2022 IEEE Eighth International Conference on Big Data Computing Service and Applications (BigDataService) | 978-1-6654-5890-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/BigDataService55688.2022.00025

Small Data Deep Learning for Lung Cancer Detection in CT

Kenji Suzuki, *Senior Member*, *IEEE*
Biomedical Artificial Intelligence Research Unit (BMAI), Institute of Innovative Research
Tokyo Institute of Technology, Tokyo, Japan
suzuki.k.di@m.titech.ac.jp

***Abstract*— Big data analytics and deep learning (DL) for big data are ones of the most promising areas of research in healthcare and medicine. These technologies analyze large amount of complex heterogeneous data such as genomics, diagnostic, therapeutic, and electronic health records data; they are successful in improving healthcare of people, diagnostic and therapeutic performance of medical doctors, as well as reducing healthcare cost. In the area of DL for medical image diagnosis, however, collecting and annotating a large number of patient cases is a big challenge. DL models require medical images of 10,000 to 100,000 patient cases to adequately train, which would take years to collect. In the past research, less attentions were given to DL models that did not require big data. In this research, we developed a DL model that can be trained with a small number of cases, which we call small-data DL. We investigated a required number of cases for the small-data DL model for computer-aided detection (CAD) of lung nodules in CT. CAD provides a decision-support service to radiologists. We demonstrated that our small-data DL model, massive-training artificial neural network (MTANN), was able to achieve a state-of-the-art performance with a small number of cases (> 100). DL models that can be trained with a small number of cases would fill in the gap between big-data DL and areas where a large number of cases are not available in medicine and healthcare.**

Keywords—small data, deep learning, decision support system, lung cancer screening, small sample size, medical images

I. Introduction

Lung cancer ranks as the leading cause of cancer death among people in the United States. Evidence suggests that early detection of lung cancer may allow more timely therapeutic intervention and thus a more favorable prognosis for the patients. Accordingly, lung cancer screening programs are conducted in the United States, Japan, and other countries with low-dose helical computed tomography (CT) as the screening modality. To help radiologists, computer-aided detection and diagnosis (CAD) for lung nodules (i.e., potential lung cancer) in low-dose CT images have been investigated as a useful tool for lung cancer screening [1, 2]. CAD provides a decision-support service to radiologists in their decision making. Many investigators have developed a number of methods for CAD for detection of lung nodules in CT scans [3, 4] based on deep learning (DL) models [5] including a 3D U-Net and contextual convolutional neural network (CNN) [6], R-CNN [7], deep reinforcement learning [8], and a deep residual network (ResNet) [9].

Big data analytics and DL [10] for big data are ones of the most promising areas of research in healthcare and medicine. These technologies analyze large amount of complex heterogeneous data such as genomics, biological, diagnostic, therapeutic, and electronic health records data; they are successful in improving healthcare of people, diagnostic and therapeutic performance of medical doctors, as well as reducing healthcare cost. In the area of DL for medical image diagnosis and therapy, however, collecting and annotating a large number of patient cases is a big challenge. DL models require medical images of 10,000 to 100,000 patient cases to adequately train, which would take years to collect, e.g., collecting 100 lung cancer patients in a CT screening program may take a few years. Once collected, annotations of medical images (e.g., identifying the locations of lesions and manually segmenting them) by radiologists needs a huge workload thus a high cost. In the past research, less attentions were given to DL models that did not require big data or a large number of patient cases.

In this research, we developed a DL model that could be trained with a small number of cases, which we call small-data DL. The purpose of this research was to develop small-data DL that could be trained with a small-sample-sized dataset. We investigated a required number of cases for the small-data DL model in lung cancer detection in CT. We evaluated our small-data DL model compared with a state-of-the-art DL model.

II. Small Data Deep Learning Model

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

An MTANN was developed by extending neural filters and neural edge enhancers [11] to accommodate various pattern-recognition tasks [12, 13]. A two-dimensional (2D) MTANN was first developed for distinguishing a specific pattern from other patterns in 2D images [12]. The 2D MTANN was applied to reduction of false positive (FPs) in CAD for detection of lung nodules in CT [12] and chest radiography [14], the separation of bones from soft tissue in chest radiographs [15-18], and the distinction between benign and malignant lung nodules on 2D CT slices [19]. MTANNs used in an end-to-end manner has been proposed for detection of lung nodules in CT [20]. For processing of three-dimensional (3D) volume data, a 3D MTANN was developed by extending of the structure of the 2D MTANN, and it was applied to 3D CT colonography data [21-23].

The architecture and training of an MTANN is illustrated in Fig. 1. An MTANN consists of an ML model such as linear-

DOI 10.1109/BigDataService55688.2022.00025

output-layer artificial neural network (ANN) regression, support vector regression [24], and nonlinear Gaussian process regression, which is capable of operating on pixel data directly. The core part of the MTANN, i.e., an ANN regressor, consists of an input layer, multiple hidden layers, and an output layer. The linear-output-layer ANN regression model employs a linear function instead of a sigmoid function as the activation function of the unit in the output layer because the characteristics of the ANN were improved significantly with a linear function when it was applied to the continuous mapping of values in image processing [11, 25]. Note that the activation functions of the units in the hidden layers are a sigmoid function for nonlinear processing. The input to the MTANN consists of pixel values in a subregion (image patch), R, extracted from an input image. The output of the MTANN is a continuous scalar value, which is associated with the center pixel in the subregion, represented by

$$O(x,y,z) = ML\{I(x-i,y-j,z-k) \mid (i,j,k) \in R\}, \qquad (1)$$

where x, y, and z are the coordinate indices, $ML(\cdot)$ is the output of the ML regression model, and $I(x,y,z)$ is a pixel value of the input image. An ML regression model rather than an ML classification model would be suited for the MTANN framework, because the output of the MTANN consists of continuous scalar values (as opposed to nominal categories or classes used generally for other DL models). The entire output image is obtained by scanning with the input subregion (local window) of the MTANN in a convolutional manner on the entire input image, as illustrated in Fig. 2. This convolutional operation offers a shift-invariant property that is desirable for image classification. The input subregion and the scanning with the MTANN are analogous to the kernel of a convolution filter and the convolutional operation of the filter, respectively. The output of the MTANN is an image that may represent a likelihood map, unlike the class of typical CNNs (although some modified CNNs can output a map or image).

For use of the MTANN as a classifier, a scoring layer is placed at the end to convert the output probability map into a single score that represents a likelihood of being a certain class for a given image. A score for a given image is defined as

$$S = \sum_{(x,y,z)\in R_E} f_W(x,y,z) \times O(x,y,z), \qquad (2)$$

where f_W is a weighting function for combining pixel-based output responses from the trained MTANN into a single score, which may often be the same distribution function used in the teaching images, and with its center corresponding to the center of the region for evaluation, R_E. This score represents the weighted sum of the estimates for the likelihood that the image (e.g., a lesion candidate) contains an object of interest (e.g., a lesion) near the center; i.e., a higher score would indicate an object (e.g., a lesion) of interest, and a lower score would indicate other patterns. Thresholding is then performed on the scores for distinction between classes.

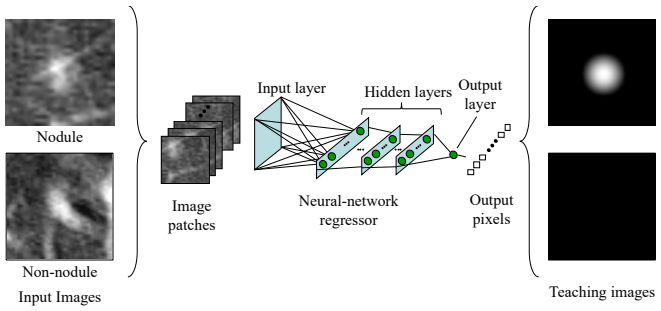

Fig. 1. Architecture and training of an MTANN small-data DL model

B. Training of an MTANN

The MTANN is trained with input images and the corresponding "teaching" (desired or ideal) images for enhancement of a specific pattern and suppression of other patterns in images. For enhancement of objects of interest (e.g., lesions), L, and suppression of other patterns (e.g., non-lesions), the teaching image contains a probability map for objects of interest, represented by

$$T(x,y,z) = \begin{cases} \textit{a certain distribution} & L \\ 0 & \textit{otherwise.} \end{cases} \qquad (3)$$

where the certain distribution may be a Gaussian distribution. For distinction between a nodule and non-nodules, the teaching image for the nodule contains a Gaussian distribution; that for non-nodules contains complete dark pixels, as illustrated in Fig. 1.

For enrichment of training samples, a training region, R_T, extracted from an input image is divided pixel by pixel into a large number of overlapping subregions (image patches), e.g., 200-500 patches. Single pixels are extracted from the corresponding teaching images as teaching values. The MTANN is massively trained by use of each of a large number of input subregions together with each of the corresponding teaching single pixels; hence the term "massive-training ANN." This massive-training scheme, i.e., training the model with a massive number of patches extracted from images of patients, allows the parameters of the model to be determined adequately with only a small number of cases [26]. In other words, the MTANN is trained on a pixel basis, as opposed to other DL models on a case/patient basis. For example, 10,000 training samples can be extracted from 10 positive and 10 negative images when 500 patches are extracted from one image. Recently, image transformers [27] employ a similar approach, namely, training based on patches. The error to be minimized by training of the MTANN is defined as:

$$E = \frac{1}{P} \sum_{(x,y,z)\in R_T} \{T(x,y,z) - O(x,y,z)\}^2 \qquad (4)$$

where P is the number of total training pixels in the training region, R_T. The MTANN is trained by a linear-output-layer back propagation algorithm which was derived for the linear-output-layer ANN model by use of the generalized delta rule. After training, the MTANN is expected to output the highest value when an object of interest is located at the center of the

subregion/local window of the MTANN, a lower value as the distance from the subregion center increases, and zero when the input subregion contains other patterns.

In order to distinguish nodules from various types of non-nodules, we extended the capability of a single MTANN and developed a mixture of multiple MTANN experts (multi-MTANN). The architecture of the multi-MTANN is shown in Fig. 2. The multi-MTANN consists of plural MTANNs that are arranged in parallel. Each MTANN is trained by use of non-nodules representing a different non-nodule type, but with the same nodules. Each MTANN acts as an expert for distinguishing nodules from a specific type of non-nodule. The scores from the expert MTANNs in the multi-MTANN are combined by use of fully-connected layers such that different types of non-nodules can be distinguished from nodules. The scores of each MTANN are entered to each input unit in the first layer; thus, the number of input units in the layer corresponds to the number of MTANNs. The scores of each MTANN function like the features for distinguishing nodules from a specific type of non-nodule with which the MTANN was trained. One unit is employed in the output layer for distinction between a nodule and a non-nodule. The teaching values for nodules are assigned the value one, and those for non-nodules are zero. After training, the fully-connected layer is expected to output a higher value for a nodule, and a lower value for a non-nodule. Thus, the output can be considered to be a value related to a "likelihood of being a nodule". By thresholding of the output, distinction between a nodule and non-nodules can be made.

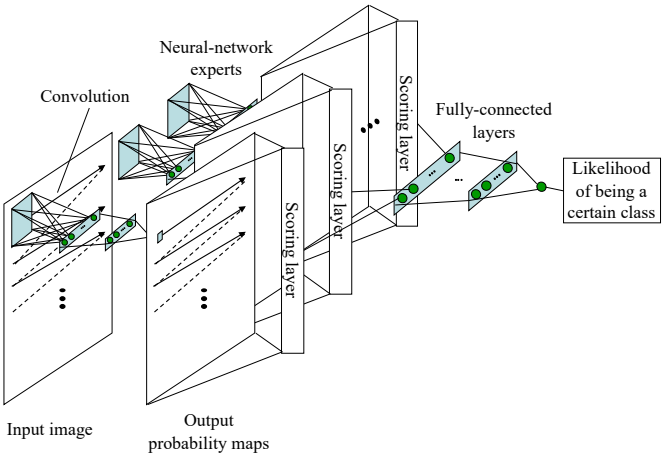

Fig. 2. Architecture of a mixture of multiple MTANN experts

III. Big Data Deep Learning Models

A. Convolutional Neural Network (CNN)

We decided to use CNNs as reference DL models in comparisons with MTANNs, because the performance of other DL models such as ResNET, U-Net, and VGGNet can be estimated by using the performance results of the CNNs. A CNN can be viewed as a simplified version of the Neocognitron model [28], which was proposed to simulate the human visual system in 1980. CNNs initially appeared in the early 1990s [29], but they did not enjoy much popularity at the time due to limited computational resources. The success of CNNs in computer vision has widely inspired investigators in the medical imaging community, which collectively demonstrates the effectiveness of CNNs for a variety of medical imaging tasks.

A CNN typically has a large number of convolutional and fully connected layers; therefore, it is not uncommon for a CNN to contain millions or billions of weights in its architecture. To determine such a large number of parameters, an CNN requires at least 10,000 to 100,000 cases for training, because it is trained on a case basis. To enrich data in a small-size dataset, data argumentation is applied to each image, generating a large number of new yet correlated training samples. We use common transformations of image scaling, translation, and rotation. Although the samples resulted from the data augmentation are correlated, they have proved effective in reducing over-fitting. In addition, we applied transfer learning to a CNN. The transfer learning is a very effective technique where the weights in a CNN are not trained from randomly initialized values, but rather from the weights of a CNN that was pre-trained on a large set of labeled training images from a different application. The above techniques would make it possible to obtain a high-performance CNN for the medical applications where only limited training data are available.

IV. Materials

A. Database of low-dose CT images with lung nodules

The database used in this research consists of 38 low-dose thoracic helical CT (LDCT) scans acquired from 31 different patients who participated voluntarily in a lung cancer screening program in Japan. The CT scans used for this research were acquired with a low-dose protocol of 120 kVp, 25 mA (11 scans) or 50 mA (27 scans), 10-mm collimation, and a 10-mm reconstruction interval at a helical pitch of two. The pixel size was 0.586 mm for 33 scans and 0.684 mm for five scans. Each reconstructed CT section had an image matrix size of 512 × 512 pixels. The 38 scans consisted of 1,057 sections and contained 50 nodules, including 38 "missed" nodules that represented biopsy-confirmed lung cancers and had not been reported or misreported during the initial clinical interpretation.

B. Initial CAD scheme for lung nodule detection in CT

We applied an initial CAD scheme for lung nodule detection in CT to the above database. The scheme consists of initial detection of lung nodule candidates by using a thresholding technique followed by mathematical morphology. The nodule candidates were provided as a form of nodule segmentation. A connected-component labeling technique [30] was applied to the segmented nodule candidates. A nine set of features such as gray-level-based features and morphological features were extracted from each candidate. Quadratic discriminant analysis was applied to the candidates in order to distinguish nodules from non-nodules.

For the above database, the initial CAD scheme yielded 40 detections of 50 nodules (80.0%) together with 1,078 FPs. In this research, we used all the 50 nodules, the locations of which were identified by a radiologist, and all 1,078 false positives generated by the initial CAD scheme. The use of radiologist-extracted true nodules with computer-generated false positives was intended to anticipate future improvements in the nodule detection sensitivity of the initial CAD scheme. It should be noted that the 1,078 false positives included in our evaluation

are considered as “difficult" false positives, because the feature-based machine learning in the initial CAD scheme was not able to distinguish them. Nodules and non-nodules (FPs) are illustrated in Fig. 3. The major source of the FPs was lung vessels, as most structures in the lungs were lung vessels.

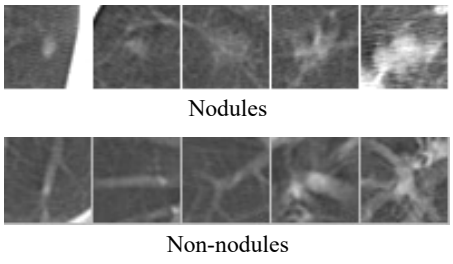

Fig. 3. Illustrations of nodules and non-nodules (FPs).

V. Experiments

A. MTANN and CNN training

Five MTANNs in the multi-MTANN were trained to distinguish nodules from various-sized vessels; and four MTANNs were applied to eliminate some other opacities. Each MTANN in the multi-MTANN was trained with a kernel of 9×9 pixels extracted from 10 nodule regions-of-interest (ROIs) and 10 non-nodule ROIs of size 50×50. The number of training ROIs increased from 10, 20, 30, 40, and 100 in order to evaluate the performance of the MTANNs with different training samples.

Five well-known CNNs including the AlexNet [31], the LeNet [32], a relatively deep CNN, a shallow CNN, and a transfer-learned AlexNet that used transfer learning from a computer-vision-trained AlexNet were trained with the same 100 cases in the same database in a 5-fold cross validation [33]. We compared the CNNs and the MTANNs extensively in the lung nodule detection problem in CT images.

VI. Results

The output images of the trained MTANN for nodules and non-nodules are illustrated in Fig. 4. As seen, the nodules are enhanced as bright pixels, whereas the non-nodules are suppressed (or disappear) in the output images.

Fig. 5 shows the performance of the MTANNs trained with 100, 40, 30, 20, and 10 cases. The performance decreased with decreasing the number of training cases. Table I shows the performance in terms of the number of FPs per patient (Pt) of the MTANNs.

Among the 5 CNN models, the best performing one was the transfer-learned AlexNet. Fig. 6 shows the comparisons of the performance of the MTANN and the best-performing CNN model (the transfer-learned AlexNet.). As seen, the performance of the MTANN is higher than that of the best-performing CNN. Noteworthy, at 100% sensitivity, the MTANN generated 2.37 FPs per patient, which was significantly lower than the best-performing CNN with 32.50 FPs per patient ($p < 0.05$), as shown in Table II. The experiment demonstrated that the performance of MTANNs was substantially higher than that of the best-performing CNN under the same condition (i.e., a small number of training cases).

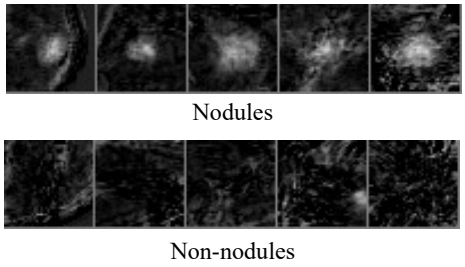

Fig. 4. Illustrations of the output images of the MTANN for nodules and non-nodules (FPs).

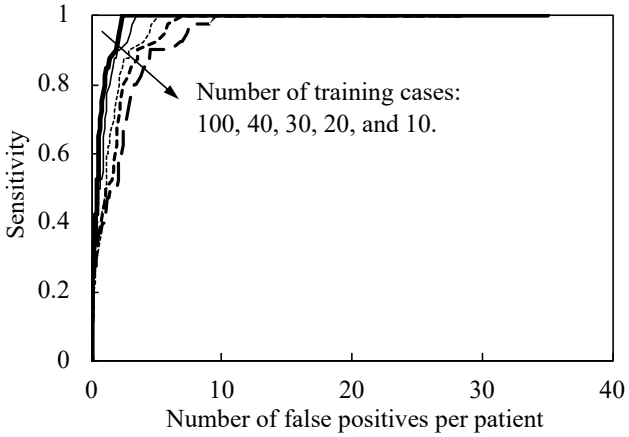

Fig. 5. Performance of MTANNs with different number of training cases.

TABLE I. Comparison among Performance of MTANNs with Different Numbers of Training Cases

# of training cases	10	20	30	40	…	100
# of FPs/Pt	9.56	7.00	5.16	3.47	…	2.37

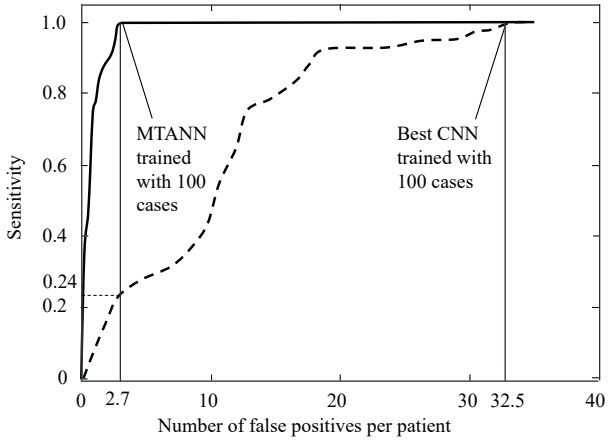

Fig. 6. Comparison between the MTANN and the best CNN trained with 100 cases.

TABLE II. Comparison between Performance of MTANN and Best CNN with 100 Training Cases

Best CNN	MTANN
32.50 FPs/Pt	2.37 FPs/Pt

VII. Conclusion

In this research, we developed a DL model that can be trained with a small number of cases, which we call small-data DL. We investigated a required number of cases for the small-data DL model, namely, MTANNs, in lung cancer detection in CT. We demonstrated that our small-data MTANN DL model was able to achieve a state-of-the-art performance with a small number of cases (> 100). DL models that require only a small number of cases would fill in the gap between big-data DL and areas where a large number of cases are not available in medicine and healthcare. We will perform more experiments with other databases to verify the robustness of our MTANN.

Acknowledgment

The author is grateful to Nima Tajbakhsh and other members of the Suzuki Lab for their valuable discussions and contributions. This research was supported by JST-Mirai Program Grant Number JPMJMI20B8, Japan.

References

[1] F. Shaukat, G. Raja, and A. F. Frangi, "Computer-aided detection of lung nodules: a review," *Journal of Medical Imaging,* vol. 6, no. 2, pp. 020901, 2019.

[2] L. M. Pehrson, M. B. Nielsen, and C. Ammitzbøl Lauridsen, "Automatic pulmonary nodule detection applying deep learning or machine learning algorithms to the LIDC-IDRI database: a systematic review," *Diagnostics,* vol. 9, no. 1, pp. 29, 2019.

[3] K. Suzuki, "A review of computer-aided diagnosis in thoracic and colonic imaging," *Quant Imaging Med Surg,* vol. 2, no. 3, pp. 163-76, Sep, 2012.

[4] A. El-Baz, G. M. Beache, G. Gimel'farb, K. Suzuki, K. Okada, A. Elnakib, A. Soliman, and B. Abdollahi, "Computer-aided diagnosis systems for lung cancer: challenges and methodologies," *Int J Biomed Imaging,* vol. 2013, pp. 942353, 2013.

[5] K. Suzuki, "Overview of deep learning in medical imaging," *Radiol Phys Technol,* vol. 10, no. 3, pp. 257-273, Sep, 2017.

[6] C. Zhao, J. Han, Y. Jia, and F. Gou, "Lung nodule detection via 3D U-Net and contextual convolutional neural network." *NaNA*, 2018, pp. 356-361.

[7] Y. Su, D. Li, and X. Chen, "Lung nodule detection based on faster R-CNN framework," *Computer Methods and Programs in Biomedicine,* vol. 200, pp. 105866, 2021.

[8] I. Ali, G. R. Hart, G. Gunabushanam, Y. Liang, W. Muhammad, B. Nartowt, M. Kane, X. Ma, and J. Deng, "Lung nodule detection via deep reinforcement learning," *Frontiers in oncology,* vol. 8, pp. 108, 2018.

[9] S. Cui, S. Ming, Y. Lin, F. Chen, Q. Shen, H. Li, G. Chen, X. Gong, and H. Wang, "Development and clinical application of deep learning model for lung nodules screening on CT images," *Scientific reports,* vol. 10, no. 1, pp. 1-10, 2020.

[10] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature,* vol. 521, no. 7553, pp. 436-444, 2015.

[11] K. Suzuki, I. Horiba, and N. Sugie, "Neural edge enhancer for supervised edge enhancement from noisy images," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 25, no. 12, pp. 1582-1596, December 2003.

[12] K. Suzuki, S. G. Armato, 3rd, F. Li, S. Sone, and K. Doi, "Massive training artificial neural network (MTANN) for reduction of false positives in computerized detection of lung nodules in low-dose computed tomography," *Med Phys,* vol. 30, no. 7, pp. 1602-17, Jul, 2003.

[13] K. Suzuki, "Pixel-based machine learning in medical imaging," *Int J Biomed Imaging,* vol. 2012, pp. 792079, 2012.

[14] K. Suzuki, J. Shiraishi, H. Abe, H. MacMahon, and K. Doi, "False-positive reduction in computer-aided diagnostic scheme for detecting nodules in chest radiographs by means of massive training artificial neural network," *Acad Radiol,* vol. 12, no. 2, pp. 191-201, Feb, 2005.

[15] A. Zarshenas, J. Liu, P. Forti, and K. Suzuki, "Separation of bones from soft tissue in chest radiographs: Anatomy-specific orientation-frequency-specific deep neural network convolution," *Med Phys,* vol. 46, no. 5, pp. 2232-2242, May, 2019.

[16] S. Chen, and K. Suzuki, "Computerized detection of lung nodules by means of "virtual dual-energy" radiography," *IEEE Trans Biomed Eng,* vol. 60, no. 2, pp. 369-78, Feb, 2013.

[17] S. Chen, and K. Suzuki, "Separation of bones from chest radiographs by means of anatomically specific multiple massive-training ANNs combined with total variation minimization smoothing," *IEEE Trans Med Imaging,* vol. 33, no. 2, pp. 246-57, Feb, 2014.

[18] S. Chen, S. Zhong, L. Yao, Y. Shang, and K. Suzuki, "Enhancement of chest radiographs obtained in the intensive care unit through bone suppression and consistent processing," *Phys Med Biol,* vol. 61, no. 6, pp. 2283-301, Mar 21, 2016.

[19] K. Suzuki, F. Li, S. Sone, and K. Doi, "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,* vol. 24, no. 9, pp. 1138-1150, Sep, 2005.

[20] K. Suzuki, "A supervised 'lesion-enhancement' filter by use of a massive-training artificial neural network (MTANN) in computer-aided diagnosis (CAD)," *Phys Med Biol,* vol. 54, no. 18, pp. S31-45, Sep 21, 2009.

[21] K. Suzuki, H. Yoshida, J. Nappi, and A. H. Dachman, "Massive-training artificial neural network (MTANN) for reduction of false positives in computer-aided detection of polyps: Suppression of rectal tubes," *Med Phys,* vol. 33, no. 10, pp. 3814-24, Oct, 2006.

[22] K. Suzuki, H. Yoshida, J. Nappi, S. G. Armato, 3rd, and A. H. Dachman, "Mixture of expert 3D massive-training ANNs for reduction of multiple types of false positives in CAD for detection of polyps in CT colonography," *Med Phys,* vol. 35, no. 2, pp. 694-703, Feb, 2008.

[23] K. Suzuki, D. C. Rockey, and A. H. Dachman, "CT colonography: advanced computer-aided detection scheme utilizing MTANNs for detection of "missed" polyps in a multicenter clinical trial," *Med Phys,* vol. 37, no. 1, pp. 12-21, Jan, 2010.

[24] J. W. Xu, and K. Suzuki, "Massive-training support vector regression and Gaussian process for false-positive reduction in computer-aided detection of polyps in CT colonography," *Medical Physics,* vol. 38, no. 4, pp. 1888-902, Apr, 2011.

[25] K. Suzuki, I. Horiba, N. Sugie, and M. Nanki, "Extraction of left ventricular contours from left ventriculograms by means of a neural edge detector," *IEEE Transactions on Medical Imaging,* vol. 23, no. 3, pp. 330-339, Mar, 2004.

[26] 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?," *Academic Radiology,* vol. 12, no. 10, pp. 1333-1341, Oct, 2005.

[27] N. Parmar, A. Vaswani, J. Uszkoreit, L. Kaiser, N. Shazeer, A. Ku, and D. Tran, "Image transformer." *ICML*, 2018, pp. 4055-4064.

[28] K. Fukushima, "Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biol Cybern,* vol. 36, no. 4, pp. 193-202, 1980.

[29] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural Comput.,* vol. 1, no. 4, pp. 541-551, 1989.

[30] L. He, Y. Chao, K. Suzuki, and K. Wu, "Fast connected-component labeling," *Pattern Recognition* vol. 42, pp. 1977-1987, 2009.

[31] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems,* vol. 25, 2012.

[32] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE,* vol. 86, no. 11, pp. 2278-2324, 1998.

[33] N. Tajbakhsh, and K. Suzuki, "Comparing Two Classes of End-to-End Learning Machines for Lung Nodule Detection and Classification: MTANNs vs. CNNs.," *Pattern Recognition,* vol. 63, pp. 476–486, 2017.