Evaluation of MTANNs for eliminating false-positive with different computer aided pulmonary nodules detection software

Zhenghao Shi[1], Jiejue Ma[1], Yaning Feng[1], Lifeng He[3] and Kenji Suzuki[3]
[1]School of Computer Science and Engineering, Xi'an University of Technology, Xi'an, China
[2]Department of Radiology, University of Chicago, Chicago, IL, USA,
[3]Aichi Prefectual University, Nagakute-cho, Aichi, Japan

Abstract: MTANN (Massive Training Artificial Neural Network) is a promising tool, which applied to eliminate false-positive for thoracic CT in recent years. In order to evaluate whether this method is feasible to eliminate false-positive of different CAD schemes, especially, when it is applied to commercial CAD software, this paper evaluate the performance of the method for eliminating false-positives produced by three different versions of commercial CAD software for lung nodules detection in chest radiographs. Experimental results demonstrate that the approach is useful in reducing FPs for different computer aided lung nodules detection software in chest radiographs.

Keywords: False positive, MTANNs, computer-aided diagnosis.

INTRODUCTION

Because too many false positives (FPs) will potentially degrade the performance of (Computer-Aided Diagnosis) scheme in the clinical application, various methods have been proposed to eliminate FPs in CAD schemes (Giger *et al.,* 1998 and Yoshida *et al.,* 1996 and Pendo *et al.,* 1998). Most of the methods used the traditional two-stage pattern-recognition mode, i.e., feature extraction firstly and then classification followed. One main limitation of these methods is that the performance of the classifier depends directly on the discriminatory power of the feature set (Pendo *et al.,* 1998) and whereas to define a suitable feature set (or space) is a challenge tasks in classification. Thus, classification performance would be limited.

Recently, Suzuki *et al*. (Suzuki *et al.,* 2003; 2005 and 2006) showed that this limitation could be overcome by using Massive-Training Artificial Neural Network (MTANN). Compared with traditional pattern classification model, MTANN has three major advantages, namely, (1) directly operate on pixel data without feature extraction; (2) good generalization performance; and (3) trained with a few cases.

Although the MTANN approach demonstrated promise in its applications to chest radiography, thoracic CT, and CT colonography, it is not clear whether the MTANNs is feasible to false-positive reduction of different CAD schemes, especially, when it is applied to commercial CAD software. For this reason, we evaluated in this paper, the performance of the MTANN when applied to commercial CAD software. In our experiment, a mixture of expert MTANNs (Suzuki *et al.,* 2007 and Suzuki *et al.,* 2008) was employed, we propose 50 expert MTANNs for 50 different types of FPs, where each of the 50 MTANNs is trained by use of input chest radiographic images and teaching images containing an estimate for the distribution for a likelihood of being a nodule. Several higher performance MTANNs are selected to construct MTANNs ensemble. The outputs of the multiple MTANNs are combined with a mixing ANN for eliminate various different types of FPs. Experimental results demonstrate that the method can provide supper effects to other approaches.

MATERIAL AND METHODS

Material

29 posterior-anterior (PA) chest radiographs of normal (i.e., nodule free) cases and 23 PA chest radiographs with nodules were used in our study. The matrix size of the chest images was 1760×1760 pixels (pixel size was 0.2 mm; gray scale was 10 bits). Nodules absence or presence in the chest radiographs were confirmed by CT examinations. An experienced chest radiologist identified nodule locations.

Mixture of expert MTANNs

Fig. 1 shows the architecture of a mixture of expert MTANNs. It consists of multiple MTANNs arranged in parallel. It consists of a modified three layer artificial neural network (ANN) where the activation function of the unit in the input is an identity function, that in hidden is a sigmoid function, and that in output layers is a linear function, respectively. Because a linear function was employed as the activation function of the unit in the output layer, the ANN can directly operate on image data.

In our experiments, 50 expert MTANNs were prepared for 50 different types of false positives. After training of all MTANNs, several higher performance MTANNs are selected to construct a mixture. In order to eliminate all

**Corresponding author:* e-mail: ylshi@xaut.edu.cn

major sources of FPs, in the mixture, the outputs of multiple MTANNs are combined with a mixing ANN. The mixing ANN consists of a modified three layer ANN with a modified BP training algorithm for processing continuous output values, i.e., the activation function of the unit in the input is an identity function, that in hidden is a sigmoid function, and that in output layers is a linear function, respectively. The scores of each selected MTANN are entered to each input unit in the mixing ANN. The scores of each expert MTANN act as the features for distinguish nodules from a specific type of non-nodule. One unit is employed in the output layer for distinction between a nodule and a non-nodule, as represented as

$$O_{\mathrm{m}} = NN[\{S_{n,c}\}] \,|\, 1 \le n \le N \qquad (1)$$

Where NN{•} is the output of the modified ANN. The teaching values for nodules are one and those for non-nodules are zero.

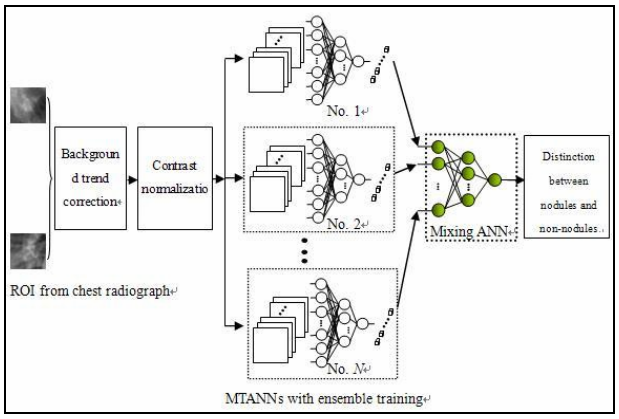

Fig. 1: The architecture of a mixture of expert MTANNs

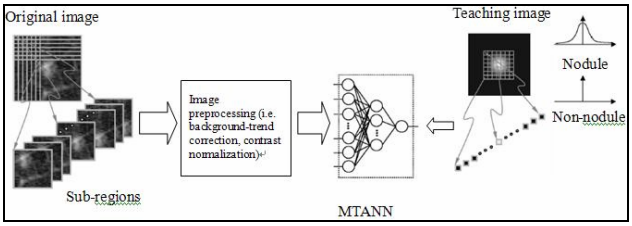

Fig. 2: Training of a massive-training artificial neural network (MTANN).

In order to remove the effects of image background on the performance of MTANNs, an image pre-processing step (including background-trend correction and contrast normalization) is utilized. Detail about background-trend correction and contrast normalization referenced to (Suzuki *et al.*, 2005).

Training of the mixture of expert MTANNs

The training for single MTANN is done by using amount of sub-regions (361=19×19 pixels) obtained by scanning pixel by pixel over the training region (27×27 pixels) in an input image together with the teaching image for nodules, the peak of which is located at the center, and that for non-nodules contains zeros, as shown in fig. 2.

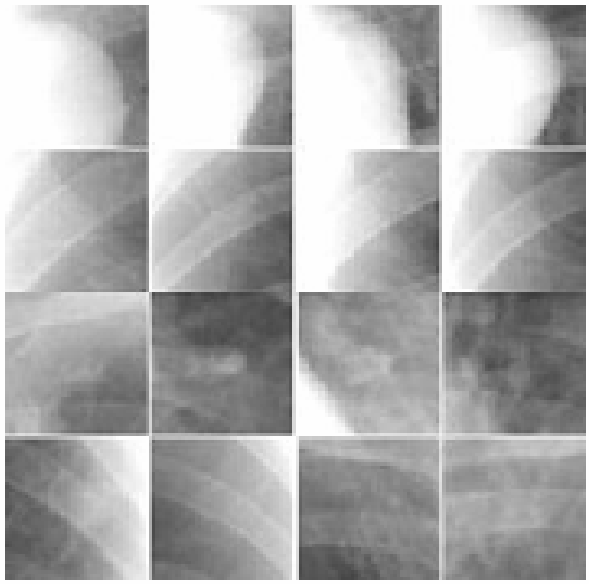

Fig. 3: Illustration of (a) nodules and (b) non-nodules used for training four expert MTANNs.

Pixel values of each sub-region of a region of interest (ROI) are employed as the inputs of the MTANN. The output of the MTANN is a continuous value corresponding to the center pixel in the sub-region, represented by

$$\mathrm{f}(x,y) = NN\{g(x-i,y-i) \,|\, i,j \in R_{\mathrm{s}}\} \qquad (2)$$

Where $f(x,\mathrm{y})$ denotes an estimatation for the teaching value; x and y indices the coordinates, NN{•} is the output of the ANN, $g(x,y)$ is a normalized input pixel value, and R_s is the local window of the ANN.

Training of the mixing ANN is performed by use of a leave-one-nodule-out cross-validation scheme. After training, the mixing ANN is expected to output a value related to the "likelihood of being a nodule". By thresholding the output, discrimination between a nodule and a non-nodule can be obtained.

Distinct nodules from non-nodules using the mixture of MTANNs

A nodule was discriminated from a non-nodule by using a score defined from the output images of the n^{th} trained MTANN and a 2D Gaussian weighting function (Hansen and Salamon 1990), as shown in fig. 3.

$$S_{\mathrm{ns}} = \sum_{x,y \in R_E} f_G(\sigma_n; x, y) \times f_{ns}(x, y) \qquad (3)$$

where S_{ns} is the score of the n^{th} trained MTANN for the s^{th} nodule candidate, R_E is the region for evaluation, $f_{ns}(x,y)$

is the output image of the n^{th} trained MTANN for the s^{th} nodule candidate where its center corresponds to the center of R_E, $f_G(\sigma_n: x,y)$ is a two dimensional Gaussian function with standard deviation σ_n where its center corresponds to the center of R_E and n is the MTANN number in the multiple MTANNs. This score represents the weighted sum of the estimate for the likelihood of the image containing a nodule near the center, i.e., a higher score would indicate a nodule, and a lower score would indicate a non-nodule.

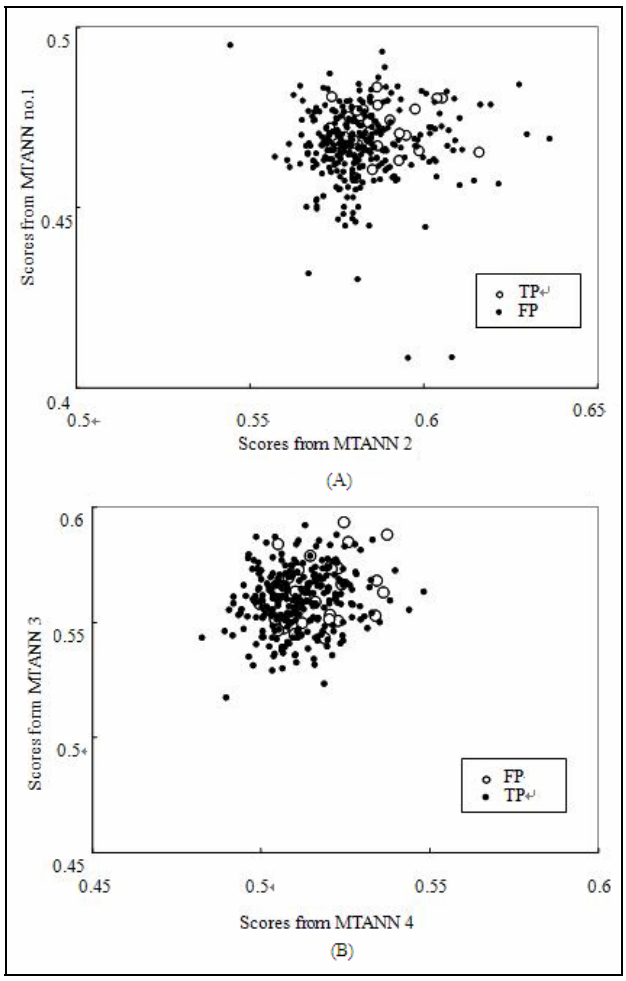

Fig. 4: Distributions of scores from expert MTANNs for nodules (white circles) and non-nodules (black dots) in validation test set 1. a) Distributions of scores from MTANN No.1 and MTANN No.2; b) Distributions of scores from MTANN No.3 and MTANN No.4.

MATERIALS AND METHODS

Training of expert MTANNs

The size of the sub region R_s of the MTANN was determined empirically to be 9x9 pixels, the standard deviation σ of the 2D Gaussian distribution was 5.0 pixels, and the size of the training region $R_T$in the teaching image was 19x19 pixels, respectively, based on our previous studies (Hansen and Salamon 1990; Mosier 1951). The number of units in the input was 81, that in hidden was 25, and that in output layers was 1, respectively. The learning rate was 0.001.

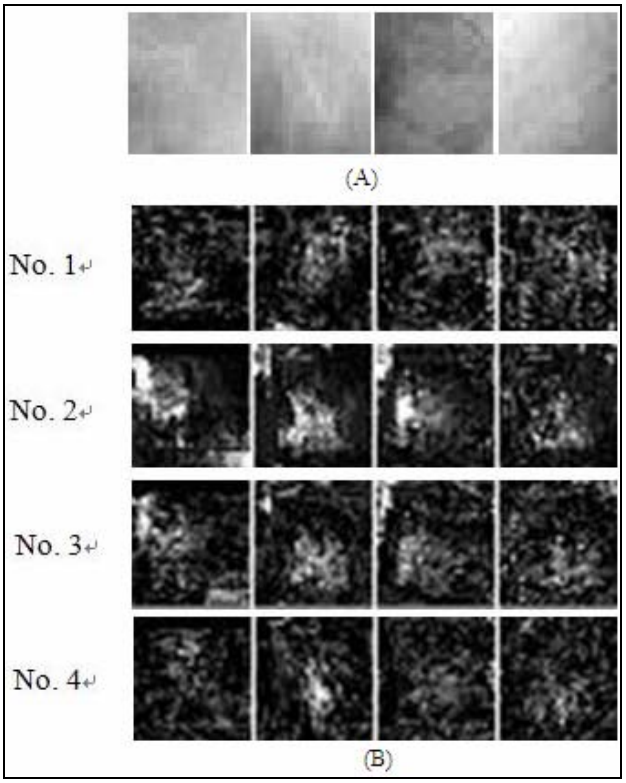

Fig. 5: Illustration of (a) non-training nodules and (b) the corresponding output images of the multiple MTANNs trained with ensemble training.

A three layer ANN was employed as the mixing ANN, where the number of units in the input layer was 8, that in hidden was 7, and that in output layer was1. The slope of the linear function of the output unit is 0.1 and the learning rate was 0.1, respectively.

Examples of the training samples are shown in fig. 3. These ROIs were extracted from chest radiographs in a training database, which is different from an independent test database. The sets of non-nodules included: (1) soft-tissue opacities, (2) right ribs, (3) small round opacities, and (4) left ribs.

Evaluation of the performance of mixture of expert MTANNs used for eliminating false positive in different pulmonary nodules CAD software

We apply the trained mixture of expert MTANNs to three validation test sets which included 281 FPs and 29 nodules (produced by CAD software 1 (Riverain CAD Ver.1.0)), 143 FPs and 33 nodules (produced by CAD software 2 (Riverain CAD Ver.3.0)), and 108 FPs and 31 nodules (produced by CAD software 3 (Riverain CAD Ver.3.0)), respectively.

Fig. 4 shows an example of distributions of scores for the output images of validation test set 1 from the expert MTANNs. From fig. 4, it can be seen that each expert removed different FPs.

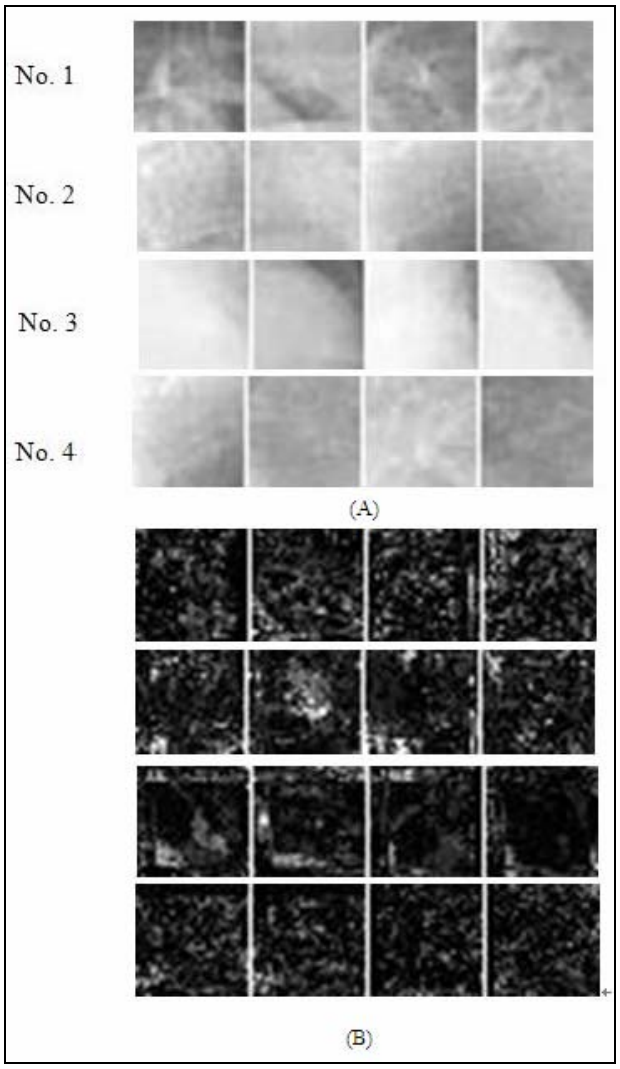

Fig. 6: Illustration of (a) non-training non-nodules and (b) the corresponding output images of the multiple MTANNs trained with ensemble training.

Fig. 5 and fig. 6 show examples of input images and corresponding output images for each one of the selected MTANNs with ensemble training for non-training cases, respectively. As shown in fig. 5, the output images of the trained MTANNs for nodules are represented by light distributions. The output images for non-nodules are relatively dark around the center, as shown in fig. 6.

Fig. 7 shows the distributions of the output values of the trained mixing ANN for validation test set 1. It can be seen that all nodules can be distinguished from the majority of FPs.

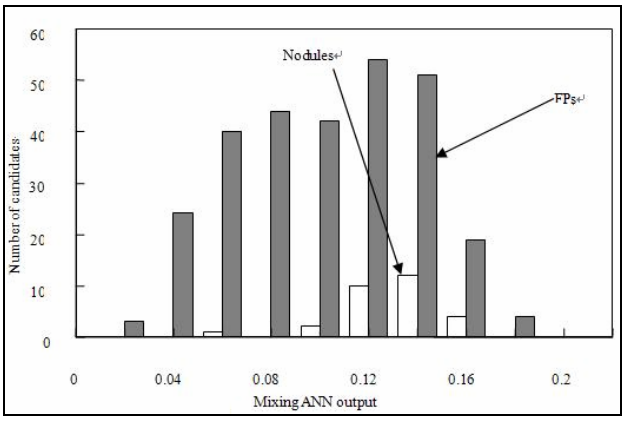

Fig. 7: Distribution of the output values from the trained mixing ANN for images of validation test set 1.

We evaluated the overall performance of the mixture of expert MTANNs for FPs reduction in three different pulmonary nodules CAD software by using free-response ROC (FROC) analysis. The experiment results are shown in fig. 8. As can be seen that with the mixture of expert MTANNs, the false-positive rate of CAD software 1 was reduced from 5.62 to 2.93 FPs per image while maintaining an original sensitivity of 60%; that of CAD software 2 was reduced from 2.86 to 1.6 positives per image while maintaining an original sensitivity of 60%; and that of CAD software 3 was reduced from 2.16 to 1.1 FPs per image while maintaining an original sensitivity of 60%; respectively. In contrast, with the best single MTANN, the FP rate of CAD software 1 was reduced from 5.62 to 3.6 FPs per image at the same sensitivity level, the FP rate of CAD software 2 was reduced from 2.86 to 2.2 FPs per image at the same sensitivity level, and CAD software 3 was reduced from 2.16 to 1.9 FPs per image at the same sensitivity level,, respectively. These results suggest that the capability of a mixture of expert MTANNs was superior to that of a single MTANN.

DISCUSSION

The performances of the trained mixture of expert MTANNs for the three validation test sets in this study were very similar, which reflects the robustness of the mixture of expert MTANNs. This observation on the generalization ability of the mixture of expert MTANNs is consistent with that of MTANNs in our previous studies (Suzuki *et al.,* 2003, 2005 and 2006). It is also consistent with the conclusion in literature (Mosier, 1951). The reasons to this high generalization ability, as discussed in previous study (Suzuki *et al.,* 2005), might be 1) the division of one nodule image into a large number (361=19x19) of subregions; 2) direct use of pixel values instead of image features as inputs of MTANN. Because the former factor, the variations in the feature components of nodules are enriched, and thus an MTANN is trained not on a case basis, but on a subregion basis, this makes

massive training with a large number (i.e., 7220=19x19x20) of subregions can contribute to the proper determination of the parameters of MTANNs. Due to the later factor, errors from incorrect image segmentation are kept out from MTANN.

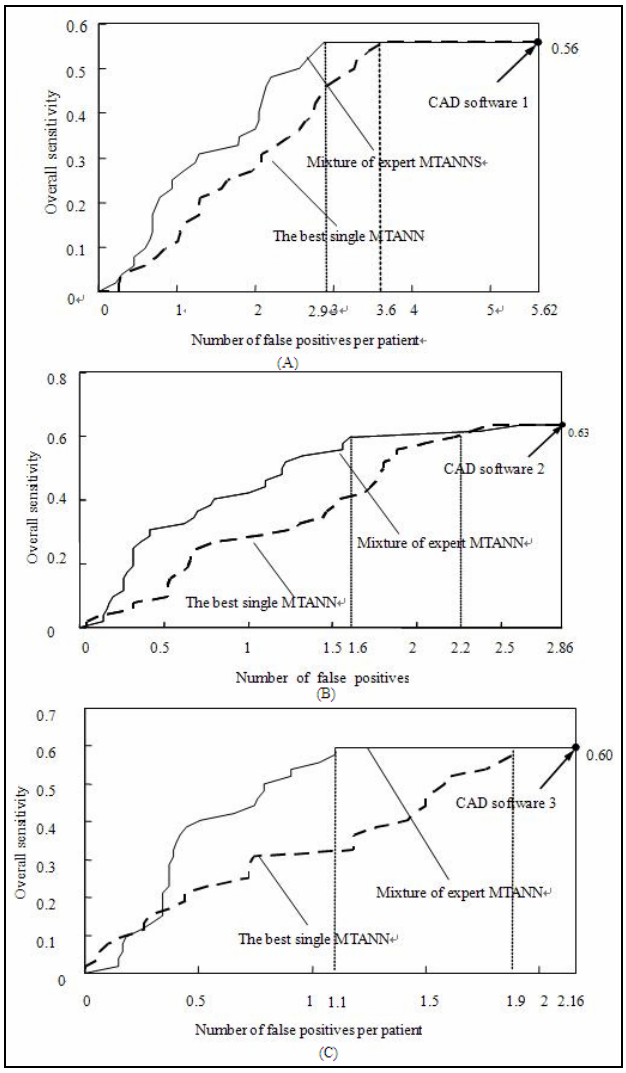

Fig. 8: FROC curve of the mixture of expert MTANNs (thick solid curve) and that of the best single MTANN (dashed curve) for a) 281 FPs and 29 nodules by CAD software 1, b) 143 FPs and 33 nodules by CAD software 2 and c) 108 FPs and 31 nodules by CAD software 3, respectively.

In previous study (Suzuki *et al.,* 2005), the performance of the mixing ANN were compared with that of the average operation, which is often used for combining multiple classifiers and would give better results compared to the majority logic. The results indicated that the performance of the mixing ANN was apparently superior to that of the average operation.

Though the time of applying a trained mixture of expert MTANNs for FPs reduction is negligibly small, training of the mixture of expert MTANNs is a time costing task, i.e., about 14 hours on a PC-based workstation (CPU: Pentium IV, 3.2 GHz) for each MTANN. Moreover, because the modified ANN algorithm employed for training MTANN is based on the BP algorithm, training of an MTANN sometimes might be trapped at local minima. So methods no matter for accelerating the convergence speed of the BP algorithm or for avoiding possible local minima will be helpful for improve training performance of an MTANN.

CONCLUSION

In this paper, we developed a mixture of expert MTANNs for reducing FPs in three different versions of commercial CAD software for detection of pulmonary nodules in chest radiographs. The mixture of expert MTANNs substantially reduced the false-positive rate of each of the three CAD softwares, while preserving a high level of sensitivity. Thus, the mixture of MTANNs based approach was effective for improving the performance of different CAD schemes for nodule detection in digital chest radiographs.

However, it should be noticed that the number of chest radiography cases with nodules in this study is limited, and use of a larger validation test database will provide more reliable evaluation results for the performance of the mixture of expert MTANNs. This will be done in the future research.

ACKNOWLEDGMENT

This work is partially supported by a grant from the Nature Science Foundation of Shaanxi Education Department (No.2013JK1136). The authors also gratefully acknowledge the valuable comments and suggestions of the reviewers.

REFERENCES

Giger ML, Doi K and MacMahon H (1988). Image feature analysis and computer aided diagnosis in digital radiograph. *Med. Phys.,* **15**(2): 158-166.

Hansen LK, Salamon P (1990). Neural network ensembles. *IEEE Transactions on Pattern analysis and Machine Intelligence,* **12**(10): 993-1001.

K Suzuki, SG Armato, F Li, S Sone and K Doi (2003). 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.

Mosier CI (1951). Problems and design of cross-validation. *Educ. Psychol. Meas.,* **11**: 5-11.

Penedo MG, Carreira MJ, Mosquera A and Cabello D (1998). Computer aided diagnosis: A neural network based approach to lung nodule detection. IEEE Trans. *Med. Imag.,* **17**(6): 872-880.

Suzuki K and Doi K (2005). 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? *Academic Radiology,* **12**(10): 1333-1341.

Suzuki K, Abe H, MacMahon H and Doi K (2006). Image-processing technique for suppressing ribs in chest radiographs by means of massive training artificial neural network (MTANN). *IEEE Transactions on Medical Imaging,* **25**(4): 406-416.

Suzuki K, He L, Khankari S, Ge L, Verceles J and Dachman AH (2007). Mixture of expert artificial neural networks with ensemble training for reduction of various sources of false positives in CAD. Proc. SPIE Medical Imaging (SPIE MI), **6514**: 651401-1-6.

Suzuki K, Li F, Sone S and Doi K (2005). 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 Transactions on Medical Imaging,* **24**(9): 1138-1150.

Suzuki K, Shiraishi J, Abe H, MacMahon H and Doi K (2005). False-positive reduction in computer-aided diagnostic scheme for detecting nodules in chest radiographs by means of massive training artificial neural network. *Academic Radiology,* **12**(2): 191-201.

Suzuki K, Yoshida H, Nappi J and Dachman AH (2006). Massive-training artificial neural network (MTANN) for reduction of false positives in computer-aided detection of polyps: Suppression of rectal tubes. *Medical Physics.,* **33**(10): 3814-3824.

Suzuki K, Yoshida H, Nappi J, Armato III SG and Dachman AH (2008). Mixture of expert 3D massive-training ANNs for reduction of multiple types of false positives in CAD for detection of polyps in CT colonography. *Medical Physics.,* **35**(2): 694-703.

Yoshida H, Doi K, Nishikawa RM, Giger ML and Schmidt RA (1996). An improved computer-assisted diagnostic scheme using wavelet transform for detecting clustered micro calcifications in digital mammograms. *Acad. Radiol.,* **3**(8): 621-627.