.) based on the visual appearance of polyps and non-polyps in terms of size, shape, and contrast.[16] The ten polyps and the ten rectal tubes were the same as those used in our previous study.[16] The number of sample volumes for each category was ten, because the performance of an expert 3D MTANN was found to be highest when the number of training sample volumes was 20 (i.e., ten polyps and ten non-polyps) in our previous study,[16] and the performance of 2D/3D MTANNs
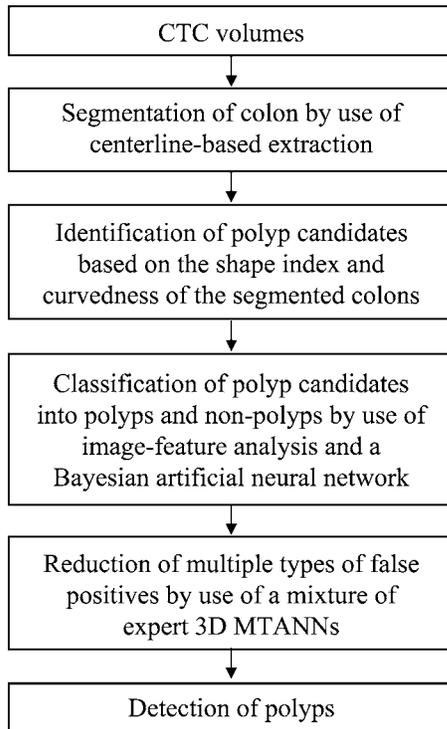
FIG. 3. Schematic diagram of CAD of polyps in CTC based on our previously reported CAD scheme with the addition of a mixture of expert 3D MTANNs for further reduction of FPs.

was not sensitive to the number of sample regions/volumes over different types of non-lesions in our previous studies.[16,35,38,44,45]

We trained eight expert 3D MTANNs with the ten polyps and ten non-polyps in each category. A three-layer structure was employed for the expert 3D MTANNs.[46] The size of the training volume and the standard deviation of the 3D Gaussian distribution in the teaching volume were $15 \times 15 \times 15$ voxels (i.e., cubic shape) and 4.5 voxels, respectively, which were determined empirically based on our previous studies.[16,35,36,44] The number of hidden units was selected to be 25 by use of a method for designing the structure of an ANN.[40,41] With the parameters above, training of the expert 3D MTANNs was performed by 500 000 iterations. We selected four among the eight expert 3D MTANNs for the mixture of expert 3D MTANNs by experimental analysis, because the mixture of these four expert 3D MTANNs [(1) rectal tubes, (2) stool with bubbles, (3) colonic walls with haustral folds, and (4) solid stool] demonstrated the highest performance (described in the next subsection). Figure 4 shows all ten training polyps and samples of non-polyps.

III.C. Evaluation of the performance for false-positive reduction

We applied the trained expert 3D MTANNs to the 27 polyps (48 true-positive volumes) and all 224 non-training FPs identified by our previously reported CAD scheme. The output volumes for these testing cases are shown in Fig. 5. The centers of the input volumes corresponded to the detection results provided by the CAD scheme (including both true positives and FPs); thus, this experiment included the effect of actual off-centering of polyp candidates produced by the initial CAD scheme. Various polyps, including a flat
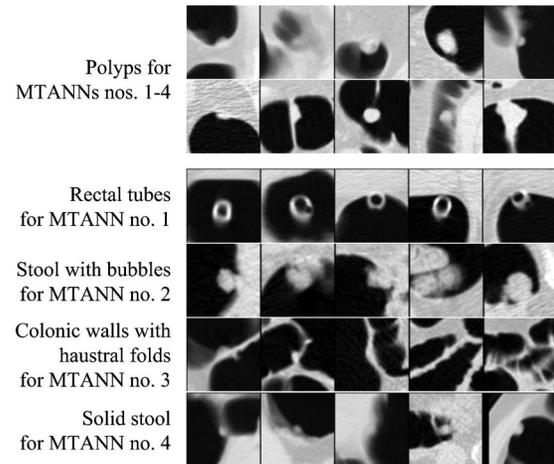
FIG. 4. Illustrations of all ten actual polyps used for training of all the expert 3D MTANNS and five of the non-polyps in each of four categories used for training four different expert 3D MTANNs. The central axial slices of the volumes are shown.
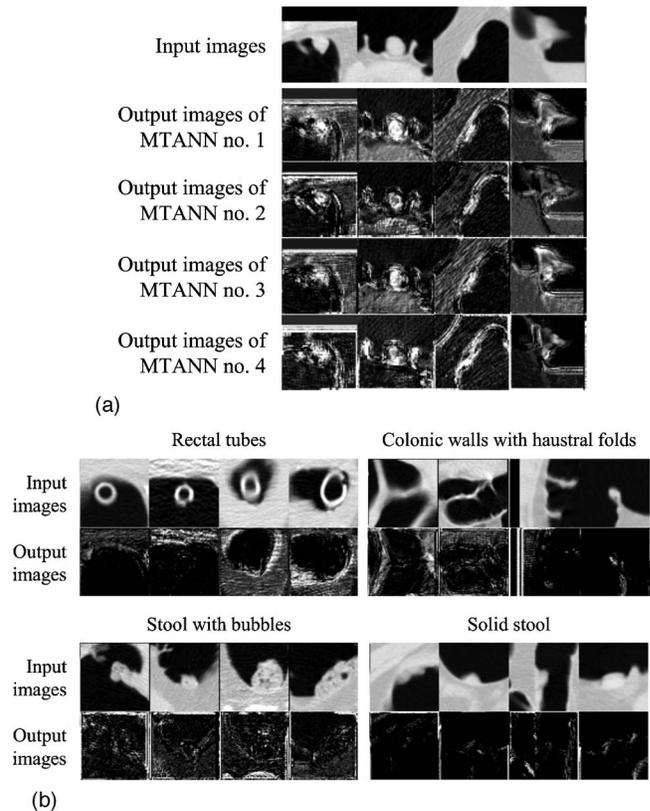
FIG. 5. Illustrations of (a) various testing polyps and the corresponding output volumes of four trained expert 3D MTANNs and (b) four different categories of testing FPs and the output volumes from corresponding expert 3D MTANNs. In the output volumes, polyps appear as distributions of bright voxels (i.e., they are enhanced), whereas different types of FPs appear as dark voxels (i.e., they are suppressed).
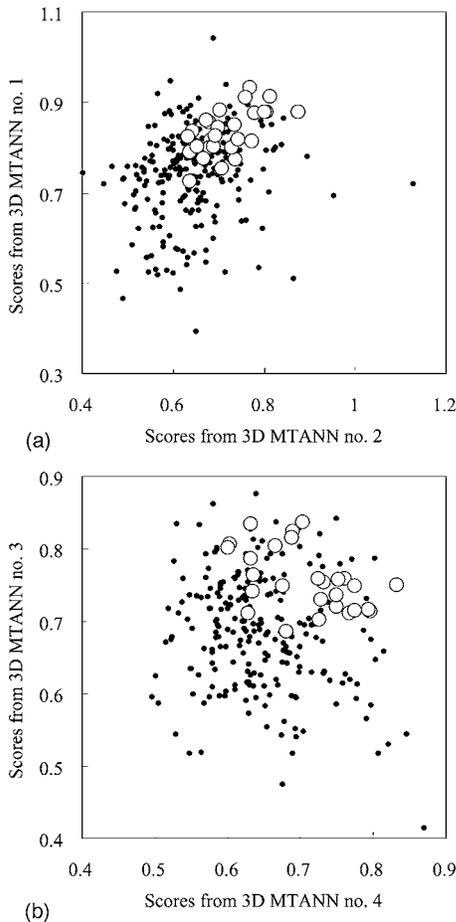
FIG. 6. Distributions of the four 3D MTANN scores for 27 polyps (open circles) and 224 FPs generated by our previously reported CAD scheme for the detection of polyps.

FIG. 7. Distributions of the output values from the trained mixing ANN for 27 polyps and 224 FPs generated by our previously reported CAD scheme in a leave-one-lesion-out cross-validation test.

lesion [the third image from the left in Fig. 5(a)], are represented in the output by distributions of bright voxels, whereas various types of non-polyps appear as darker voxels, indicating the ability of the expert 3D MTANNs to enhance polyps and suppress different types of non-polyps. The flat lesion was enhanced by the 3D MTANN, because the 3D MTANN approach is basically voxel-based determination of likelihood of being a polyp, which is less shape dependent compared to a morphologic-feature-based approach. We applied the 3D scoring method to the output volumes for polyps and non-polyps. The 3D Gaussian weighting function used the same standard deviation as that for the 3D Gaussian distribution in the polyp teaching volume, i.e., $\sigma_{n,} = \sigma_{T}$. Distributions of scores from the expert 3D MTANNs for the 27 polyps and all FPs are shown in Fig. 6. Although the two distributions in each graph overlap, a substantial fraction of FPs can be eliminated by use of the expert 3D MTANNs.

To distinguish between polyps and FPs, we merged the scores from the four individual expert 3D MTANNs with a mixing ANN. Figure 7 shows the distributions of the output values of the trained mixing ANN in a leave-one-lesion-out cross-validation test.[42] In the leave-one-lesion-out cross-validation test, the two instances of a polyp in supine and prone scans are left out simultaneously for testing in order to

remove the bias associated with the use of multiple instances from the same lesion for training. Thus, a training set and a testing set were completely separated in each run of the leave-one-lesion-out cross-validation test. Although the two distributions overlap, all polyps can be distinguished from the majority of non-polyps. We evaluated the overall performance of the mixture of expert 3D MTANNs for FP reduction by use of free-response receiver-operating-characteristic (FROC) analysis.[47] The FROC curve of the trained mixture of expert 3D MTANNs is shown in Fig. 8. The FROC curve was obtained by a change in the threshold value for the output of the mixing ANN. This FROC curve indicates that the mixture of expert 3D MTANNs was able to eliminate 63% (142/224) of non-polyps (FPs) without removal of any of the
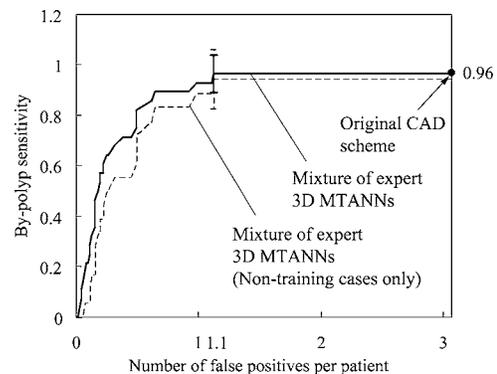
FIG. 8. The solid curve is an FROC curve that shows the overall performance of the mixture of expert 3D MTANNs when it was applied to the entire database of 27 polyps (48 true-positive volumes) and 224 FPs, and the dotted curve shows the performance of the mixture of expert 3D MTANNs when it was tested on the non-training-case-only database from which the ten training polyps were excluded. The error bars indicate that the difference between the sensitivities at an FP rate of 1.1 per patient under the two testing conditions is not statistically significant. It should be noted that the sensitivity for the non-training-case-only database was slightly lower because of the different number of polyps used, not because of detection misses. These FROC curves indicate that the mixture of expert 3D MTANNs yielded a reduction of 63% (142/224) of non-polyps (FPs) without removal of any true positives, i.e., it achieved 100% (27/27 or 17/17) classification sensitivity.

FIG. 9. Typical examples of sources of the remaining FPs that were not removed by the mixture of expert 3D MTANNs at the specific operating point reported in the text. From left to right, a haustral fold, medium-sized solid stool, a part of the rectum attached to a rectal tube, and stool adhering to the colonic wall.

27 polyps, i.e., a 96.4% (27/28) overall by-polyp sensitivity was achieved at an FP rate of 1.1 (82/73) per patient.

Because the 27 polyps included the ten training polyps, this evaluation could be biased. In order to reduce this bias, we excluded the ten training polyps from the evaluation of the mixture of expert 3D MTANNs. The FROC curve for only non-training cases is shown in Fig. 8. The FROC curves indicate that the performance of the mixture of expert 3D MTANNs for non-training cases only is similar to that for the entire database, e.g., a classification sensitivity of 100% (17/17) was achieved at an FP rate of 1.1 (82/73) per patient (there is no statistically significant difference between the sensitivities for the entire database and the non-training-case-only database, as indicated by the error bars representing 95% confidence intervals[48]). The 17 polyps included a polyp on a haustral fold, polyps attached to folds, and polyps touching stool. Thus, the mixture of expert 3D MTANNs was able to distinguish polyps under such varied situations from normal structures. Typical examples of FPs that were not removed by the mixture of expert 3D MTANNs are shown in Fig. 9; they include a haustral fold, medium-sized solid stool, a part of the rectum attached to a rectal tube, and stool adhering to the colonic wall.

We investigated the effect of the change in the number of expert 3D MTANNs on the performance of the mixture of expert 3D MTANNs. The performance was evaluated by receiver-operating-characteristic (ROC) analysis.[49,50] It should be noted that the number of expert 3D MTANNs corresponds to the number of input units in the mixing ANN. The mixing ANN was evaluated by use of a leave-one-lesion-out cross-validation test. We examined all possible combinations of expert 3D MTANNs. Figure 10 shows the average area under the ROC curve (AUC) values[51] of the mixture of expert 3D MTANNs over various numbers of expert 3D MTANNs. The results show that the average performance of the mixture of expert 3D MTANNs was highest when the number of expert 3D MTANNs was four, although the difference was very small. The differences in any combinations of the AUC values for the mixtures of expert 3D MTANNs were not statistically significant (two-tailed p value >0.05).[52]

We also investigated the effect of the change in the number of hidden units in the mixing ANN. The mixing ANN was evaluated by use of a leave-one-lesion-out cross-validation test. The number of expert 3D MTANNs (i.e., the number of input units) was four. We examined all possible combinations of four expert 3D MTANNs. Figure 11 shows the average performance of the mixing ANN with various
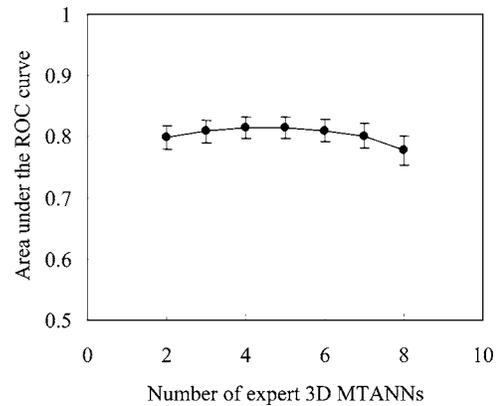
FIG. 10. Effect of change in the number of expert 3D MTANNs on the performance of the mixture of expert 3D MTANNs in a leave-one-lesion-out cross-validation test. Average performance over possible combinations of expert 3D MTANNs is shown. The standard deviations are shown as error bars. The AUC value was highest when four expert 3D MTANNs were used.

numbers of hidden units (an average was taken over possible combinations of four expert 3D MTANNs). The differences of any combinations of the AUC values for the mixing ANNs were not statistically significant (two-tailed p value >0.05).[52] Thus, the performance did not demonstrate a sensitivity to the number of hidden units.

IV. DISCUSSION

A limitation of this study is the number of CTC cases with polyps: use of a larger database will provide more reliable evaluation results for the performance of the mixture of expert 3D MTANNs. However, it should be noted that, although the mixture of expert 3D MTANNs was trained with only ten polyps, the performance for 27 polyps including the ten polyps and that for only 17 non-training polyps were very similar, which reflects the robustness of the mixture of expert 3D MTANNs. This observation on the generalization ability of the mixture of expert 3D MTANNs is consistent with that of MTANNs in our previous studies[16,35,38,44] which
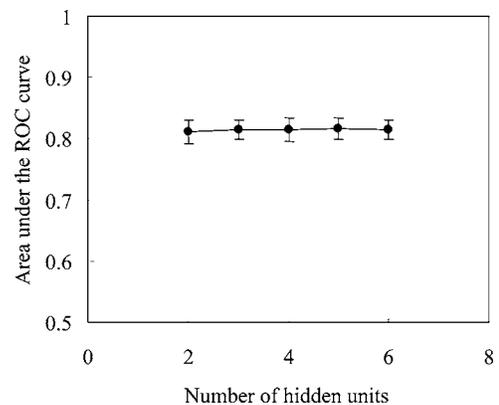
FIG. 11. Effect of change in the number of hidden units on the performance of the mixing ANN in a leave-one-lesion-out cross-validation test. Average performance over possible combinations of four expert 3D MTANNs is shown. The standard deviations are shown as error bars. The performance was not sensitive to the number of hidden units.

involved 109 lung nodules in thoracic CT,[35,36] 76 malignant nodules and 413 benign nodules in thoracic CT,[38] and 91 lung nodules in chest radiography.[37] Therefore, we expect that the performance of the mixture of expert 3D MTANNs for a large database would potentially be comparable to that obtained in this study.

The four categories used for the four expert 3D MTANNs were rectal tubes, stool with bubbles, colonic walls with haustral folds, and solid stool, because this combination achieved the highest performance among all combinations of FP categories in terms of the AUC value. The other four categories (small bulbous folds, elongated folds, strip-shaped folds, and the ileocecal valve) were not selected, probably because the category of colonic walls with haustral folds includes patterns in these four categories and thus correlates well with the four categories. In effect, three of the four categories are subcategories of haustral folds, and the ileocecal valve generally appears similar to a thickened haustral fold.

In previous study,[16] we investigated the effect of the intra- and inter-observer variations in selecting training cases on the performance of an expert 3D MTANN, because the performance would depend on the manual selection of training cases. The differences in the performance of the trained 3D MTANNs with the three different sets selected by the same observer at different times (i.e., the intra-observer variation) were not statistically significant (two-tailed p-value >0.05). The differences in the performance of the trained 3D MTANNs depended on which observer selected the training cases (i.e., the inter-observer variation), although these differences were not statistically significant (two-tailed p-value >0.05). Moreover, we investigated the effect of the number of training cases on the performance of an expert 3D MTANN.[16] The expert 3D MTANN trained with 20 cases had the highest AUC value, which provides the rationale for using 20 training cases in this study. This 20-case training approach was consistent with our previous MTANN studies in the distinction between nodules and non-nodules in thoracic CT.[35–38,44]

The MTANN is trained not on a case (whole volume) basis, but on a subvolume (i.e., the 7-by-7-by-7 input subvolume) basis. The MTANN does not see a whole polyp at a time, but sees local subvolumes one by one and gathers the local pattern features in the subvolumes. The local pattern features may include the mean CT value, the gray-level distribution, edges, texture, curvedness, and shape, all of which can, in theory, be calculated from voxels in a local volume. The MTANN is trained to learn those local pattern features in different-sized polyps in an effort to distinguish polyps from non-polyps. If the local pattern features are similar, the trained MTANN outputs similar values for both a medium-sized polyp and a larger polyp. The performance of the MTANN would, thus, be less dependent on the sizes of objects, but dependent on local pattern features.

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 non-polyps. 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. In conjunction with the training with the same-sized Gaussian distribution, we used the Gaussian weighting function with a fixed size for scoring the output volume of the trained 3D MTANN for each polyp candidate. This fixed-size approach allows 3D MTANNs to be relatively robust against size variations of polyps.

The 3D MTANNs were able to be trained with ten polyps and ten non-polyps. The key to this property is the division of cases into a large number of overlapping subvolumes.[44] By dividing a case (volume) into a large number of subvolumes, a 3D MTANN can be trained not on a case basis, but on a subvolume basis. The massive training with a large number of training subvolume samples (3375 samples $\times$ 20 cases $=67\,500$ samples) allows a 3D MTANN to avoid the "over-fitting" problem[53] of ANNs. This problem often occurs when the number of training samples is too small for determining the parameters of an ANN. The number of training samples required is, in general, greater than the number of parameters of an ANN. For example, an ANN with a 7-9-1 structure (7 inputs $\times$ 9 input-hidden-layer weights $+$ 9 hidden-output-layer weights $+$ 10 offsets $+$ 100 parameters) required 400–800 training samples to achieve an adequate performance for non-training cases.[54] The massive training with a large number of training samples (67 500), which, however, are not completely independent, would contribute to the proper determination of the parameters (343 inputs $\times$ 25 input-hidden-layer weights $+$ 25 hidden-output-layer weights $+$ 21 offsets $=8621$ parameters) of the 3D MTANN, and avoid the over-fitting problem. Thus, the massive training with a large number of training subvolume samples allows a 3D MTANN to be trained with a small number of cases.

In our previous study,[38] we performed an experiment to gain insight into the enrichment of the input information to an MTANN by the division of cases into subregions. We examined the relationship between ten training lesions and 76 lesions in the entire database in the input multidimensional vector space. We applied principal-component analysis (also referred to as Karhune–Loeve analysis)[55] on the input vector to the MTANN. The result of our analysis showed that the ten training cases represented the entire database of 76 cases very well in the input vector space, i.e., the ten training lesions covered, on average, 94% of the components of each lesion. Because all components of each lesion are combined with the scoring method in the MTANN, the non-covered 6% of components would not be critical at

all for the classification accuracy. Thus, the division of each lesion into a large number of subregions enriched the input information on lesions for the MTANN.

In our previous study,[38] we did experiments to examine the effect of the number of MTANNs and that of the number of hidden units on the performance of multiple MTANNs. In the experiments, we used a relatively large database containing 76 malignant nodules and 413 benign nodules in thoracic CT. The result showed that, as the number of MTANNs increased from two to eight, the AUC value went up from 0.81 and peaked at 0.88 when the number of MTANNs was six, and then declined to 0.84. As to the number of hidden units, as the number of hidden units increased from two to seven, the AUC value went up from 0.84 and peaked at 0.88 when the number of MTANNs was four, and then declined to 0.86. Thus, the trend of the changes on the performance was similar when the number of positive cases was larger, but the amount of the changes was slightly larger.

Wang *et al.*[26] have developed an FP reduction method consisting of extraction of three types of internal features (i.e., geometric, morphologic, and texture features) and classification based on the features, whereas the first FP reduction method in our CAD scheme consists of extraction of geometric (including morphologic) and texture features and a Bayesian ANN with the features for classification. Although the features and the classifiers used are different, both FP reduction methods can be considered as standard methods in a CAD scheme. On the other hand, some researchers have developed a specialized method for the reduction of FPs due to a specific FP source. Such a method can be used as a second FP reduction method (or an additional FP reduction method) in a CAD scheme. Iordanescu and Summers[24] have developed an image-segmentation-based approach for the reduction of FPs due to rectal tubes. Summers *et al.*[25] have developed a method for the reduction of FPs caused by the ileocecal valve. As a second FP reduction method, we have developed a single 3D MTANN specifically designed for removal of rectal-tube-induced FPs.[16] With this single 3D MTANN, we were able to eliminate all rectal-tube-induced FPs without removal of any true positives. Our mixture of expert 3D MTANNs was not specifically trained for the ileocecal value, but an expert 3D MTANN eliminated FPs due to the ileocecal valve. Therefore, the two FP reduction methods by Iordanescu and Summers *et al.* can be considered as alternatives to two expert 3D MTANNs.

In our previous study,[16] we have investigated the effect of the inter-observer variation in selecting training cases on the performance of a 3D MTANN. The results indicated that the performance of the 3D MTANNs trained with the cases selected by three different observers was slightly different (AUC values of 0.78, 0.74, and 0.73; the differences among the AUC values were not statistically significant). Thus, the performance of the 3D MTANN did depend on the observer who selected the training cases, but the difference was relatively small.

V. CONCLUSION

We developed a mixture of expert 3D MTANNs for the reduction of multiple types of FPs in a CAD scheme for the detection of polyps in CTC. With this mixture of expert 3D MTANNs, we were able to eliminate various types of FPs without removal of any true positives and achieved a classification sensitivity of 100% (17/17) with a reduction in the number of FPs by 63% (142/224). Overall, a 96.4% (27/28) by-polyp sensitivity with an FP rate of 1.1 (82/73) per patient was achieved. Thus, our 3D MTANN could be useful for improving the performance of a CAD scheme. We believe that a CAD scheme for the detection of polyps with a low FP rate would be useful for radiologists in detecting polyps in CTC.

[a] Author to whom correspondence should be addressed. Electronic mail: suzuki@uchicago.edu

[1] A. Jemal *et al.*, "Cancer statistics, 2005," CA-Cancer J. Clin. **55**(1), 10–30 (2005).

[2] S. J. Winawer *et al.*, "Colorectal cancer screening: clinical guidelines and rationale," Gastroenterology **112**(2), 594–642 (1997).

[3] A. H. Dachman, *Atlas of Virtual Colonoscopy* (Springer-Verlag, New York, 2003).

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

[5] J. G. Fletcher *et al.*, "CT colonography: Unraveling the twists and turns," Curr. Opin. Gastroenterol. **21**(1), 90–98 (2005).

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

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

[8] H. Yoshida *et al.*, "Computerized detection of colonic polyps at CT colonography on the basis of volumetric features: Pilot study," Radiology **222**(2), 327–336 (2002).

[9] H. Yoshida *et al.*, "Computer-aided diagnosis scheme for detection of polyps at CT colonography," Radiographics **22**(4), 963–979 (2002).

[10] H. Yoshida and J. Näppi, "Three-dimensional computer-aided diagnosis scheme for detection of colonic polyps," IEEE Trans. Med. Imaging **20**(12), 1261–1274 (2001).

[11] R. M. Summers *et al.*, "Automated polyp detection at CT colonography: Feasibility assessment in a human population," Radiology **219**(1), 51–59 (2001).

[12] D. S. Paik *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).

[13] G. Kiss *et al.*, "Computer-aided diagnosis in virtual colonography via combination of surface normal and sphere fitting methods," Eur. Radiol. **12**(1), 77–81 (2002).

[14] J. Näppi *et al.*, "A new high-performance CAD scheme for the detection of polyps in CT colonography," Med. Imaging 2004: Image Processing **5370**, 839–848 (2004).

[15] R. M. Summers *et al.*, "Computed tomographic virtual colonoscopy computer-aided polyp detection in a screening population," Gastroenter-

ology **129**(6), 1832–1844 (2005).

[16] K. Suzuki *et al.*, "Massive-training artificial neural network (MTANN) for reduction of false positives in computer-aided detection of polyps: Suppression of rectal tubes," Med. Phys. **33**(10), 3814–3824 (2006).

[17] S. B. Gokturk *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).

[18] J. 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).

[19] J. Nappi *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).

[20] B. Acar *et al.*, "Edge displacement field-based classification for improved detection of polyps in CT colonography," IEEE Trans. Med. Imaging **21**(12), 1461–1467 (2002).

[21] A. K. Jerebko *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).

[22] A. K. Jerebko *et al.*, "Multiple neural network classification scheme for detection of colonic polyps in CT colonography data sets," Acad. Radiol. **10**(2), 154–160 (2003).

[23] A. K. Jerebko *et al.*, "Support vector machines committee classification method for computer-aided polyp detection in CT colonography," Acad. Radiol. **12**(4), 479–486 (2005).

[24] G. 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).

[25] R. 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).

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

[27] H. Yoshida and J. Nappi, "Three-dimensional computer-aided diagnosis scheme for detection of colonic polyps," IEEE Trans. Med. Imaging **20**(12), 1261–1274 (2001).

[28] J. 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).

[29] M. A. Kupinski *et al.*, "Ideal observer approximation using Bayesian classification neural networks," IEEE Trans. Med. Imaging **20**(9), 886–899 (2001).

[30] C. D. Johnson and A. H. Dachman, "CT colonography: The next colon screening examination?" Radiology **216**(2), 331–341 (2000).

[31] K. 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).

[32] K. Suzuki *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).

[33] K. 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).

[34] K. Suzuki *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).

[35] K. Suzuki *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).

[36] H. Arimura *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).

[37] K. Suzuki *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).

[38] K. Suzuki *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).

[39] K. Suzuki, H. Abe, and 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).

[40] K. 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).

[41] K. Suzuki, "Determining the receptive field of a neural filter," J. Neural Eng. **1**(4), 228–237 (2004).

[42] C. I. Mosier, "Problems and designs of cross-validation," Educ. Psychol. Meas. **11**, 5–11 (1951).

[43] H. Frimmel, J. Nappi, and H. Yoshida, "Fast and robust computation of colon centerline in CT colonography," Med. Phys. **31**(11), 3046–3056 (2004).

[44] K. 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).

[45] K. Suzuki *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).

[46] K. Funahashi, "On the approximate realization of continuous mappings by neural networks," Neural Networks **2**, 183–192 (1989).

[47] J. P. Egan, G. Z. Greenberg, and A. I. Schulman, "Operating characteristics, signal detectability, and the method of free response," J. Acoust. Soc. Am. **33**, 993–1007 (1961).

[48] D. C. Edwards *et al.*, "Maximum likelihood fitting of FROC curves under an initial-detection-and-candidate-analysis model," Med. Phys. **29**(12), 2861–2870 (2002).

[49] C. E. Metz, "ROC methodology in radiologic imaging," Invest. Radiol. **21**(9), 720–733 (1986).

[50] C. 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).

[51] J. 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).

[52] C. E. Metz, B. A. Herman, and C. A. Roe, "Statistical comparison of two ROC-curve estimates obtained from partially-paired datasets," Med. Decis. Making **18**(1), 110–121 (1998).

[53] C. M. Bishop, *Neural Networks for Pattern Recognition* (Oxford University Press, New York, 1995).

[54] H. P. Chan *et al.*, "Classifier design for computer-aided diagnosis: Effects of finite sample size on the mean performance of classical and neural network classifiers," Med. Phys. **26**(12), 2654–2668 (1999).

[55] E. Oja, *Subspace Methods of Pattern Recognition* (Research Studies Press, Wiley, Letchworth, Hertfordshire, England, 1983).