Overview of deep learning in medical imaging

Kenji Suzuki

Springer

Springer

CrossMark

Overview of deep learning in medical imaging

Kenji Suzuki[1,2]

Abstract The use of machine learning (ML) has been increasing rapidly in the medical imaging field, including computer-aided diagnosis (CAD), radiomics, and medical image analysis. Recently, an ML area called deep learning emerged in the computer vision field and became very popular in many fields. It started from an event in late 2012, when a deep-learning approach based on a convolutional neural network (CNN) won an overwhelming victory in the best-known worldwide computer vision competition, ImageNet Classification. Since then, researchers in virtually all fields, including medical imaging, have started actively participating in the explosively growing field of deep learning. In this paper, the area of deep learning in medical imaging is overviewed, including (1) what was changed in machine learning before and after the introduction of deep learning, (2) what is the source of the power of deep learning, (3) two major deep-learning models: a massive-training artificial neural network (MTANN) and a convolutional neural network (CNN), (4) similarities and differences between the two models, and (5) their applications to medical imaging. This review shows that ML with feature input (or feature-based ML) was dominant before the introduction of deep learning, and that the major and essential difference between ML before and after deep learning is the learning of image data directly without object segmentation or feature extraction; thus, it is the source of the power of deep learning, although the depth of the model is an important attribute. The class of ML with image input (or image-based ML) including deep learning has a long history, but recently gained popularity due to the use of the new terminology, deep learning. There are two major models in this class of ML in medical imaging, MTANN and CNN, which have similarities as well as several differences. In our experience, MTANNs were substantially more efficient in their development, had a higher performance, and required a lesser number of training cases than did CNNs. "Deep learning", or ML with image input, in medical imaging is an explosively growing, promising field. It is expected that ML with image input will be the mainstream area in the field of medical imaging in the next few decades.

1 Introduction

The use of machine learning (ML) has been increasing rapidly in the medical imaging field [1–13], including computer-aided detection (CADe) and diagnosis (CADx) [14–17], radiomics, and medical image analysis, because objects such as lesions and organs in medical images may be too complex to be represented accurately by a simple equation or model. For example, a lung nodule is generally modeled as a solid sphere, but there are nodules of various shapes and inhomogeneous nodules, such as spiculated nodules and ground-glass nodules. A polyp in the colon is modeled as a bulbous object, but there are also colorectal lesions that have a flat shape [18, 19]. Modeling of such

✉ Kenji Suzuki
  ksuzuki@iit.edu

1 Medical Imaging Research Center and Department of Electrical and Computer Engineering, Illinois Institute of Technology, 3440 South Dearborn Street, Chicago, IL 60616, USA

2 World Research Hub Initiative (WRHI), Tokyo Institute of Technology, Tokyo, Japan

complex objects needs a complex model with a large number of parameters. Determining such a large number of parameters cannot be accomplished manually, but has to be determined from data. Thus, detection and diagnostic tasks in medical imaging essentially require "learning from examples (or data)" for determination of a large number of parameters in a complex model. Therefore, ML plays an essential role in the medical imaging field.

One of the most popular uses of ML in computer-aided diagnosis (CAD), including CADe and CADx, and medical image analysis [6, 20] is the classification of objects such as lesions into certain classes (e.g., lesions or non-lesions, and malignant or benign) based on input features (e.g., contrast, area, and circularity) obtained from segmented objects. This class of ML is referred to as ML with feature input or feature-based ML. The task of ML is to determine "optimal" boundaries for separating classes in the multi-dimensional feature space that is formed by the input features [21].

A class of ML that is different from the class of the ML with feature input (feature-based ML) was proposed and developed in the field of CAD. The name of the ML technique in the new class is a "massive-training artificial neural network" (MTANN) [22]. This class of ML uses images as input, as opposed to features extracted from a segmented object. The ML class is referred to as ML with image input (or image-based ML) [23]. Thus, ML with image input (image-based ML) does not require feature calculation or object segmentation. This class of ML can be classified as end-to-end ML that goes through the entire process from input images to the final classification. MTANNs were applied to the reduction of false positives (FPs) in CADe for detection of lung nodules in CT [22, 24–27] and in chest radiographs [chest X-ray (CXR)] [28, 29], the separation of bones from soft tissue in CXR [29–34], the distinction between benign and malignant lung nodules in CT [35], and the FP reduction in CADe for detection of polyps in CT colonography [36–40].

Recently, an ML area called deep learning emerged in the computer vision field and became very popular in virtually all fields. It started from an event in late 2012. A deep-learning approach based on a CNN [41] won an overwhelming victory in the best-known worldwide computer vision competition, ImageNet Classification, with the error rate smaller by 11% than that in the 2nd place of 26% [42]. Consequently, the MIT Technology Review named it one of the top 10 breakthrough technologies in 2013. Since then, researchers in virtually all fields have started actively participating in the explosively growing field of deep learning [43]. Because deep learning directly uses images as input, it belongs to the class of ML with image input (image-based ML).

This paper presents an overview of the area of deep learning in medical imaging, including (1) what has changed in machine learning before and after the introduction of deep learning, (2) where the source of the power of deep learning comes from, (3) two major deep-learning models: a massive-training artificial neural network (MTANN) and a convolutional neural network (CNN), (4) similarities and differences between the two models, and (5) their applications to medical imaging.

## 2 Overview of machine learning in medical imaging

2.1 "Standard" ML: ML with feature input

One of the most popular uses of the ML algorithms is probably classification. In this use, an ML algorithm is often called a classifier. A standard ML approach in the computer vision field is illustrated in Fig. 1a. First, objects (e.g., lesions) in an image are segmented by use of a segmentation technique such as thresholding, edge-based segmentation, and an active contour model. Next, features such as contrast, circularity, and size are extracted from the segmented lesions by use of a feature extractor. Then, the extracted features are entered as input to an ML model such as linear or quadratic discriminant analysis (LDA or QDA) [44], a multilayer perceptron (MLP) [45], and a support vector machine (SVM) [46]. The ML model is trained with sets of input features and known class labels. The training is performed for determination of "optimal" boundaries for separating classes such as cancer or non-cancer in the multi-dimensional feature space that is formed by the input features. After training, the trained ML model determines to which class a new unknown lesion belongs. Thus, this class of ML can be referred to as ML with feature input, feature-based ML, object/feature-based ML, or a classifier.

2.2 "New" ML class: ML with image input

A class of ML that is different from the class of ML with feature input (feature-based ML or a classifier) was proposed and developed in the field of CAD. Suzuki et al. invented and developed massive-training artificial neural networks (MTANNs) for classification between lesions and non-lesions in medical images [22]. MTANNs use images as input, as opposed to features extracted from a segmented lesion. Thus, this class of ML with image input (or image-based ML) does not require feature calculation or lesion segmentation, and it can be classified as end-to-end ML that does the entire process from input images to the final classification.

A term, deep learning, was proposed for ML models for a high-level representation of objects by Hinton in 2007, but it was not recognized widely until late 2012. Deep
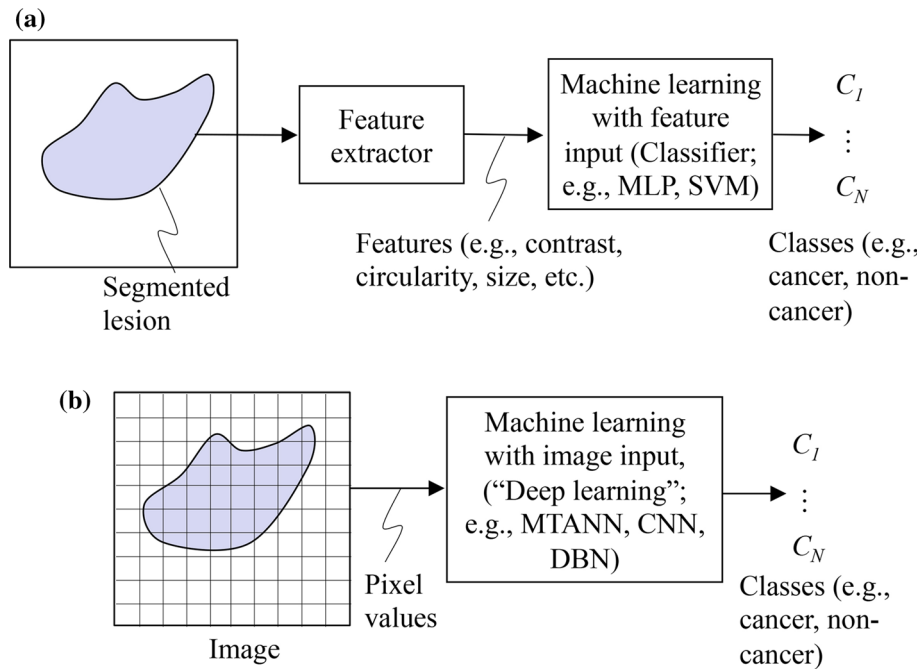
Fig. 1 Machine learning in the field of computer vision before and after the introduction of deep learning. **a** ML before learning: Standard ML for classifying lesions (i.e., ML with feature input or feature-based ML) in the field of computer vision. Features (e.g., contrast, circularity, and effective diameter) are extracted from a segmented lesion in an image. Those features are entered as input to an ML model with feature input (classifier) such as a multilayer perceptron (MLP) and a support vector machine (SVM). The output of the classifier consists of class categories such as cancer or non-cancer. **b** ML after deep learning: "New" ML class, ML with image input (image-based ML). Pixel values from an image are directly entered as input to an ML model with image input model such as a massive-training artificial neural network (MTANN), a convolutional neural network (CNN), and a deep belief net (DBN). This class of ML with image input (image-based ML) includes deep learning. Thus, the major and essential difference between ML approaches before and after the introduction of deep learning is direct training of pixels in images

learning became very popular in the computer vision field after late 2012, when a deep-learning approach based on a CNN [41] won an overwhelming victory in the best-known computer vision competition, ImageNet [42]. Deep learning such as deep belief nets (DBNs) [47] and deep CNNs uses pixel values in images directly instead of features calculated from segmented objects as input information; thus, feature calculation or object segmentation is not required, as shown in Fig. 1b. Although the development of segmentation techniques has been studied for a long time, segmentation of objects is still challenging, especially for complicated objects, subtle objects, and objects in a complex background. In addition, defining and extracting relevant features for a given task is a challenging task, as calculated features may not have the discrimination power that is sufficient for classifying objects of interest. Because deep learning can avoid errors caused by the inaccurate feature calculation and segmentation that often occur for subtle or complex objects, the performance of deep learning is generally higher for such objects than that of common classifiers (i.e., ML with feature input or object/feature-based MLs). Deep learning has multiple layers (>4) of nonlinear or quasi-nonlinear processing to acquire a high-level representation of objects or features in images. Compared to ML with feature input (object/feature-based ML, or a common classifier), deep learning skips steps of segmentation of objects, feature extraction from the segmented objects, and feature selection for determining "effective features", as shown in Fig. 2. Deep learning is also called end-to-end ML, because it enables the entire process to map from raw input images to the final classification, eliminating the need for hand-crafted features. It is interesting to note that people do not refer to the use of MLP with deep layers in the object/feature-based approach as deep learning, and they still call a shallow CNN with only a few layers of deep learning, which is the evidence that people are confused by the terminology, deep learning.

As shown in Fig. 1a, b, the major and essential difference between ML with feature input (feature-based ML) and ML with image input (image-based ML) is the use of pixels in images directly as input to ML models, as opposed to features extracted from segmented objects. This is true for ML approaches before and after the introduction of deep learning. Therefore, the terminology "deep learning" may mislead people to think that the power of deep learning comes from its depth. A proper terminology for
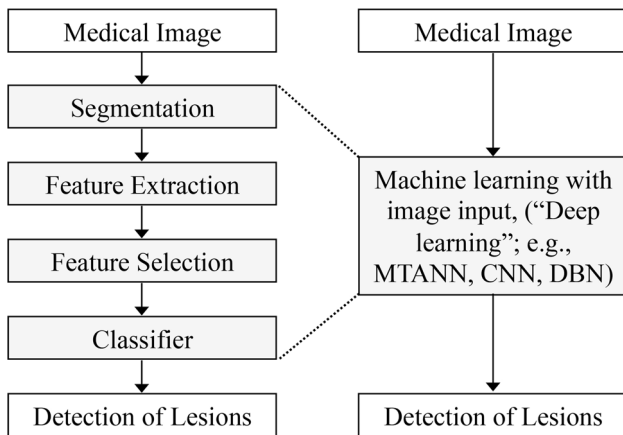
Fig. 2 Changes in ML approaches before and after the introduction of ML with image input (image-based ML) including "deep learning". Compared to ML with feature input (feature-based ML, or a classifier with features), ML with image input including deep learning skips steps of segmentation of objects, feature extraction from the segmented objects, and feature selection for determining "effective features", which offers an end-to-end ML paradigm

the "deep learning" that people use right now would be ML with image input or image-based ML [23]. The depth of MLs is still an important attribute that determines the characteristics or properties of ML models or applications. When the architecture is deep, the ML model should be called deep ML with image input (image-based ML) or deep ML with feature input (feature-based ML).

2.3 History of ML in computer vision and medical imaging fields

Figure 3 summarizes the history of ML in the fields of computer vision and medical imaging. Before the

popularity of "deep learning" starting in 2013, ML with feature input (feature-based ML) as dominant in their fields. Before 1980, even when the term "machine learning" did not exist, classical classifiers such as LDA, QDA, and a k-nearest neighbor classifier (k-NN) were used for classification. In 1986, MLP was proposed by Rumelhart and Hinton [45]. The MLP created the 2nd neural network (NN) research boom (the 1st one was in 1960s). In 1995, Vapnik proposed an SVM [46] and became the most popular classifier for a while, partially because of publicly available code on the Internet. Various ML methods were proposed, including random forests by Ho et al. in 1995 [48], and dictionary learning by Mairal et al. in 2009 [49]. On the other hand, various ML with image input (image-based ML) techniques were proposed before the introduction of the term "deep learning". It started from the Neocognitron by Fukushima in 1980 [50]. In 1989, LeCun et al. simplified the Neocognitron and proposed a CNN [51], but he did not study CNNs very much until recently. In 1994, Suzuki et al. applied an MLP to cardiac images in a convolutional way [52]. Two years later, Suzuki et al. proposed neural filters based on a modified MLP to reduce noise in images [53], and in 2000, they proposed neural edge enhancers [54]. Suzuki et al. proposed MTANN for classification of patterns in 2003 [22], detection of objects in 2009 [26], separation of specific patters from other patterns in X-ray images in 2006 [31], and reduction of noise and artifacts on CT images in 2013 [55]. Hinton et al. proposed a deep brief network (DBN) in 2006 [47], and they created the term "deep learning" a year later. Deep learning was not recognized much until late 2012. In late 2012, a CNN won in the ImageNet competition [42]. Among them, Neocognitron, MLP, CNN, neural filters, MTANN, and DBN are capable of deep architecture. Thus,

Fig. 3 The history of ML in the fields of computer vision and medical imaging. There are two distinct ML approaches in these fields. Before the popularity of "deep learning" starting in 2013, ML with feature input (feature-based ML) was dominant in the fields. After that, ML with image input (image-based ML) including deep learning gained enthusiastic popularity, but it has a long history
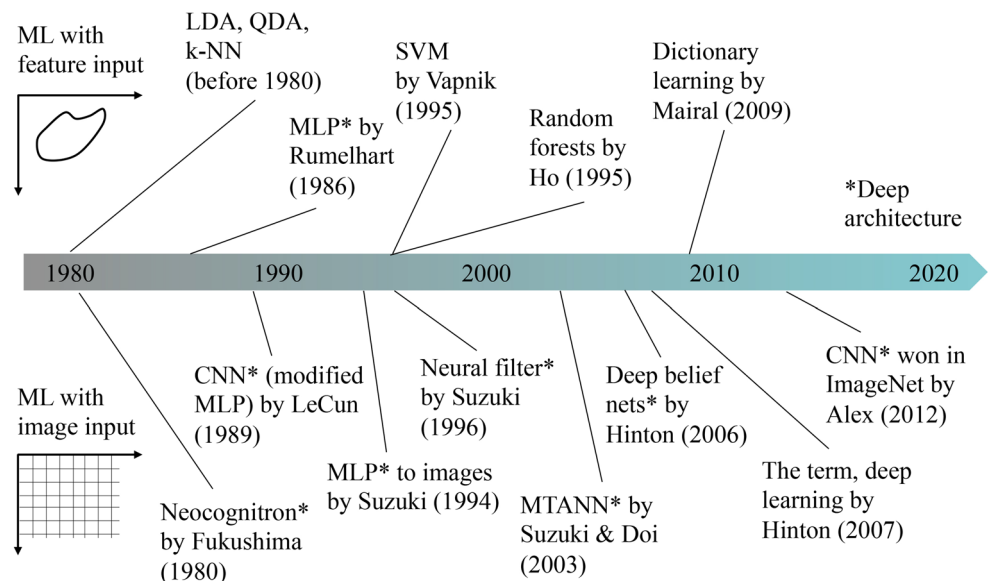
at present the term "deep learning", which is ML with image input (image-based ML) with deep architecture, to be accurate, does not offer new ML models, but rather it is essentially a collection of earlier work on ML (namely, ML with image input) that was recently recognized again with a different terminology.

3 Two major models in the class of ML with image input

3.1 Massive-training artificial neural network (MTANN)

In the field of signal and image processing, supervised nonlinear filters based on an MLP model (or a multilayer NN), called neural filters, have been proposed [56, 57]. The neural filter employs a linear-output-layer NN model as a convolution kernel of a filter. The inputs to the neural filter are an object pixel and spatially/spatiotemporally adjacent pixels in a subregion (or local window, image patch, kernel). The output of the neural filter is a single pixel. The neural filter is trained with input images and corresponding "teaching" (desired or ideal) images. The class of neural filters is used for image processing tasks such as edge-preserving noise reduction in radiographs and other digital pictures [56, 57], edge enhancement from noisy images [58], and enhancement of subjective edges traced by a physician ("semantic segmentation") in left ventriculograms [59].

An MTANN was developed by extending of neural filters to accommodate various pattern-recognition tasks [22], including classification [22, 28, 31, 35–40], pattern enhancement and suppression [31], and object detection [26]. In other words, neural filters are a subclass or a special case of MTANNs. A two-dimensional (2D) MTANN was first developed for distinguishing a specific pattern from other patterns in 2D images [22]. The 2D MTANN was applied to reduction of FPs in CAD for detection of lung nodules on 2D CT slices in a slice-by-slice way [22, 24, 25], and in chest radiographs (CXR) [28], the separation of bones from soft tissue in CXR [30–32], and the distinction between benign and malignant lung nodules on 2D CT slices [35]. 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 [36–40]. Various MTANN architectures were developed, including multiple MTANNs [22, 24, 28, 35, 56, 57], a mixture of expert MTANNs [36, 37], a multi-resolution MTANN [31], a Laplacian eigenfunction MTANN [40], as well as a massive-training support vector regression (MTSVR) and a massive-training nonlinear Gaussian process regression [39].

The general architecture of an MTANN is illustrated in Fig. 4. An MTANN consists of an ML model such as linear-output-layer artificial NN (ANN) regression, support vector regression [46, 60], and nonlinear Gaussian process regression, which is capable of operating on pixel data directly [58]. The core part of the MTANN consists of an input layer, multiple hidden layers, and an output layer, as illustrated in Fig. 4a. 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 an ANN were improved significantly with a linear function when it was applied to the continuous mapping of values in image processing [58]. Note that the activation functions of the units in the hidden layers are a sigmoid function for non-linear 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

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

where x, y, and z are the coordinate indices, ML(.) is the output of the ML model, and $I(x,y,z)$ is a pixel value of the input image. The structure of input units and the number of hidden units in the ANN may be designed by use of sensitivity-based unit-pruning methods [61, 62]. ML regression models rather than ML classification models 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). The entire output image is obtained by scanning with the input subregion of the MTANN in a convolutional manner on the entire input image, as illustrated in Fig. 4b. 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 CNNs.

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, as shown in Fig. 4c. A score for a given image is defined as

$$S = \sum_{(x,y,z) \in R_{\text{E}}} f_{\text{w}}(x, y, z) \times O(x, y, z), \quad (2)$$

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

Fig. 4 Architecture of an MTANN consisting of an ML model (e.g., linear-output-layer ANN regression) with subregion (or image patch, local kernel) input and single-pixel output. **a** Architecture of the core part of the MTANN. **b** Entire architecture of the MTANN for predicting a likelihood map. The entire output image representing a likelihood map is obtained by scanning with the input subregion of the MTANN in a convolutional manner on the entire input image. **c** The MTANN for classification. A scoring layer is placed in the end to convert the output likelihood map into a single score that represents the likelihood of being a certain class for a given input image
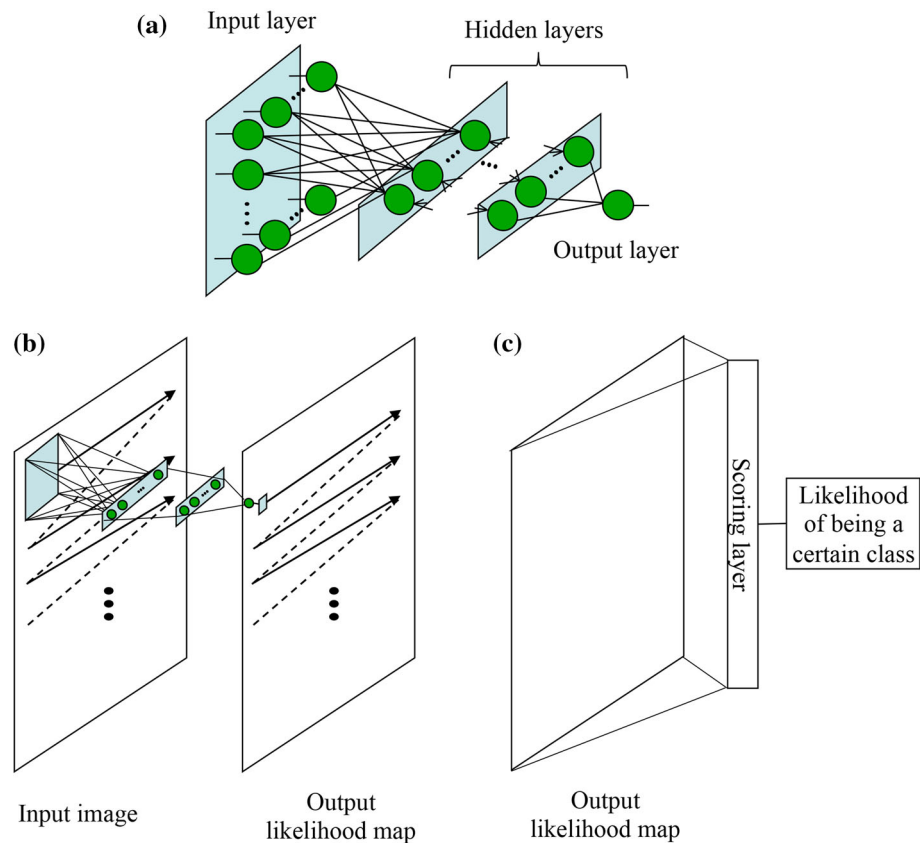
score represents the weighted sum of the estimates for the likelihood that the image (e.g., a lesion candidate) contains an object of interest (e.g., a lesion) near the center, i.e., a higher score would indicate an object (e.g., a lesion) of interest, and a lower score would indicate other patterns. Thresholding is then performed on the scores for distinction between classes.

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

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

For enrichment of training samples, a training region, R_T, extracted from the input images is divided pixel by pixel into a large number of overlapping subregions. 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." The error to be minimized by training of the MTANN is defined as:

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

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 (BP) algorithm [58] which was derived for the linear-output-layer ANN model by use of the generalized delta rule [45]. After training, the MTANN is expected to output the highest value when an object of interest is located at the center of the subregion of the MTANN, a lower value as the distance from the subregion center increases, and zero when the input subregion contains other patterns.

3.2 Convolutional neural networks (CNNs)

A CNN can be considered as a simplified version of the Neocognitron model that was proposed to simulate the human visual system in 1980 by Fukushima [50]. LeCun et al. developed a CNN called LeNet for handwritten ZIP-code recognition [51]. The LeNet has 5 layers: 1 input layer, 3 hidden layers, and 1 output layer. The input layer has a small 16×16 pixel image. The 3 hidden layers consist of 2 convolutional layers and 1 fully connected layer. The architecture of a general CNN is illustrated in Fig. 5. The input to the CNN is an image, and the outputs

are class categories such as cancer or non-cancer. The layers are connected with local shift-invariant inter-connections (or convolution with a local kernel). Unlike the Neocognitron, the CNN has no lateral inter-connections or feedback loops, and the BP algorithm [45] is used for training. Each unit (neuron) in a subsequent layer is connected with the units of a local region in the preceding layer, which offers the shift-invariant property; in other words, forward data propagation is similar to a shift-invariant convolution operation. The data from the units in a certain layer are convolved with the weight kernel, and the resulting value of the convolution is collected into the corresponding unit in the subsequent layer. This value is processed further by the unit through an activation function and produces an output datum. The activation function between two layers is a nonlinear or quasi-nonlinear function such as a rectified linear function and a sigmoid function. As layers go deeper (close to the output layer), the size of the local region in a layer is reduced in a pooling layer. In the pooling layer, the pixels in the local region are sub-sampled with a maximum operation. For deriving the training algorithm for the CNN, the generalized delta rule [45] is applied to the architecture of the CNN. For distinguishing an image containing an object of interest from an image without it, a class label for the object, the number 1, is assigned to the corresponding output unit, and zeros to other units. A softmax function is often used in the output layer, called a softmax layer.

3.3 Comparisons between the two "deep learning" models

3.3.1 Architecture

CNNs and MTANNs are in the class of ML with image input (image-based ML) or "deep learning". Both models use pixel values in images directly as input information,

instead of features calculated from segmented objects; thus, they can be classified as end-to-end ML models that do the entire process from input images to the final classification. Both models can have deep layers (>4 layers). There are major differences between CNNs and MTANNs in (1) architecture, (2) output, and (3) the required number of training samples. (1) In CNNs, convolutional operations are performed within the network, whereas the convolutional operation is performed outside the network in MTANNs, as shown in Figs. 4 and 5. (2) The output of CNNs consists, in principle, of class categories, whereas that of MTANNs consists of images (continuous values in a map). (3) Another major difference is the required number of training samples. CNNs require a huge number of training images (e.g., 1,000,000 images) because of a large number of parameters in the model, whereas MTANNs require a very small number of training images (e.g., 12 images for classification between lung nodules and non-nodules in CAD for detection of lung nodules in CT [22], and 4 images for separation of bone components from soft tissue components in CXR [30, 31]).

3.3.2 Performance

The performance of well-known CNNs (including the AlexNet, the LeNet, a relatively deep CNN (rd-CNN), a shallow CNN (sh-CNN), and a fine-tuned AlexNet (Fine-Tuned AlexNet) which used transfer learning from a computer-vision-trained AlexNet) and MTANNs was compared extensively in focal lesion detection and classification problems in medical imaging [63]. Comparison experiments were done for detection of lung nodules and classification of detected lung nodules into benign and malignant in CT with the same databases. The experiments demonstrated that the performance of MTANNs was substantially higher than that of the best-performing CNN under the same condition, as demonstrated in Fig. 6. With a

Fig. 5 Architecture of a CNN. The layers in the CNN are connected with local shift-invariant inter-connections (or convolution with a local kernel). The input and output of the CNN are images and class labels (e.g., Class A and Class B), respectively
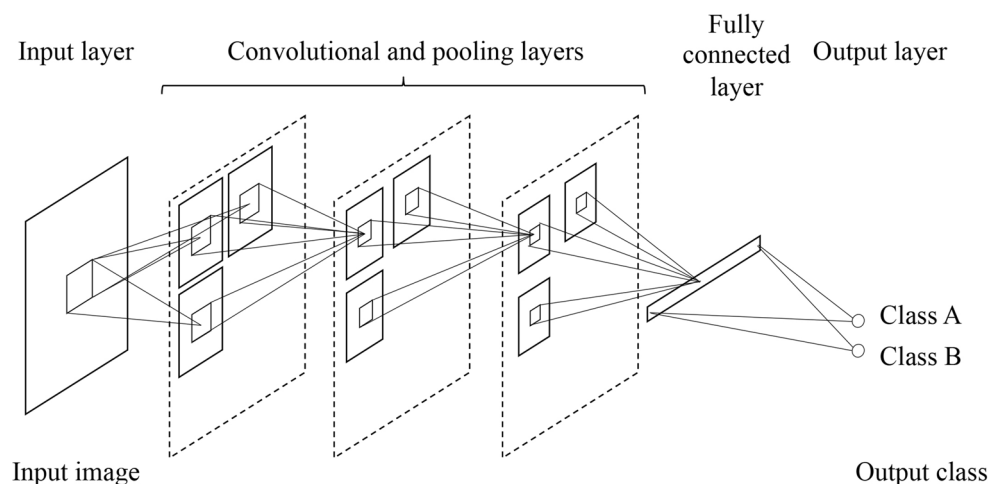
larger training dataset used only for CNNs, the performance gap became less evident, even though the margin was still significant. Specifically, for nodule detection, MTANNs generated 2.7 FPs per patient at 100% sensitivity, which was significantly ($p < 0.05$) lower than that for the best-performing CNN model (FineTuned AlexNet), with 22.7 FPs per patient at the same level of sensitivity. For nodule classification, MTANNs yielded an area under the receiver-operating-characteristic curve (AUC) of 0.881, which was significantly ($p < 0.05$) greater than that for the best-performing CNN model, with an AUC of 0.776.

4 Applications of ML to medical imaging

4.1 Applications of ML with feature input

There has been a large number of papers that reported applications of ML with feature input (feature-based ML, common classifiers) in medical imaging, such as applications to lung nodule detection in CXR [64–67] and thoracic CT [24, 68–70], classification of lung nodules into benign or malignant in CXR [71] and thoracic CT [72, 73], detection of microcalcifications in mammography [74–77], detection of masses [78] and classification of masses into benign or malignant [79–81] in mammography, polyp detection in CT colonography [82–84], and detection of aneurysms in brain MRI [85]. There are applications of ML for regression problems such as determining subjective similarity measure of mammographic images [86–88]. A survey of studies on ML in computer-aided diagnosis has been reported [8], and a collection of studies on ML in computer-aided diagnosis and medical image analysis can be found in books [6, 20].
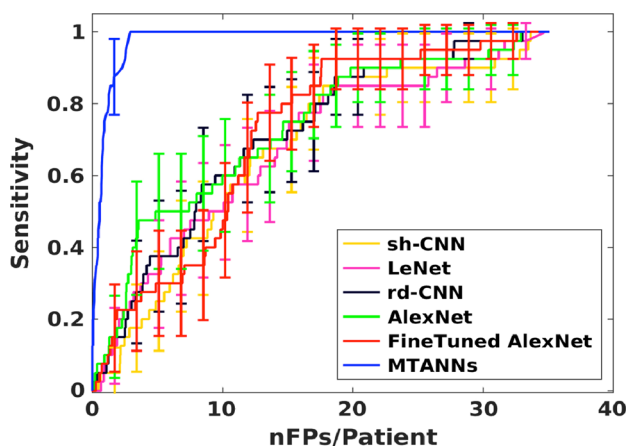
Fig. 6 Comparative evaluation of the performance of MTANNs and major CNN models in the classification between lung nodules and non-nodules in CADe in CT under the same conditions with the same database

4.2 Applications of ML with image input ("deep learning")

4.2.1 Classification between lesions and non-lesions

The class of "deep" MTANNs with 4–7 layers has been used for classification, such as FP reduction in CAD schemes for detection of lung nodules in CXR [28] and CT [22, 24, 25], and FP reduction in a CAD scheme for polyp detection in CT colonography [36–40]. Figure 7a shows a schematic diagram of an MTANN for classification between lesions and non-lesions in medical images. For enhancement of lesions and suppression of non-lesions, the teaching image contains a distribution of values that represent a map for the probability of being a lesion. For example, the teaching volume contains a 3D Gaussian distribution with standard deviation for a lesion and zero (i.e., completely dark) for non-lesions, as illustrated in Fig. 7a. After training, a scoring method is used for combining of output voxels from the trained MTANNs, as illustrated in Fig. 7b. Thresholding is then performed on the scores for distinction between lesions and non-lesions. Figure 8 shows the output images of the MTANN trained to enhance lung nodules and suppress various types of non-nodules in CAD for CT. Various lung nodules are enhanced in the MTANN output images, whereas various types of non-nodules are suppressed. Nodules such as a solid nodule, a part-solid (mixed-ground-glass) nodule, and a non-solid (ground-glass) nodule are enhanced, whereas non-nodules such as different-sized lung vessels and soft tissue opacity are suppressed around the centers of ROIs. With those nodule-enhanced images, distinction between nodules and non-nodules is made by use of the scoring method described above. In other words, classification between a particular pattern and other patterns is made by enhancement of the particular pattern, which may be referred to as "classification by enhancement".

Before the introduction of the term deep learning, "deep" CNNs had been used for FP reduction in CAD for lung nodule detection in CXRs [89–91]. A convolution NN was trained with 28 CXRs for distinguishing lung nodules from non-nodules (i.e., FPs produced by an initial CAD scheme). The trained CNN reduced 79% of FPs (which is equivalent to 2–3 FPs per patient), whereas 80% of true-positive detections were preserved. CNNs have been applied to FP reduction in CAD for detection of microcalcifications [92] and masses [93] in mammography. A CNN was trained with 34 mammograms for distinguishing microcalcifications from FPs. The trained CNN reduced 90% of FPs, which resulted in 0.5 FPs per image, whereas a true-positive detection rate of 87% was preserved [92]. Shift-invariant NNs which are almost identical to CNNs, have been used for FP reduction in CAD for

Fig. 7 a Training of an MTANN for distinction between lesions and non-lesions in a CAD scheme for detection of lesions in medical images. The teaching image for a lesion contains a Gaussian distribution that for a non-lesion contains zero (completely dark). After the training, the MTANN expects to enhance lesions and suppress non-lesions. **b** Scoring method for combining pixel-based output responses from the trained MTANN into a single score for each ROI
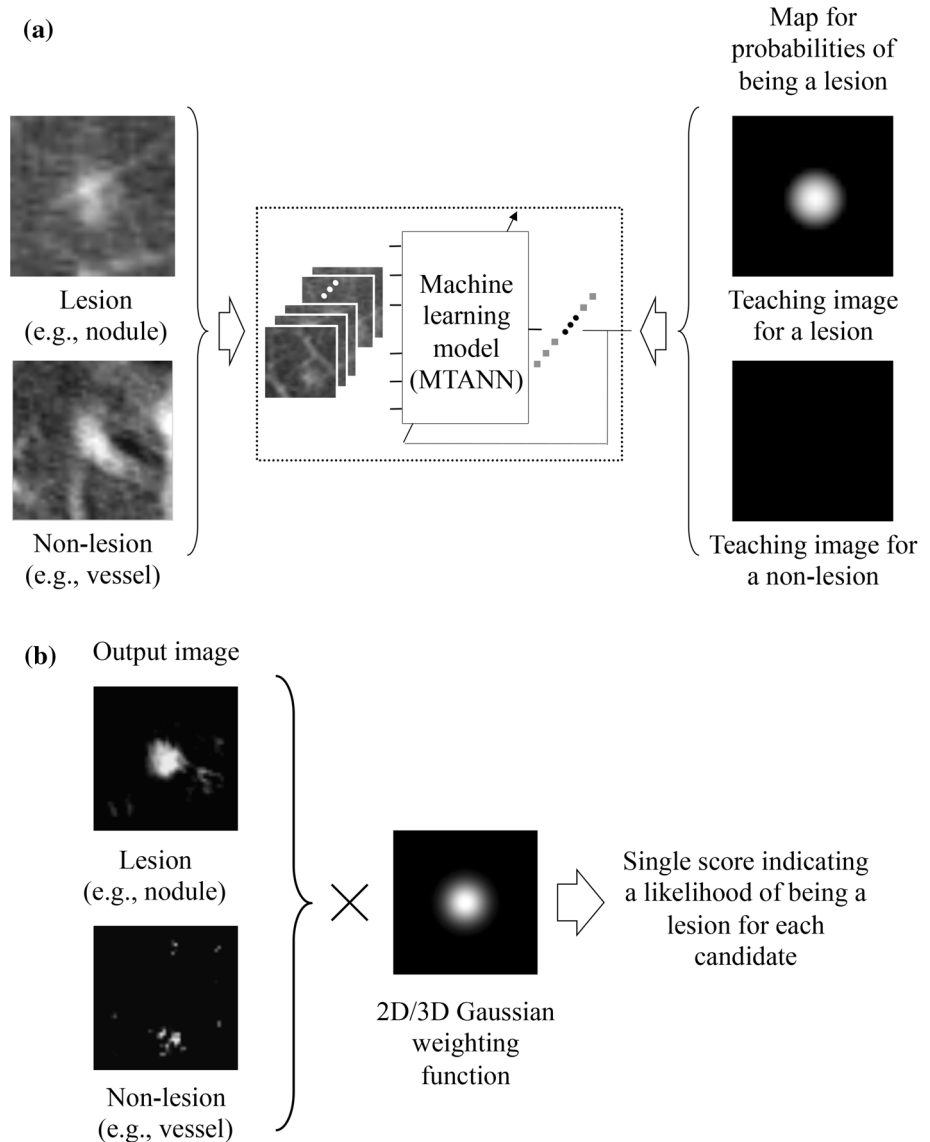
detection of microcalcifications [94, 95]. A shift-invariant NN was trained to detect microcalcifications in ROIs. Microcalcifications were detected by thresholding of the output images of the trained shift-invariant NN. When the number of detected microcalcifications was greater than a predetermined number, the ROI was considered as a microcalcification ROI. With the trained shift-invariant NN, 55% of FPs was removed without any loss of true positives.
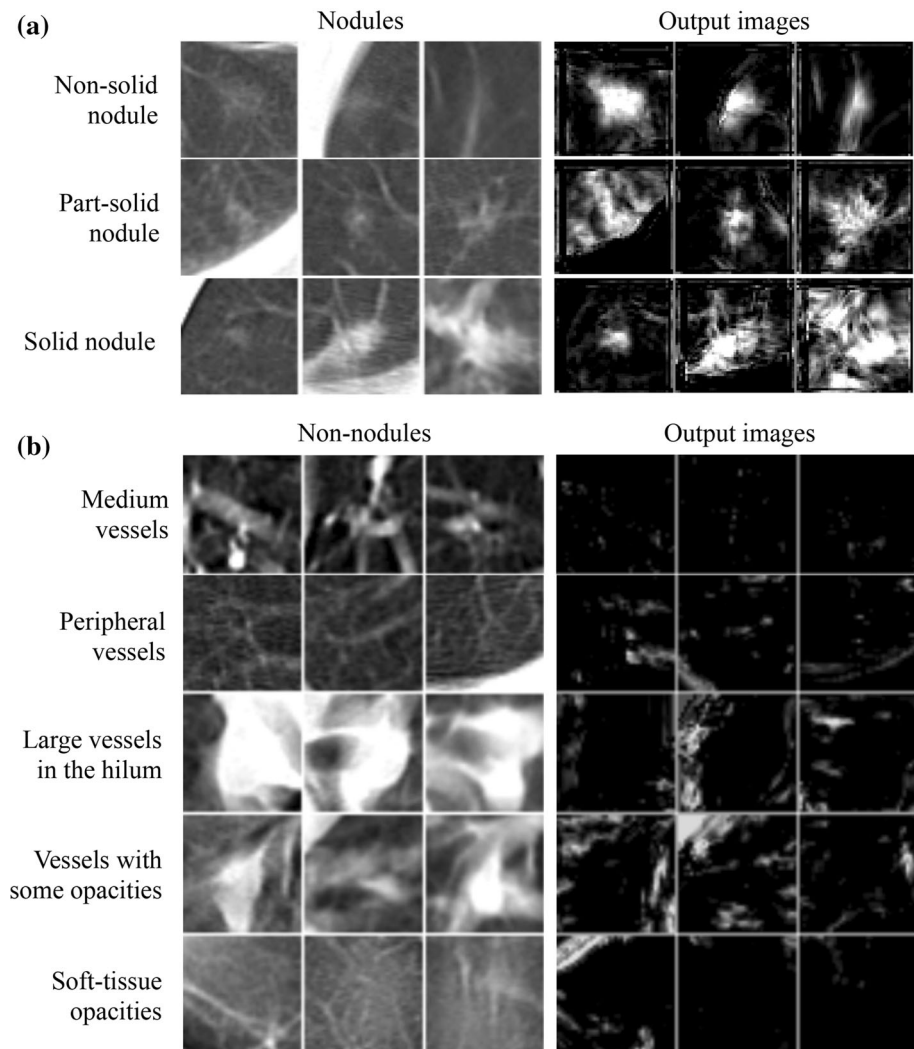
After the introduction of the term "deep learning", a CNN was used for classification of masses and non-masses in digital breast tomosynthesis images [96]. The CNN for digital breast tomosynthesis was trained by use of transfer learning from the CNN for mammography. The CNN achieved an AUC of 0.90 in the classification of mass ROIs and non-mass ROIs in digital breast tomosynthesis images. A CNN was used for FP reduction in lung nodule detection

in PET/CT [97]. The CNN was used for feature extraction, and classification was done by SVM with the CNN-extracted and hand-crafted features. With the FP reduction method, the performance was improved from a sensitivity of 97.2% with 72.8 FPs/case to a sensitivity of 90.1% with 4.9 FPs/case. Because there are a growing number of papers for applications of CNNs in this area, those papers are not reviewed in this paper.

4.2.2 Classification of lesion types

Before the introduction of the term "deep learning", "deep" MTANNs with seven layers were applied to the distinction between benign and malignant lung nodules in low-dose screening CT [35]. The MTANNs achieved an AUC value of 0.882 in the classification between 76 malignant and 413 benign lung nodules, whereas an AUC

Fig. 8 Illustrations of various types of nodules and non-nodules and corresponding output images of the trained MTANN in CAD for detection of lung nodules in CT. **a** Results for various types of nodules. **b** Results for various types of non-nodules. Nodules are represented by bright pixels, whereas non-nodules are almost dark around the centers of ROIs

value for chest radiologists for the same task with a subset of the database was 0.56.

After the introduction of the term "deep learning", a CNN was used for classification between perifissural nodules and non-perifissural nodules in CT [98]. A pre-trained 2D CNN was used. The CNN achieved a performance in terms of AUC of 0.868. A pre-trained CNN was used for classification between cysts from soft tissue lesions in mammography [99]. The CNN achieved an AUC value of 0.80 in the classification between benign solitary cysts and malignant masses. A CNN was used for classification of plaque compositions in carotid ultrasound [100]. CNN's classification achieved a correlation value of about 0.90 with the clinical assessment for the estimation of lipid-core, fibrous-cap, and calcified-tissue areas in carotid ultrasound. A CNN was used for classifying of tooth types in cone-beam CT [101]. The CNN achieved a classification accuracy of 88.8% in classification of 7 tooth types in ROIs.

4.2.3 Detection of lesions

A "lesion-enhancement" filter-based MTANN was developed for enhancement of actual lesions in CAD for detection of lung nodules in CT [26]. For enhancement of lesions and suppression of non-lesions in CT images, the teaching image contained a probability map for being a lesion. For enhancement of a nodule in an input CT image, a 2D Gaussian distribution was placed at the location of the nodule in the teaching image, as a model of the lesion probability map. For testing of the performance, the trained MTANN was applied to non-training lung CT images. As shown in Fig. 9, the nodule is enhanced in the output image of the trained MTANN filter, whereas normal structures such as lung vessels are suppressed. After large and small remaining regions were removed by use of area information obtained with connected-component labeling [102–104], accurate nodule detection was achieved with no

FP, which means that one MTANN functions as a complete CAD scheme with high accuracy.

After the introduction of the term "deep learning", deep CNNs were used for detection of lymph nodes in CT [105]. Detection of lymph nodes is a challenging task, as evidenced by the fact that ML with feature input (feature-based ML) achieved approximately 50% sensitivity with 3 FPs/volume. With use of deep CNNs, the performance reached at 70 and 83% sensitivities with 3 FPs/volume in the mediastinum and abdomen areas, respectively.

4.2.4 Segmentation of lesions or organs

Neural edge enhancers (NEEs; predecessor of MTANNs) enhanced subjective edges traced by a physician ("semantic segmentation") in left ventriculograms [59], as illustrated in Fig. 10. As seen, the edge enhancement performance of the NEE was superior to that of the Marr–Hildreth operator in this challenging segmentation problem. The segmentation by the NEE agreed excellently with the "gold-standard" manual segmentation by an experienced cardiologist.

Shift-invariant NNs were used for detection of the boundaries of the human corneal endothelium in photomicrographs [106]. In addition, a CNN was used for segmentation of the bladder in CT urography [107]. The CNN achieved a Jaccard index of $76.2 \pm 11.8\%$ for bladder segmentation, compared with "gold-standard" manual segmentation. A CNN was used for segmentation of tissues in MR brain images [108]. The CNN achieved average Dice coefficients of 0.82–0.91 in five different datasets.

4.2.5 Separation of bones from soft tissue in CXR

Studies showed that 82–95% of the lung cancers missed by radiologists in CXR were partly obscured by overlying bones such as ribs and/or a clavicle [109, 110]. To prevent such misses, MTANNs were developed for separation of bones from soft tissues in CXR [30, 31]. To this end, the MTANNs were trained with input CXRs with overlapping bones and the corresponding "teaching" dual-energy bone images acquired with a dual-energy radiography system [111]. Figure 11 shows a non-training original CXR and a "virtual" dual-energy soft tissue image obtained by use of the trained MTANN. The contrast of ribs is suppressed substantially in the MTANN soft tissue image, whereas the contrast of soft tissue such as lung vessels is maintained. A filter learning in the class of ML with image input (image-based ML) was developed for suppression of ribs in CXR [112].

4.2.6 Analysis of a trained ML model

Some researchers refer to a trained NN as a "black box", but there are ways to analyze or look inside a trained NN. With such methods, trained NNs are not "black boxes". Analysis of a trained ML model is very important for revealing what was trained in the trained ML model. Suzuki et al. analyzed an NEE that was trained to enhance edges from noisy images [58]. Figure 12a shows the receptive field of the trained NEE [61] that was revealed by the application of a method for designing the optimal structure of an NN [62] to the trained NEE. The receptive field shows which input pixels were used for enhancement of edges from noisy images. Furthermore, they analyzed

(a)   **(b)**   **(c)**
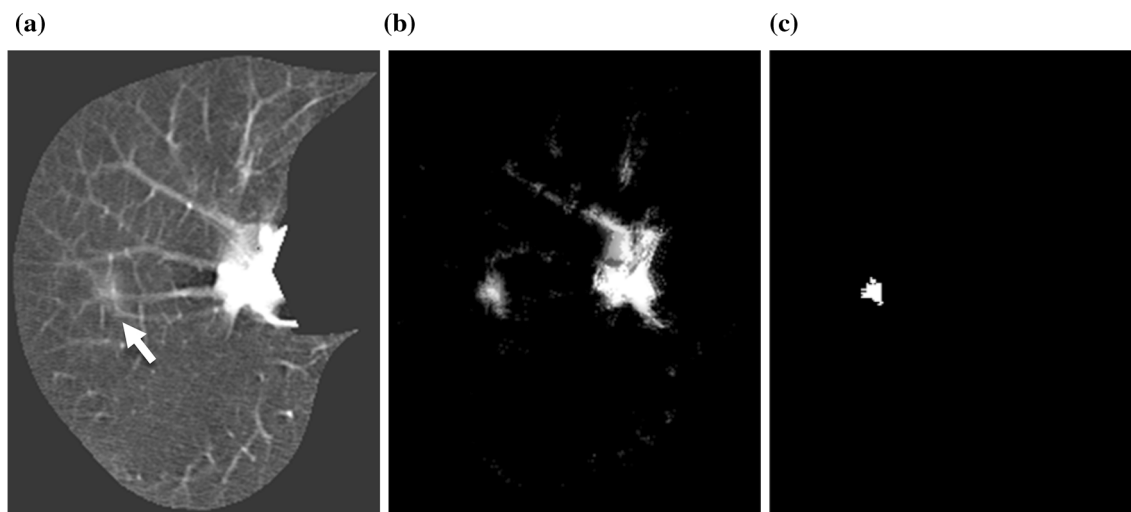
**Fig. 9** Enhancement of a lunge nodule by use of a trained lesion-enhancement MTANN filter for a non-training case. **a** Original chest CT image of the segmented lung with a nodule (indicated by an arrow). **b** Output image of the trained lesion-enhancement MTANN filter. **c** Result after thresholding and removal of large and small regions by area, which achieves accurate segmentation and detection of the lung nodule

(a)   **(b)**

(c)   **(d)**
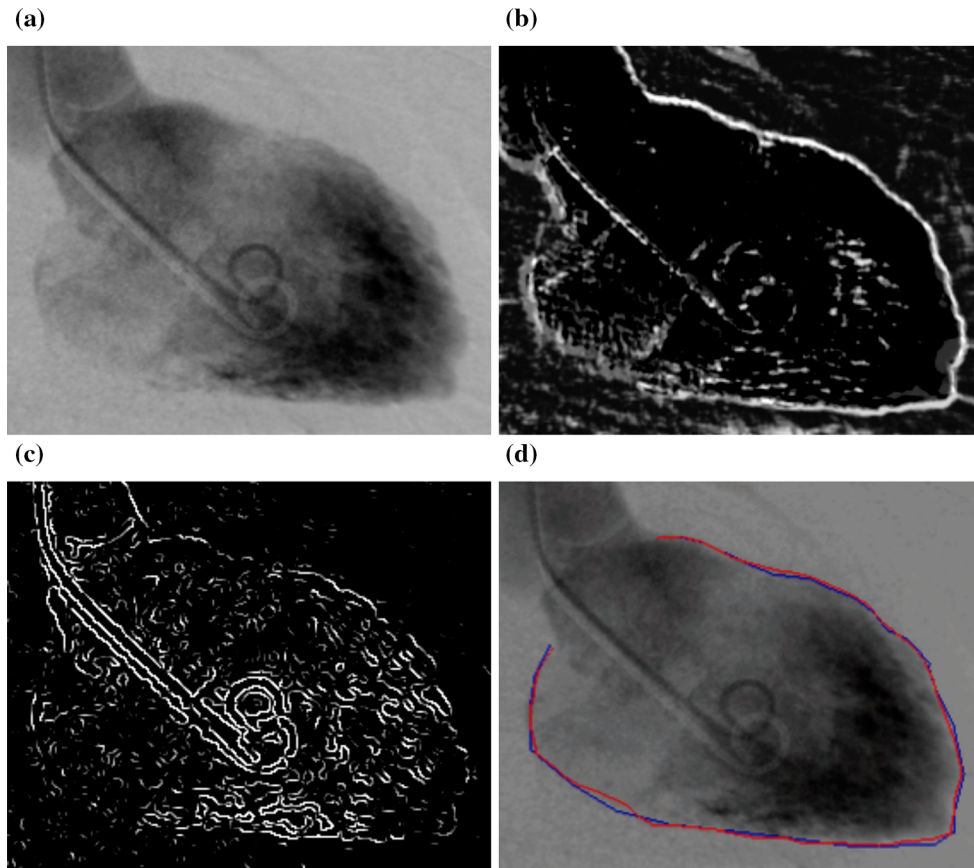
Fig. 10 Segmentation of the left vernicle in a left ventriculogram. **a** Input left ventriculogram. This segmentation is a very challenging problem, because some portions of the boundary of the left ventricle have of low-contrast, less edge contrast, and fuzzy borders. **b** Edge enhancement by the trained NEE. The boundary of the left ventricle is enhanced well even for challenging very subtle edges. **c** Edges enhanced by the Marr–Hildreth operator. Some boundary portions are missing in the result. **d** Comparison between segmentation by our NEE and "gold-standard" manual segmentation by a cardiologist. The two segmentations agree very well

(a)   **(b)**
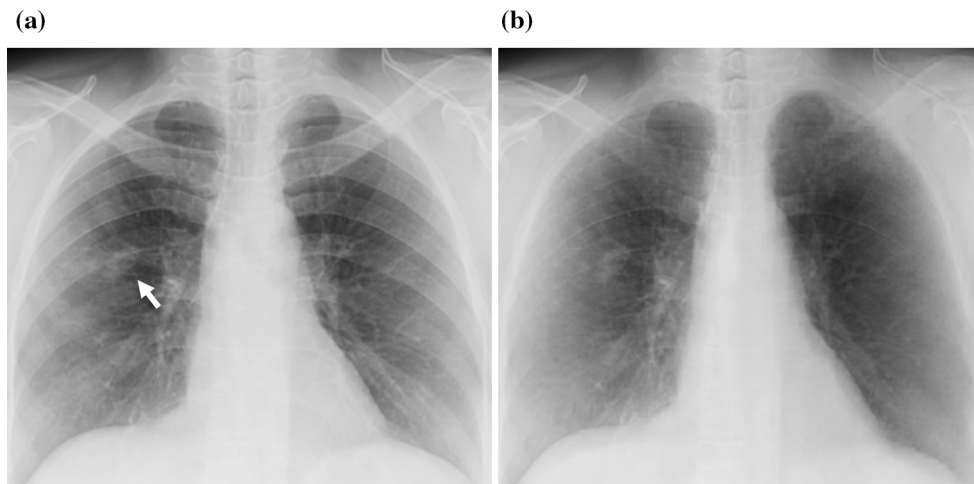
Fig. 11 Separation of bone components from soft tissue components in CXR by use of an MTANN. **a** Input CXR with a nodule (indicated by an *arrow*). **b** "Virtual dual-energy" soft-tissue image: Result of application of the trained MTANN

the units in the hidden layer of the trained NEE. Figure 12b shows the analysis results of the internal representation of the trained NEE, showing the strength of the weights from each hidden unit to the input units. The black square indicates a pixel having a negative weight. The pixels having the same sign correspond to a smoothing operation,

whereas the pixels having the opposite sign correspond to an edge enhancement operation. It is interesting to note that Fig. 12b is reminiscent of the receptive fields of various simple units in the cat and monkey cerebral cortex discovered by Hubel and Wiesel [113]. The weights in the figure indicate the operations for diagonal edge enhancement together with smoothing in hidden unit no. 1, vertical edge enhancement together with horizontal smoothing in hidden unit no. 2, and edge enhancement with smoothing for another diagonal orientation in hidden unit no. 3. The results of the analysis suggest that the trained NEE uses directional gradient operators with smoothing. These directional gradient operators with smoothing, followed by integration with nonlinearity, lead to robust edge enhancement against noise.

5 Advantages and limitations of "deep learning"

As described earlier, the major difference between ML with image input (image-based ML) including "deep learning" and ML with feature input (feature-based ML,

common classifiers) is the direct use of pixel values with the ML model. In other words, unlike ordinary classifiers (ML with feature input), feature calculation from segmented objects is not necessary. Because the ML with image input can avoid errors caused by inaccurate feature calculation and segmentation, the performance of the ML with image input can be higher than that of ordinary feature-based classifiers. ML with image input learns pixel data directly, and thus all information on pixels should not be lost before the pixel data are entered into the ML with image input, whereas ordinary feature-based classifiers learn the features extracted from segmented lesions and thus important information can be lost with this indirect extraction; also, inaccurate segmentation often occurs for complicated patterns. In addition, because feature calculation is not required for the ML with image input, development and implementation of segmentation and feature calculation, and selection of features are unnecessary; this offers fast and efficient development.

The characteristics of the ML with image input which uses pixel data directly would generally differ from those of ordinary feature-based classifiers (ML with feature input).
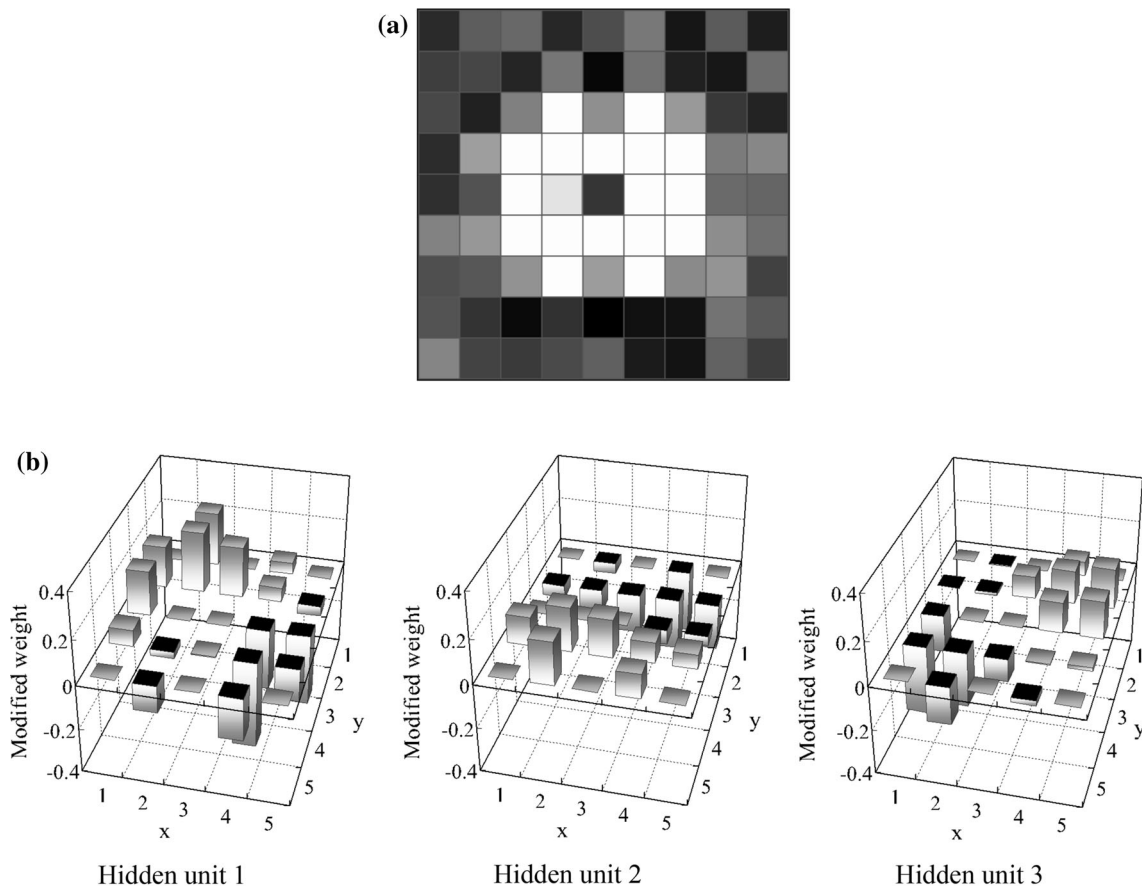
Fig. 12 Analysis of the NEE trained to enhance edge from noisy images. **a** The receptive field of the trained NEE. **b** Analysis results of the internal representation of the trained NEE. The weights from each hidden unit to the input units are shown

Therefore, combining an ordinary feature-based classifier with ML with image input would yield a higher performance than that of a classifier alone or ML with image input alone. Indeed, in previous studies, both classifier and ML with image input were used successfully for classification of lesion candidates into lesions and non-lesions.

Limitations of "deep" CNNs (in ML with image input) include (1) a very high computational cost for training because of the high dimensionality of input data and (2) the required large number of training images. Because "deep" CNNs use pixel data in images directly, the number of input dimensions is generally large. A CNN requires a huge number of training images (e.g., 1,000,000) for determining a large number of parameters in the CNN. However, an MTANN requires a small number of training images (e.g., 20) because of its simpler architecture. With GPU implementation, an MTANN completes training in a few hours, whereas a deep CNN takes several days.

6 Conclusion

This paper provides an overview of the area of deep learning in medical imaging is overviewed, including (1) what was changed in machine learning before and after the introduction of deep learning, (2) what is the source of the power of deep learning, (3) two major deep-learning models: an MTANN and a CNN, (4) similarities and differences between the two models, and (5) their applications to medical imaging. This review reveals that ML with feature input (or feature-based ML) was dominant before the introduction of deep learning, and that the major and essential difference between ML before and after deep learning is learning image data directly without object segmentation or feature extraction; thus, it is the source of the power of deep learning, although the deepness of the model is an important attribute. There is a long history of deep-learning techniques, including the Neocognitron, CNNs, neural filters, and MTANNs in the class of ML with image input, except a new term, "deep learning". ML with image input including deep learning is a very powerful, versatile technology with higher performance, which can bring the current state-of-the-art performance level of medical image analysis to the next level. ML including deep learning in medical imaging is an explosively growing, promising field. It is expected that ML with image input including deep learning will be the mainstream technology in medical imaging in the next few decades.

Compliance with ethical standards

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the Institutional Review Board (IRB) and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the studies. The studies complied with the Health Insurance Portability and Accountability Act (HIPAA).

References

1. Suzuki K. Machine learning for medical imaging, vol. 3. Algorithms. 2010. http://www.mdpi.com/journal/algorithms/special_issues/machine-learning-for-medical-imaging. Accessed 5 July 2017.
2. Wang F, Yan P, Suzuki K, Shen D, eds. Machine learning in medical imaging (MLMI), vol. 6357. Lecture notes in computer science. Berlin: Springer; 2010.
3. Suzuki K, Wang F, Shen D, Yan P, eds. Machine learning in medical imaging (MLMI), vol. 7009. Lecture notes in computer science. Berlin: Springer; 2011.
4. Suzuki K. Machine learning for medical imaging 2012, vol. 5. Algorithms. 2012. http://www.mdpi.com/journal/algorithms/special_issues/medical_imaging_2012. Accessed 5 July 2017.
5. Suzuki K, Yan P, Wang F, Shen D. Machine learning in medical imaging. Int J Biomed Imaging. 2012;2012:123727.
6. Suzuki K. Machine learning in computer-aided diagnosis: medical imaging intelligence and analysis. Hershey: IGI Global; 2012.
7. Wang F, Shen D, Yan P, Suzuki K, editors. Machine learning in medical imaging (MLMI), vol. 7588. Lecture notes in computer science. Berlin: Springer; 2012.
8. Suzuki K. Machine learning in computer-aided diagnosis of the thorax and colon in CT: a survey. IEICE Trans Inf Syst. 2013;E96-D(4):772–83.
9. Wu G, Zhang D, Shen D, Yan P, Suzuki K, Wang F, editors. Machine learning in medical imaging (MLMI), vol. 8184. Lecture notes in computer science. Berlin: Springer; 2013.
10. Yan P, Suzuki K, Wang F, Shen D. Machine learning in medical imaging. Mach Vision Appl. 2013;24(7):1327–9.
11. Shen D, Wu G, Zhang D, Suzuki K, Wang F, Yan P. Machine learning in medical imaging. Comput Med Imaging Graph. 2015;41:1–2.
12. Suzuki K, Zhou L, Wang Q. Machine learning in medical imaging. Pattern Recognit. 2017;63:465–7.
13. El-Baz A, Gimel'farb G, Suzuki K. Machine learning applications in medical image analysis. Comput Math Methods Med. 2017;2017:2.
14. Doi K. Overview on research and development of computer-aided diagnostic schemes. Semin Ultrasound CT MRI. 2004;25(5):404–10.
15. Doi K. Current status and future potential of computer-aided diagnosis in medical imaging. Br J Radiol. 2005;78(Spec No 1):S3–19.

16. Doi K. Diagnostic imaging over the last 50 years: research and development in medical imaging science and technology. Phys Med Biol. 2006;51(13):R5–27.

17. Doi K. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. Comput Med Imaging Graph. 2007;31(4–5):198–211.

18. Lostumbo A, Wanamaker C, Tsai J, Suzuki K, Dachman AH. Comparison of 2D and 3D views for evaluation of flat lesions in CT colonography. Acad Radiol. 2010;17(1):39–47.

19. Lostumbo A, Suzuki K, Dachman AH. Flat lesions in CT colonography. Abdom Imaging. 2010;35(5):578–83.

20. Suzuki K. Computational intelligence in biomedical imaging. New York: Springer; 2014.

21. Duda RO, Hart PE, Stork DG. Pattern recognition. 2nd ed. Hoboken: Wiley Interscience; 2001.

22. Suzuki K, Armato SG 3rd, Li F, Sone S, Doi K. Massive training artificial neural network (MTANN) for reduction of false positives in computerized detection of lung nodules in low-dose computed tomography. Med Phys. 2003;30(7):1602–17.

23. Suzuki K. Pixel-based machine learning in medical imaging. Int J Biomed Imaging. 2012;2012:792079.

24. Arimura H, Katsuragawa S, Suzuki K, et al. Computerized scheme for automated detection of lung nodules in low-dose computed tomography images for lung cancer screening. Acad Radiol. 2004;11(6):617–29.

25. Li F, Arimura H, Suzuki K, et al. Computer-aided detection of peripheral lung cancers missed at CT: ROC analyses without and with localization. Radiology. 2005;237(2):684–90.

26. Suzuki K. A supervised 'lesion-enhancement' filter by use of a massive-training artificial neural network (MTANN) in computer-aided diagnosis (CAD). Phys Med Biol. 2009;54(18):S31–45.

27. Suzuki K, Doi K. 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. 2005;12(10):1333–41.

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

29. Chen S, Suzuki K. Computerized detection of lung nodules by means of "virtual dual-energy" radiography. IEEE Trans Biomed Eng. 2013;60(2):369–78.

30. Suzuki K, Abe H, Li F, Doi K. Suppression of the contrast of ribs in chest radiographs by means of massive training artificial neural network. Paper presented at Proc. SPIE medical imaging (SPIE MI), San Diego, CA, May 2004.

31. Suzuki K, Abe H, MacMahon H, Doi K. Image-processing technique for suppressing ribs in chest radiographs by means of massive training artificial neural network (MTANN). IEEE Trans Med Imaging. 2006;25(4):406–16.

32. Oda S, Awai K, Suzuki K, et al. Performance of radiologists in detection of small pulmonary nodules on chest radiographs: effect of rib suppression with a massive-training artificial neural network. AJR Am J Roentgenol. 2009;193(5):W397–402.

33. Chen S, Suzuki K. 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. 2014;33(2):246–57.

34. Chen S, Zhong S, Yao L, Shang Y, Suzuki K. Enhancement of chest radiographs obtained in the intensive care unit through bone suppression and consistent processing. Phys Med Biol. 2016;61(6):2283–301.

35. Suzuki K, Li F, Sone S, Doi K. 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. 2005;24(9):1138–50.

36. Suzuki K, Rockey DC, Dachman AH. CT colonography: advanced computer-aided detection scheme utilizing MTANNs for detection of "missed" polyps in a multicenter clinical trial. Med Phys. 2010;37(1):12–21.

37. Suzuki K, Yoshida H, Nappi J, Armato SG 3rd, Dachman AH. 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. 2008;35(2):694–703.

38. Suzuki K, Yoshida H, Nappi J, Dachman AH. Massive-training artificial neural network (MTANN) for reduction of false positives in computer-aided detection of polyps: suppression of rectal tubes. Med Phys. 2006;33(10):3814–24.

39. Xu J, Suzuki K. Massive-training support vector regression and Gaussian process for false-positive reduction in computer-aided detection of polyps in CT colonography. Med Phys. 2011;38:1888–902.

40. Suzuki K, Zhang J, Xu J. Massive-training artificial neural network coupled with Laplacian-eigenfunction-based dimensionality reduction for computer-aided detection of polyps in CT colonography. IEEE Trans Med Imaging. 2010;29(11):1907–17.

41. Lawrence S, Giles CL, Tsoi AC, Back AD. Face recognition: a convolutional neural-network approach. IEEE Trans Neural Netw. 1997;8(1):98–113.

42. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Paper presented at advances in neural information processing systems, 2012.

43. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–44.

44. Fukunaga K. Introduction to statistical pattern recognition. 2nd ed. San Diego: Academic Press; 1990.

45. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature. 1986;323:533–6.

46. Vapnik VN. The nature of statistical learning theory. Berlin: Springer; 1995.

47. Hinton G, Osindero S, Teh Y-W. A fast learning algorithm for deep belief nets. Neural Comput. 2006;18(7):1527–54.

48. Ho TK. Random decision forests. Paper presented at document analysis and recognition, 1995, proceedings of the third international conference on, 1995.

49. Mairal J, Bach F, Ponce J, Sapiro G. Online dictionary learning for sparse coding. In: Proceedings of the 26th annual international conference on machine learning, Montreal, Quebec, Canada, 2009.

50. Fukushima K. Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol Cybern. 1980;36(4):193–202.

51. LeCun Y, Boser B, Denker JS, et al. Backpropagation applied to handwritten zip code recognition. Neural Comput. 1989;1(4):541–51.

52. Suzuki K, Horiba I, Ikegaya K, Nanki M. Recognition of coronary artery stenosis using neural network on DSA system. IEICE Trans Inf Syst. 1994;J77-D-II:1910–6.

53. Suzuki K, Horiba I, Sugie N, Ikeda S. Improvement of image quality of X-ray fluoroscopy using spatiotemporal neural filter which learns noise reduction, edge enhancement and motion compensation. Paper presented at Proc. Int. Conf. signal processing applications and technology (ICSPAT), Boston, MA, October, 1996.

54. Suzuki K, Horiba I, N. S. Edge detection from noisy images using a neural edge detector. In: Proc. IEEE Int. workshop on neural networks for signal processing (NNSP). December 2000;10:487–496.

55. Suzuki K, Liu Y, Higaki T, Funama Y, Awai K. Supervised conversion of ultra-low-dose to higher-dose CT images by using

pixel-based machine learning: phantom and initial patient studies. In: Program of scientific assembly and annual meeting of Radiological Society of North America (RSNA), SST14-06, Chicago, IL, 2013.

56. Suzuki K, Horiba I, Sugie N, Nanki M. Neural filter with selection of input features and its application to image quality improvement of medical image sequences. IEICE Trans Inf Syst. 2002;E85-D(10):1710–8.

57. Suzuki K, Horiba I, Sugie N. Efficient approximation of neural filters for removing quantum noise from images. IEEE Trans Signal Process. 2002;50(7):1787–99.

58. Suzuki K, Horiba I, Sugie N. Neural edge enhancer for supervised edge enhancement from noisy images. IEEE Trans Pattern Anal Mach Intell. 2003;25(12):1582–96.

59. Suzuki K, Horiba I, Sugie N, Nanki M. Extraction of left ventricular contours from left ventriculograms by means of a neural edge detector. IEEE Trans Med Imaging. 2004;23(3):330–9.

60. Vapnik VN. Statistical learning theory. New York: Wiley; 1998.

61. Suzuki K. Determining the receptive field of a neural filter. J Neural Eng. 2004;1(4):228–37.

62. Suzuki K, Horiba I, Sugie N. A simple neural network pruning algorithm with application to filter synthesis. Neural Process Lett. 2001;13(1):43–53.

63. Tajbakhsh N, Suzuki K. Comparing two classes of end-to-end learning machines for lung nodule detection and classification: MTANNs vs. CNNs. Pattern Recognit. 2017;63:476–86.

64. Shiraishi J, Li Q, Suzuki K, Engelmann R, Doi K. Computer-aided diagnostic scheme for the detection of lung nodules on chest radiographs: localized search method based on anatomical classification. Med Phys. 2006;33(7):2642–53.

65. Coppini G, Diciotti S, Falchini M, Villari N, Valli G. Neural networks for computer-aided diagnosis: detection of lung nodules in chest radiograms. IEEE Trans Inf Technol Biomed. 2003;7(4):344–57.

66. Hardie RC, Rogers SK, Wilson T, Rogers A. Performance analysis of a new computer aided detection system for identifying lung nodules on chest radiographs. Med Image Anal. 2008;12(3):240–58.

67. Chen S, Suzuki K, MacMahon H. A computer-aided diagnostic scheme for lung nodule detection in chest radiographs by means of two-stage nodule-enhancement with support vector classification. Med Phys. 2011;38:1844–58.

68. Armato SG 3rd, Giger ML, MacMahon H. Automated detection of lung nodules in CT scans: preliminary results. Med Phys. 2001;28(8):1552–61.

69. Ye X, Lin X, Dehmeshki J, Slabaugh G, Beddoe G. Shape-based computer-aided detection of lung nodules in thoracic CT images. IEEE Trans Biomed Eng. 2009;56(7):1810–20.

70. Way TW, Sahiner B, Chan HP, et al. Computer-aided diagnosis of pulmonary nodules on CT scans: improvement of classification performance with nodule surface features. Med Phys. 2009;36(7):3086–98.

71. Aoyama M, Li Q, Katsuragawa S, MacMahon H, Doi K. Automated computerized scheme for distinction between benign and malignant solitary pulmonary nodules on chest images. Med Phys. 2002;29(5):701–8.

72. Aoyama M, Li Q, Katsuragawa S, Li F, Sone S, Doi K. Computerized scheme for determination of the likelihood measure of malignancy for pulmonary nodules on low-dose CT images. Med. Phys. 2003;30(3):387–94.

73. Shah SK, McNitt-Gray MF, Rogers SR, et al. Computer aided characterization of the solitary pulmonary nodule using volumetric and contrast enhancement features. Acad. Radiol. 2005;12(10):1310–9.

74. Wu Y, Doi K, Giger ML, Nishikawa RM. Computerized detection of clustered microcalcifications in digital mammograms: applications of artificial neural networks. Med Phys. 1992;19(3):555–60.

75. El-Naqa I, Yang Y, Wernick MN, Galatsanos NP, Nishikawa RM. A support vector machine approach for detection of microcalcifications. IEEE Trans Med Imaging. 2002;21(12):1552–63.

76. Yu SN, Li KY, Huang YK. Detection of microcalcifications in digital mammograms using wavelet filter and Markov random field model. Comput Med Imaging Graph. 2006;30(3):163–73.

77. Ge J, Sahiner B, Hadjiiski LM, et al. Computer aided detection of clusters of microcalcifications on full field digital mammograms. Med Phys. 2006;33(8):2975–88.

78. Wu YT, Wei J, Hadjiiski LM, et al. Bilateral analysis based false positive reduction for computer-aided mass detection. Med Phys. 2007;34(8):3334–44.

79. Huo Z, Giger ML, Vyborny CJ, Wolverton DE, Schmidt RA, Doi K. Automated computerized classification of malignant and benign masses on digitized mammograms. Acad Radiol. 1998;5(3):155–68.

80. Delogu P, Evelina Fantacci M, Kasae P, Retico A. Characterization of mammographic masses using a gradient-based segmentation algorithm and a neural classifier. Comput Biol Med. 2007;37(10):1479–91.

81. Shi J, Sahiner B, Chan HP, et al. Characterization of mammographic masses based on level set segmentation with new image features and patient information. Med Phys. 2008;35(1):280–90.

82. Yoshida H, Nappi J. Three-dimensional computer-aided diagnosis scheme for detection of colonic polyps. IEEE Trans Med Imaging. 2001;20(12):1261–74.

83. Jerebko AK, Summers RM, Malley JD, Franaszek M, Johnson CD. Computer-assisted detection of colonic polyps with CT colonography using neural networks and binary classification trees. Med Phys. 2003;30(1):52–60.

84. Wang S, Yao J, Summers RM. Improved classifier for computer-aided polyp detection in CT colonography by nonlinear dimensionality reduction. Med Phys. 2008;35(4):1377–86.

85. Arimura H, Li Q, Korogi Y, et al. Computerized detection of intracranial aneurysms for three-dimensional MR angiography: feature extraction of small protrusions based on a shape-based difference image technique. Med Phys. 2006;33(2):394–401.

86. Muramatsu C, Li Q, Schmidt RA, et al. Determination of subjective similarity for pairs of masses and pairs of clustered microcalcifications on mammograms: comparison of similarity ranking scores and absolute similarity ratings. Med Phys. 2007;34(7):2890–5.

87. Muramatsu C, Li Q, Schmidt R, et al. Experimental determination of subjective similarity for pairs of clustered microcalcifications on mammograms: observer study results. Med Phys. 2006;33(9):3460–8.

88. Muramatsu C, Li Q, Suzuki K, et al. Investigation of psychophysical measure for evaluation of similar images for mammographic masses: preliminary results. Med Phys. 2005;32(7):2295–304.

89. Lo SB, Lou SA, Lin JS, Freedman MT, Chien MV, Mun SK. Artificial convolution neural network techniques and applications for lung nodule detection. IEEE Trans Med Imaging. 1995;14(4):711–8.

90. Lo SCB, Chan HP, Lin JS, Li H, Freedman MT, Mun SK. Artificial convolution neural network for medical image pattern recognition. Neural Netw. 1995;8(7–8):1201–14.

91. Lin JS, Lo SB, Hasegawa A, Freedman MT, Mun SK. Reduction of false positives in lung nodule detection using a two-level neural classification. IEEE Trans Med Imaging. 1996;15(2):206–17.

92. Lo SC, Li H, Wang Y, Kinnard L, Freedman MT. A multiple circular path convolution neural network system for detection of

mammographic masses. IEEE Trans Med Imaging. 2002;21(2):150–8.

93. Sahiner B, Chan HP, Petrick N, et al. Classification of mass and normal breast tissue: a convolution neural network classifier with spatial domain and texture images. IEEE Trans Med Imaging. 1996;15(5):598–610.

94. Zhang W, Doi K, Giger ML, Nishikawa RM, Schmidt RA. An improved shift-invariant artificial neural network for computerized detection of clustered microcalcifications in digital mammograms. Med Phys. 1996;23(4):595–601.

95. Zhang W, Doi K, Giger ML, Wu Y, Nishikawa RM, Schmidt RA. Computerized detection of clustered microcalcifications in digital mammograms using a shift-invariant artificial neural network. Med Phys. 1994;21(4):517–24.

96. Samala RK, Chan HP, Hadjiiski L, Helvie MA, Wei J, Cha K. Mass detection in digital breast tomosynthesis: deep convolutional neural network with transfer learning from mammography. Med Phys. 2016;43(12):6654.

97. Teramoto A, Fujita H, Yamamuro O, Tamaki T. Automated detection of pulmonary nodules in PET/CT images: ensemble false-positive reduction using a convolutional neural network technique. Med Phys. 2016;43(6):2821.

98. Ciompi F, de Hoop B, van Riel SJ, et al. Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2D views and a convolutional neural network out-of-the-box. Med Image Anal. 2015;26(1):195–202.

99. Kooi T, van Ginneken B, Karssemeijer N, den Heeten A. Discriminating solitary cysts from soft tissue lesions in mammography using a pretrained deep convolutional neural network. Med Phys. 2017;44(3):1017–27.

100. Lekadir K, Galimzianova A, Betriu A, et al. A convolutional neural network for automatic characterization of plaque composition in carotid ultrasound. IEEE J Biomed Health Inform. 2017;21(1):48–55.

101. Miki Y, Muramatsu C, Hayashi T, et al. Classification of teeth in cone-beam CT using deep convolutional neural network. Comput Biol Med. 2017;80:24–9.

102. He L, Chao Y, Suzuki K, Wu K. Fast connected-component labeling. Pattern Recognit. 2009;42:1977–87.

103. He L, Chao Y, Suzuki K. A run-based two-scan labeling algorithm. IEEE Trans Image Process. 2008;17(5):749–56.

104. Suzuki K, Horiba I, Sugie N. Linear-time connected-component labeling based on sequential local operations. Comput Vis Image Underst. 2003;89(1):1–23.

105. Roth HR, Lu L, Seff A, et al. A new 2.5D representation for lymph node detection using random sets of deep convolutional neural network observations. Med Image Comput Comput Assist Interv. 2014;17(Pt 1):520–7.

106. Hasegawa A, Itoh K, Ichioka Y. Generalization of shift invariant neural networks: image processing of corneal endothelium. Neural Netw. 1996;9(2):345–56.

107. Cha KH, Hadjiiski L, Samala RK, Chan HP, Caoili EM, Cohan RH. Urinary bladder segmentation in CT urography using deep-learning convolutional neural network and level sets. Med Phys. 2016;43(4):1882.

108. Moeskops P, Viergever MA, Mendrik AM, de Vries LS, Benders MJ, Isgum I. Automatic segmentation of MR brain images with a convolutional neural network. IEEE Trans Med Imaging. 2016;35(5):1252–61.

109. Austin JH, Romney BM, Goldsmith LS. Missed bronchogenic carcinoma: radiographic findings in 27 patients with a potentially resectable lesion evident in retrospect. Radiology. 1992;182(1):115–22.

110. Shah PK, Austin JH, White CS, et al. Missed non-small cell lung cancer: radiographic findings of potentially resectable lesions evident only in retrospect. Radiology. 2003;226(1):235–41.

111. Ishigaki T, Sakuma S, Horikawa Y, Ikeda M, Yamaguchi H. One-shot dual-energy subtraction imaging. Radiology. 1986;161(1):271–3.

112. Loog M, van Ginneken B. Segmentation of the posterior ribs in chest radiographs using iterated contextual pixel classification. IEEE Trans Med Imaging. 2006;25(5):602–11.

113. Hubel DH, Wiesel TN. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol. 1962;160(1):106–54.