[Invited Review Article]

Machine Learning in Medical Imaging Before and After Introduction of Deep Learning

Kenji SUZUKI, Ph.D.

Medical Imaging Research Center & Department of Electrical and Computer Engineering, Illinois Institute of Technology
3440 South Dearborn Street, Chicago, IL 60616, USA
(Received on April 4, 2017. In final form on April 10, 2017)

Abstract : Because a huge amount of new data is generated every day, the data that we treat are "big data", and they are not something that we can handle manually. Machine learning (ML) that can handle such "big data" automatically becomes a rapidly growing, indispensable area of research in the fields of medical imaging and computer vision. Recently, a terminology, deep learning emerged and became very popular in the computer vision field. It started from an event in 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 field of machine learning in medical imaging before and after the introduction of deep learning is reviewed to make clear 1) what deep learning is exactly, 2) what was changed before and after the introduction of deep learning, and 3) what is the source of the power of deep learning. This review reveals that object/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. The class of image/pixel-based ML including deep learning has a long history, but gained the popularity recently due to the new terminology, deep learning. The image/pixel-based ML is a versatile technology with substantially high performance. ML including deep learning in medical imaging is an explosively growing, promising field. It is expected that image/pixel-based ML including deep learning will be the mainstream technology in the field of medical imaging in the next few decades.
Keywords : deep learning, convolutional neural network, massive-training artificial neural network, classification, machine learning, computer-aided diagnosis, medical image analysis.

1. Introduction

Machine learning (ML) is a research area that deals with construction of data-processing models by learning from data or examples. Because a huge amount of new data is generated every day, data now are "big data", and they are not something that we can handle manually. For example, hundreds of millions of images are created and shared in YouTube or social media. A large number of CT exams is performed each year : 85 and 63 million CT scans in the U.S. and Japan, respectively, each of which contains 100-700 images (slices). ML that can handle such "big data" becomes a rapidly growing, indispensable area of research in the fields of medical imaging and computer vision.

ML plays an essential role in the medical imaging field, including medical image analysis, computer-aided diagnosis (CAD) [1, 2], and radiomics, because objects such as organs and lesions in medical images may be too complex to be represented accurately by a simple equation. Modeling of such complex objects requires a large number of parameters that have to be determined by data. 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 exhibit a flat shape [3]. Thus, detection and diagnostic tasks in medical images essentially require "learning from examples (or data)" to determine a large number of parameters in a complex model.

One of the most popular uses of ML in medical image analysis 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 object- or feature-based ML. The task of ML is to determine "optimal" boundaries for separating classes in the multi-dimensional feature space which is formed by the input features [4].

Recently, a terminology, deep learning emerged and became very popular in the computer vision field. It started from an event in 2012. A deep learning approach based on a convolutional neural network (CNN) [5] won the best-known worldwide computer-vision competition, ImageNet Classification, with the error rate smaller by 11% than the 2nd place of 26% [6]. Consequently, 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 [7].

In this paper, the field of machine learning in medical imaging before and after the introduction of deep learning is reviewed to make clear 1) what deep learning is exactly, 2) what was changed before and after the introduction of deep learning, 3) what is the source of the power of deep learning, 4) advantages and limitations of deep learning, 5) applications of deep learning, and 6) a prospect for the field of machine learning in medical imaging.

2. Machine learning in computer vision and medical imaging

2.1. Standard ML approaches before deep learning – Object- or feature-based classifiers

One of the most popular uses of ML algorithms would probably be classification. In this use, an ML algorithm is called a classifier. A standard ML approach in the field of computer vision is illustrated in Fig.1. First, objects in an image are segmented by using 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 objects by using a feature extractor. Then, extracted features are entered as input to an ML model such as linear discriminant analysis (LDA) [8], quadratic discriminant analysis (QDA) [8], a multilayer perceptron (MLP) [9], and a support vector machine (SVM) [10]. The ML model is trained with sets of input features and known class labels (i.e., C_1, C_2, ..., C_N for N classes). The training is done for determining "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 which class a new unknown object belongs to.

2.2. ML Approaches after deep learning – Image- or pixel-based ML

A terminology called deep learning emerged in 2007, and it became very popular in the computer vision field after 2012 when a deep learning approach based on a CNN [5] won an overwhelming victory in the best-known computer-vision ImageNet competition [6]. Deep learning such as deep belief nets (DBNs) [11] and deep CNNs uses pixel values in images directly instead of features calculated from segmented objects as input information ; thus, feature calculation or segmentation is not required, as shown in Fig.2. 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 discrimination power sufficient to classify objects of interest. Because deep learning can avoid errors caused by inaccurate feature calculation and segmentation which 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., object/feature-based MLs). Deep learning has multiple layers (>4) of nonlinear or quasi-nonlinear processing to acquire high-level representation of objects or features in images. Compared to object/feature-based MLs (or common classifiers), 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.3. Deep learning is also called end-to-end ML, because it enables the entire process to map from

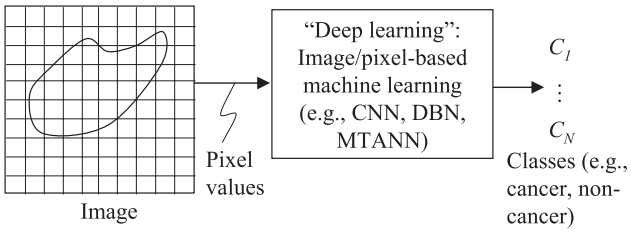


Fig.2 ML approach after the introduction of deep learning. Pixel values from an image are directly entered as input to an image/pixel-based ML model such as a convolutional neural network (CNN), a deep belief net (DBN), and a massive-training artificial neural network (MTANN). This class of ML should be called image/pixel-based ML that includes deep learning, because the major and essential difference between ML approaches before and after the introduction of deep learning is direct training of pixels in images.

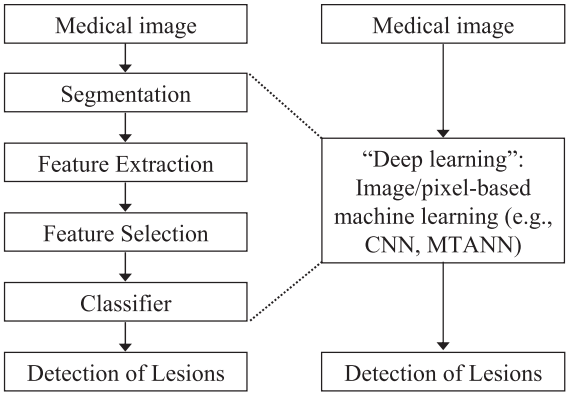


Fig.3 Changes in ML approaches before and after the introduction of deep learning. Compared to object/feature-based ML (i.e., a classifier with features), "deep learning" (or image/pixel–based ML) skips steps of segmentation of objects, feature extraction from the segmented objects, and feature selection for determining "effective features", which offers an end-to-end ML approach.

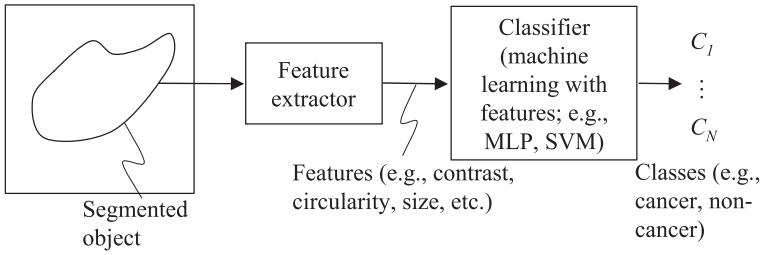


Fig.1 Standard ML approach to classification of objects (i.e., object/feature-based ML), before the introduction of deep learning, in the field of computer vision. Features (e.g., contrast, circularity, and effective diameter) are extracted from a segmented object in an image. Those features are entered as input to a classifier such as a multilayer perceptron (MLP) and a support vector machine (SVM). The output of the classifier is class categories such as cancer or non-cancer.

raw input images to the final classification, eliminating the need for hand-crafted features. It is interesting to note that people do not call the use of MLP with deep layers in the object/feature-based approach deep learning, and call a shallow CNN with only a few layers still deep learning, which is the evidence that people are confused by the terminology, deep learning.

As it is obvious from Figs.1 and 2, the major and essential difference between ML approaches before and after deep learning is the use of pixels in images directly as input to ML models, as opposed to features extracted from segmented objects. Therefore, the terminology deep learning may mislead people to the mindset that the power of deep learning comes from the deepness. A proper terminology for the "deep learning" that people call right now would be image/pixel-based ML. The deepness of MLs is still an important attribute that determines the characteristics or properties of ML models or applications. When the architecture is deep, it should be called deep image/pixel-based ML or deep object/feature-based ML.

Fig.4 summarizes the history of ML in the fields of computer vision and medical imaging. Before the popularity of "deep learning" in 2012-2013, object/feature-based ML were dominant in the fields. Before 1980 even when the terminology, 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 Rumelhard and Hinton [9]. That created the 2nd neural network (NN) research boom (the 1st one was in 1960's). In 1995, Vapnik proposed an SVM [10] and became the most popular classifier. Various ML methods were proposed including random forests by Ho et al. in 1995 [12] and dictionary learning by Mairal et al. in 2009 [13]. On the other hand, various image/pixel-based ML techniques were proposed before the introduction of the terminology, "deep learning". It started from Neocognitron by Fukushima in 1980 [14]. In 1989, LeCun et al. simplified the Neocognitron and proposed CNN [15]. In 1994, Suzuki et al. applied MLP to cardiac images in a convolutional way [16]. Two years later, Suzuki proposed neural filters to reduce noise in images [17], and in 2000, neural edge enhancers [18]. Suzuki et al. proposed MTANN for classification of patterns in 2003 [19], detection of an object in 2009 [20], and separation of specific patters from other patterns in x-ray images in 2006 [21]. Hinton et al. proposed a DBN in 2006 [11], and he created the terminology, deep learning a year later. In 2012, a CNN won in the ImageNet competition [6]. Among them, Neocognitron, MLP, CNN, neural filters, MTANN, and DBN are capable of deep architecture. Thus, the "deep learning", which is image/pixel-based ML with deep architecture to be accurate, is not new ML models, but rather it is essentially a collection of earlier ML work that was re-recognized by a different terminology recently. Deep learning researchers in the fields of computer vision and machine learning started saying so very recently.

2.3. Two "deep learning" (image/pixel-based ML) models

2.3.1. 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 [14]. LeCun et al. has developed a CNN called LeNet for handwritten ZIP-code recognition [15]. 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 interconnections or feedback loops; and the error back-propagation (BP) algorithm [9] 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

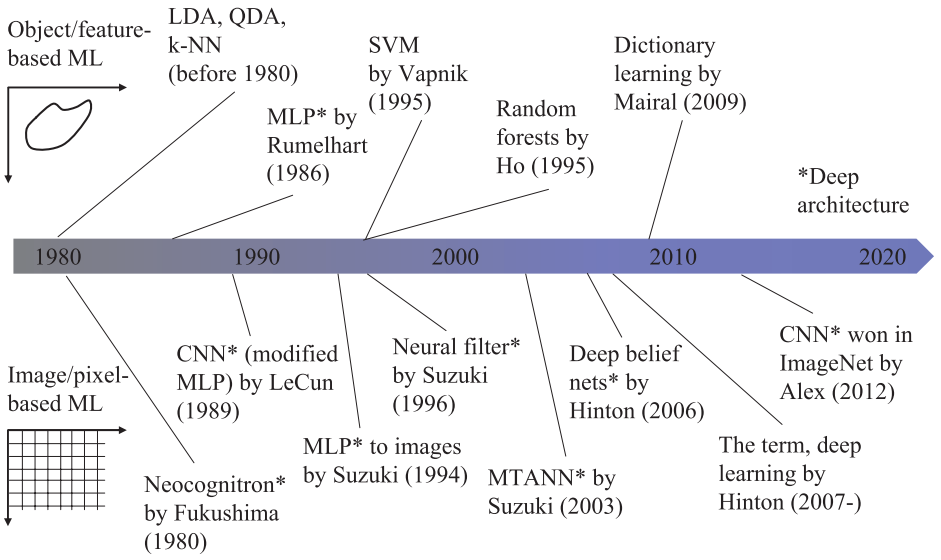


Fig.4 The history of ML in the fields of computer vision and medical imaging. There are two distinct ML approaches in the fields. Before the popularity of "deep learning" in 2012-2013, object/feature-based ML were dominant in the fields. After that, image/pixel-based ML including deep learning gained enthusiastic popularity, but it has indeed a long history.

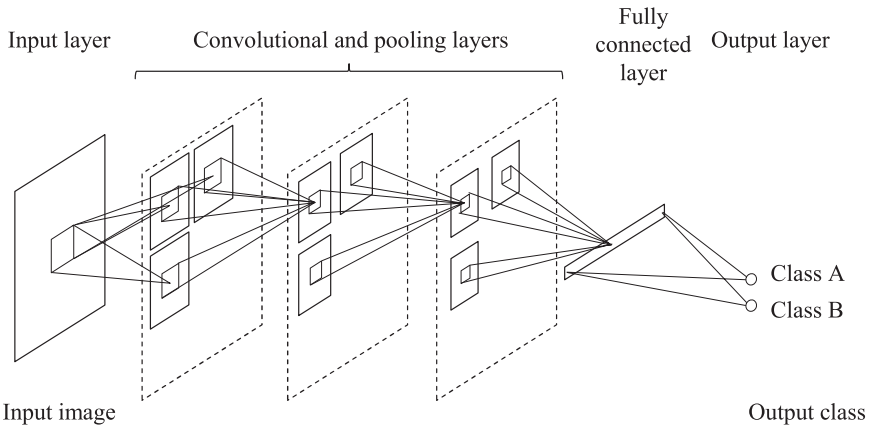


Fig.5 Architecture of a CNN. The layers in the CNN are connected with local shift-invariant interconnections (or convolution with a local kernel). The input and output of the CNN are images and class labels (e.g., Class A and Class B), respectively.

the convolution is collected into the corresponding unit in the subsequent layer. This value is further processed by the unit through an activation function and produces an output data. 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 [9] 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 (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.

2.3.2. Massive-training artificial neural network (MTANN)

In the field of signal/image processing, supervised nonlinear filters based on an MLP model (or a multilayer NN), called neural filters, were proposed [22, 23]. 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 [22, 23], edge enhancement from noisy images [24], and enhancement of subjective edges traced by a physician ("semantic segmentation") in left ventriculograms [25].

An MTANN was developed by extending neural filters to accommodate various pattern-recognition tasks [19], including classification [19, 21, 26-32], pattern enhancement and suppression [21], and object detection [20]. 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 [19]. The 2D MTANN was applied to reduction of false positives (FPs) in CAD for detection of lung nodules on 2D CT slices in a slice-by-slice way [19, 33, 34] and in chest radiographs (chest x-ray: CXR) [30], the separation of bones from soft tissue in CXR [21, 35, 36], and the distinction between benign and malignant lung nodules on 2D CT slices [26]. For processing of three-dimensional (3D) volume data, a 3D MTANN was developed by extending the structure of the 2D MTANN, and it was applied to 3D CT colonography data [27-29, 31, 32]. Various MTANN architectures were developed, including multiple MTANNs [19, 22, 23, 26, 30, 33], a mixture of expert MTANNs [27, 28], a multi-resolution MTANN [21], a Laplacian eigenfunction MTANN [32], and a massive-training support vector regression (MTSVR) [31].

The general architecture of an MTANN is illustrated in Fig.6. An MTANN consists of an ML model such as linear-output-layer artificial NN (ANN) regression, support vector regression [10, 37], and nonlinear Gaussian process regression, which is capable of operating on pixel data directly [24]. The core part of the MTANN consists of an input layer, multiple hidden layers, and an output layer, as illustrated in Fig.6(a). 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 applied to the continuous mapping of values in image processing [24]. 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 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 [38, 39]. ML regression models rather than ML classification models would be suited for the MTANN framework, because the output of the MTANN are 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

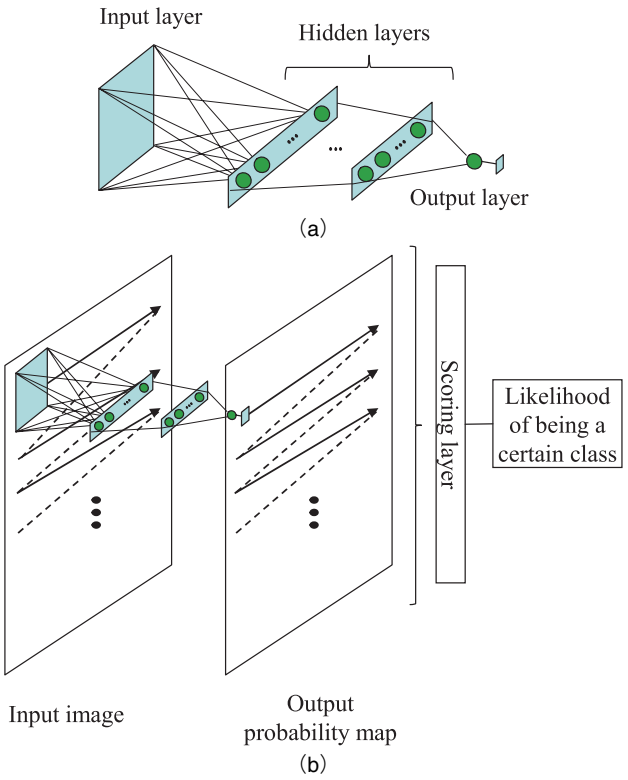


Fig.6 Architecture of an MTANN consisting of an ML model (e.g., linear-output-layer ANN regression) with subregion (or image patch, local kernel) input and single-pixel output. (a) Architecture of the core part of the MTANN. (b) Entire architecture of the MTANN. The entire output image representing a probability map is obtained by scanning with the input subregion of the MTANN in a convolutional manner on the entire input image. A scoring layer is placed in the end to convert the output probability map into a single score that represents a likelihood of being a certain class for the given input image.

on the entire input image, as illustrated in Fig.6(b). This convolution operation offers a shift-invariant property that is desirable for image classification. The input subregion and the scanning with the MTANN can be 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 probability map, unlike the class of CNNs. To use the MTANN as a classifier, a scoring layer is placed in 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. 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 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. To enrich 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 MTANN is trained by a linear-output-layer BP algorithm [24] which was derived for the linear-output-layer ANN model by use of the generalized delta rule [9]. 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.

Fig.7 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. With those nodule-enhanced images, distinction between nodules and non-nodules is made by using 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 "classification by enhancement."

2.3.3. Comparisons between the two "deep learning" models

CNNs and MTANNs are in the class of image/pixel-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 architecture, output, and the required number of training samples. In CNNs, convolutional operations are performed within the network, whereas the convolutional operation is performed outside the network in MTANNs, as shown in Figs.5 and 6. The output of CNNs is, in principle, class categories, whereas that of MTANNs is images (continuous values in a map). 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 [19]; 4 images for separation of bone components from soft-tissue components in CXR [21, 35]).

The performance of well-known CNNs (including AlexNet, LeNet, deep CNNs, and shallow CNNs) and MTANNs was extensively compared in focal lesion detection and classification problems in medical imaging [40]. 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. With a 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 a 100% sensitivity, which was significantly ($p<0.05$) lower than the best performing CNN model with 22.7 false positives per patient at the same level of

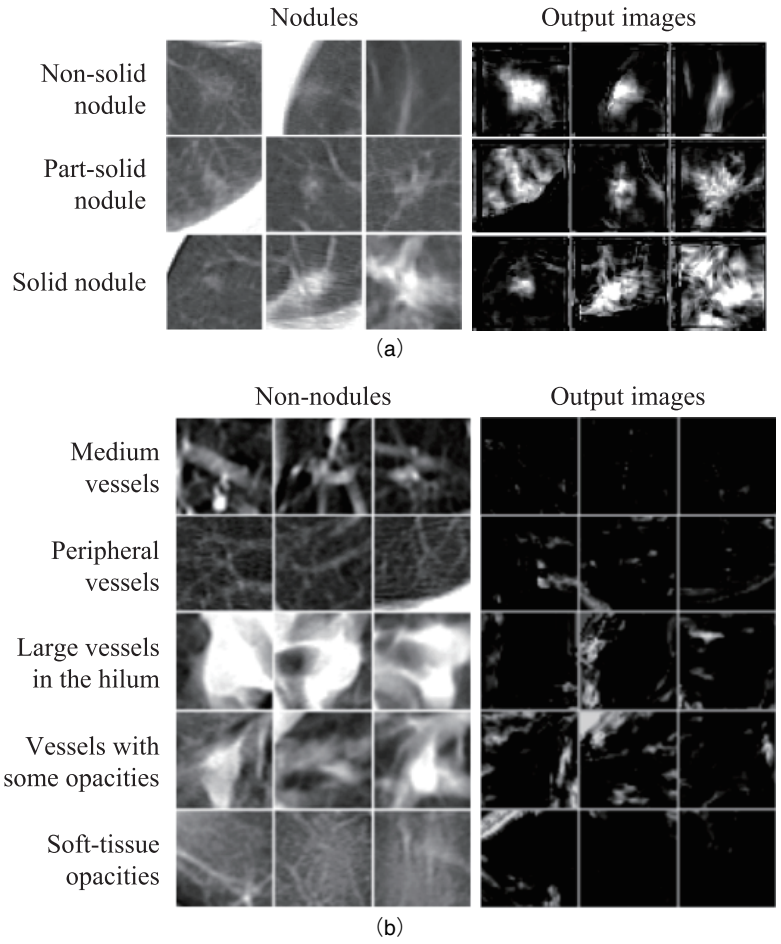


Fig.7 Illustrations of various types of nodules and non-nodules and corresponding output images of the trained MTANN in CAD for detection of lung nodules in CT. (a) Results for various types of nodules. (b) Results for various types of non-nodules. Nodules are represented by bright pixels, whereas non-nodules are almost dark around the centers of ROIs.

sensitivity. For nodule classification, MTANNs yielded an area under the receiver-operating-characteristic curve (AUC) of 0.8806, which was significantly ($p < 0.05$) greater than the best performing CNN model with an AUC of 0.7755.

3. Applications of ML in medical imaging

3.1. Applications of object/feature-based MLs

There are a large number of papers that reported applications of object/feature-based ML (common classifiers) in medical imaging, such as applications to lung nodule detection in CXR [41-44] and thoracic CT [33, 45-47], classification of lung nodules into benign or malignant in CXR [48] and thoracic CT [49, 50], detection of microcalcifications in mammography [51-54], detection of masses [55] and classification of masses into benign or malignant [56-58] in mammography, polyp detection in CT colonography [59-61], and detection of aneurysms in brain MRI [62]. In addition to the applications of ML for classification problems, there are applications of ML for regression problems such as determining the subjective similarity measure of mammographic images [63-65].

3.2. Applications of image/pixel-based MLs ("deep learning")

3.2.1 Classification between lesions and non-lesions

Before the introduction of the term, deep learning, "deep" CNNs have been used for FP reduction in CAD for lung nodule detection in CXRs [66-68]. 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), while 80% of true-positive detections were preserved. CNNs have been applied to FP reduction in CAD for detection of microcalcifications [69] and masses [70] 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, while a true-positive detection rate of 87% was preserved [69]. Shift-invariant NNs which are almost identical to CNNs have been used for FP reduction in CAD for detection of microcalcifications [71, 72]. A shift-invariant NN was trained to detect microcalcifications in regions-of-interest (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.

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 [30] and CT [19, 33, 34], and FP reduction in a CAD scheme for polyp detection in CT

colonography [27-29, 31, 32].

After the introduction of the term, deep learning, a CNN was used for classification of masses and non-masses in digital breast tomosynthesis images [73]. The CNN for digital breast tomosynthesis was trained by using transfer learning from the CNN for mammography. The CNN achieved an AUC of 0.90 in 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 [74]. 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.

3.2.2. Classification of lesion types

Before the introduction of the term, deep learning, "deep" MTANNs with seven layers were applied to distinction between benign and malignant lung nodules in low-dose screening CT [26], The MTANNs achieved an AUC value of 0.882 in classification between 76 malignant and 413 benign lung nodules, whereas an AUC value of 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 [75]. A pre-trained 2D CNN was used. The CNN achieved the performance in terms of AUC of 0.868. A pre-trained CNN was used for classification between cysts from soft tissue lesions in mammography [76]. The CNN achieved an AUC value of 0.80 in classification between benign solitary cysts and malignant masses. A CNN was used for classification of plaque compositions in carotid ultrasound [77]. 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 teeth types in cone-beam CT [78]. The CNN achieved a classification accuracy of 88.8% in classification of 7 teeth types in ROIs.

3.2.3. Detection of lesions

A "lesion-enhancement" filter based MTANNs was developed for enhancement of actual lesions in CAD for detection of lung nodules in CT [20]. For enhancement of lesions and suppression of non-lesions in CT images, the teaching image contains a probability map for 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. The nodule is enhanced in the output image of the trained MTANN filter, while normal structures such as lung vessels are suppressed.

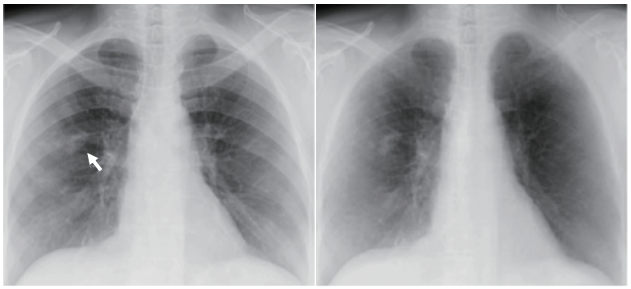
After the introduction of the term, deep learning, deep CNNs were used for detection of lymph nodes in CT [79]. Detection of lymph nodes is a challenging task, as evidenced by the fact that object/feature-based ML achieved approximately 50% sensitivity with 3 FPs/volume. By using the deep CNNs, the performance reached at 70% and 83% sensitivities with 3 FPs/volume in mediastinum and abdomen areas, respectively.

3.2.4 Segmentation of lesions or organs

Neural edge enhancers (predecessor of MTANNs) enhanced subjective edges traced by a physician ("semantic segmentation") in left ventriculograms [25]. Shift-invariant NNs were used for detection of the boundaries of the human corneal endothelium in photomicrographs [80]. A CNN was used for segmentation of the bladder in CT urography [81]. The CNN achieved a Jaccard index of 76.2% +/− 11.8% for bladder segmentation, compared with "gold-standard" manual segmentation. A CNN was used for segmentation of tissues in MR brain images [82]. The CNN achieved average Dice coefficients of 0.82-0.91 in five different datasets.

3.2.5. Separation of bones from soft tissue in CXR

Studies showed that 82 to 95% of the lung cancers missed by radiologists in CXR were partly obscured by overlying bones such as ribs and/or a clavicle [83, 84]. To prevent from such misses, MTANNs were developed for separation of bones from soft tissue in CXR [21, 35]. To this end, the MTANN were trained with input CXRs with overlapping bones and the corresponding "teaching" dual-energy bone images acquired with a dual-energy radiography system [85]. Fig.8 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 image/pixel-based ML was developed for suppression of ribs in CXR [86].



(a) Input chest radiograph with a nodule (arrow)　(b) "Virtual" dual-energy soft-tissue image by MTANN

Fig.8 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) Result of an application of the trained MTANN.

4. Advantages and limitations of "deep learning"

As described earlier, the major difference between image/pixel-based ML (e.g., "deep learning") and object/feature-based ML (i.e., classifiers) is the direct use of pixel values with the image/pixel-based ML. In other words, unlike ordinary classifiers, feature calculation from segmented objects is not necessary. Because the image/pixel-based ML can avoid errors caused by inaccurate feature calculation and segmentation, the performance of the image/pixel-based ML can be higher than that of ordinary feature-based classifiers. Image/pixel-based MLs learn pixel data directly, and thus all information on pixels should not be lost before the pixel data are entered into the image/pixel-based ML, 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 image/pixel-based ML, development and implementation of segmentation and feature calculation, and selection of features are unnecessary.

Limitations of "deep learning" (image/pixel-based MLs) are 1) very high computational cost for training because of the high dimensionality of input data and 2) a required large number of training images. Because image/pixel-based MLs use pixel data in images directly, the number of input dimensions is generally large. To address the issue of high dimension of input data, Laplacian-eigenfunction-based dimensionality reduction of the input vectors to a 3D MTANN was proposed. With the dimensionality reduction, the training time was reduced by a factor of 8.5. A CNN requires a huge number of training images (e.g., 1,000,000) to determine a large number of parameters in the CNN, whereas an MTANN requires a small number of training images (e.g., 10) because of its simpler architecture. With GPU implementation, an MTANN completes training for a couple of hours, whereas a deep CNN does in several days.

5. Conclusion

In this paper, ML in medical imaging before and after the introduction of deep learning is reviewed to make clear 1) what deep learning is exactly, 2) what was changed before and after the introduction of deep learning, and 3) what is the source of the power of deep learning. This review reveals that object/feature-based ML (i.e., classifiers with features) were dominant before the introduction of deep learning, and that the major and essential difference between ML approaches before and after deep learning is training of image data directly without object segmentation or feature extraction; thus, it is the source of the power of deep learning. There is a long history of deep learning techniques including Neocognitron, CNN, neural filters, and MTANN in the class of image/pixel-based ML, except a new terminology, deep learning. Image/pixel-based ML including deep learning is a very powerful, versatile technology with higher performance, which can make 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 image/pixel-based ML including deep learning will be the mainstream technology in medical imaging in the next few decades.

6. Acknowledgments

The authors are grateful to all members in Computational Intelligence in Biomedical Imaging Lab at Illinois Institute of Technology and Suzuki Lab at University of Chicago for their valuable contributions to research in ML including deep learning in medical imaging in the past 15 years or so.

REFERENCES

[1] Giger ML, Suzuki K: Computer-Aided Diagnosis (CAD), In: Biomedical Information Technology, Edited by: Feng DD, Academic Press, 359-374, 2007.

[2] Doi K: Current status and future potential of computer-aided diagnosis in medical imaging, Br J Radiol, 78 Spec No 1, S3-S19, 2005.

[3] Lostumbo A, Wanamaker C, Tsai J, et al.: Comparison of 2D and 3D views for evaluation of flat lesions in CT colonography, Acad. Radiol., 17(1), 39-47, 2010.

[4] Duda RO, Hart PE, Stork DG: In: Pattern Recognit., 2nd, Hoboken, NJ, Wiley Interscience, 2001.

[5] Lawrence S, Giles CL, Tsoi AC, et al.: Face recognition: A convolutional neural-network approach, IEEE Transactions on Neural Networks, 8(1), 98-113, 1997.

[6] Krizhevsky A, Sutskever I, Hinton GE: Imagenet classification with deep convolutional neural networks, Advances in neural information processing systems, 1097-1105, 2012.

[7] LeCun Y, Bengio Y, Hinton G: Deep learning, Nature, 521(7553), 436-444, 2015.

[8] Fukunaga K: Introduction to Statistical Pattern Recognition, San Diego: Academic Press, 1990.

[9] Rumelhart DE, Hinton GE, Williams RJ: Learning representations by back-propagating errors, Nature, 323, 533-536, 1986.

[10] Vapnik VN: The Nature of Statistical Learning Theory, Berlin: Springer-Verlag, 1995.

[11] Hinton G, Osindero S, Teh Y-W: A fast learning algorithm for deep belief nets, Neural computation, 18(7), 1527-1554, 2006.

[12] Ho TK: Random decision forests, Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on, IEEE, 278-282, 1995.

[13] Mairal J, Bach F, Ponce J, et al.: Online dictionary learning for sparse coding, Proceedings of the 26th Annual International Conference on Machine Learning, Montreal, Quebec, Canada, ACM, 689-696, 2009.

[14] Fukushima K: Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position, Biol Cybern, 36(4), 193-202, 1980.

[15] LeCun Y, Boser B, Denker JS, et al.: Backpropagation applied to handwritten zip code recognition, Neural Comput., 1(4), 541-551, 1989.

[16] Suzuki K, Horiba I, Ikegaya K, et al.: Recognition of Coronary Artery Stenosis Using Neural Network on DSA System, IEICE Transactions on Information Systems, J77-D-II, 1910-1916, 1994.

[17] Suzuki K, Horiba I, Sugie N, et al.: Improvement of image quality of x-ray fluoroscopy using spatiotemporal neural filter which learns noise reduction, edge enhancement and motion compensation, Proc. Int. Conf. Signal Processing Applications and Technology (ICSPAT), Boston, MA, 1382-1386, 1996.

[18] Suzuki K, Horiba I, N. S: Edge detection from noisy images using a neural edge detector, Proc. IEEE Int. Workshop on Neural Networks for Signal Processing (NNSP), 10, 487-496, 2000.

[19] Suzuki K, Armato SG, 3rd, Li F, et al.: Massive training artificial neural network (MTANN) for reduction of false positives in computerized detection of lung nodules in low-dose computed tomography, Med. Phys., 30(7), 1602-

1617, 2003.
[20] 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., 54(18), S31-45, 2009.
[21] Suzuki K, Abe H, MacMahon H, et al.: Image-processing technique for suppressing ribs in chest radiographs by means of massive training artificial neural network (MTANN), IEEE Trans. Med. Imaging, 25(4), 406-416, 2006.
[22] Suzuki K, Horiba I, Sugie N, et al.: Neural filter with selection of input features and its application to image quality improvement of medical image sequences, IEICE Trans. Inf. Syst., E85-D(10), 1710-1718, 2002.
[23] Suzuki K, Horiba I, Sugie N: Efficient approximation of neural filters for removing quantum noise from images, IEEE Trans. Signal Process., 50(7), 1787-1799, 2002.
[24] Suzuki K, Horiba I, Sugie N: Neural edge enhancer for supervised edge enhancement from noisy images, IEEE Trans. Pattern Anal. Mach. Intell., 25(12), 1582-1596, 2003.
[25] Suzuki K, Horiba I, Sugie N, et al.: Extraction of left ventricular contours from left ventriculograms by means of a neural edge detector, IEEE Trans. Med. Imaging, 23 (3), 330-339, 2004.
[26] Suzuki K, Li F, Sone S, et al.: Computer-aided diagnostic scheme for distinction between benign and malignant nodules in thoracic low-dose CT by use of massive training artificial neural network, IEEE Trans. Med. Imaging, 24(9), 1138-1150, 2005.
[27] 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., 37(1), 12-21, 2010.
[28] Suzuki K, Yoshida H, Nappi J, et al.: 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., 35(2), 694-703, 2008.
[29] Suzuki K, Yoshida H, Nappi J, et al.: Massive-training artificial neural network (MTANN) for reduction of false positives in computer-aided detection of polyps: Suppression of rectal tubes, Med. Phys., 33(10), 3814-3824, 2006.
[30] Suzuki K, Shiraishi J, Abe H, et al.: False-positive reduction in computer-aided diagnostic scheme for detecting nodules in chest radiographs by means of massive training artificial neural network, Acad. Radiol., 12(2), 191-201, 2005.
[31] 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., 38, 1888-1902, 2011.
[32] 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, 29(11), 1907-1917, 2010.
[33] 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., 11(6), 617-629, 2004.
[34] 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, 237(2), 684-690, 2005.
[35] Suzuki K, Abe H, Li F, et al.: Suppression of the contrast of ribs in chest radiographs by means of massive training artificial neural network, Proc. SPIE Medical Imaging (SPIE MI), San Diego, CA, 1109-1119, 2004.
[36] 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., 193(5), W397-402, 2009.
[37] Vapnik VN: Statistical Learning Theory, New York: Wiley, 1998.
[38] Suzuki K: Determining the receptive field of a neural filter, Journal of Neural Engineering, 1(4), 228-237, 2004.
[39] Suzuki K, Horiba I, Sugie N: A simple neural network pruning algorithm with application to filter synthesis, Neural Process. Lett., 13(1), 43-53, 2001.
[40] Tajbakhsh N, Suzuki K: Comparing Two Classes of End-to-End Learning Machines for Lung Nodule Detection and Classification: MTANNs vs. CNNs., Pattern Recognit., 63, 476-486, 2017.
[41] Shiraishi J, Li Q, Suzuki K, et al.: Computer-aided diagnostic scheme for the detection of lung nodules on chest radiographs: localized search method based on anatomical classification, Med. Phys., 33(7), 2642-2653, 2006.
[42] Coppini G, Diciotti S, Falchini M, et al.: Neural networks for computer-aided diagnosis: detection of lung nodules in chest radiograms, IEEE Trans Inf Technol Biomed, 7(4), 344-357, 2003.
[43] Hardie RC, Rogers SK, Wilson T, et al.: Performance analysis of a new computer aided detection system for identifying lung nodules on chest radiographs, Med. Image Anal., 12(3), 240-258, 2008.
[44] 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., 38, 1844-1858, 2011.
[45] Armato SG, 3rd, Giger ML, MacMahon H: Automated detection of lung nodules in CT scans: preliminary results, Med. Phys., 28(8), 1552-1561, 2001.
[46] Ye X, Lin X, Dehmeshki J, et al.: Shape-based computer-aided detection of lung nodules in thoracic CT images, IEEE Trans. Biomed. Eng., 56(7), 1810-1820, 2009.
[47] 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., 36(7), 3086-3098, 2009.
[48] Aoyama M, Li Q, Katsuragawa S, et al.: Automated computerized scheme for distinction between benign and malignant solitary pulmonary nodules on chest images, Med. Phys., 29(5), 701-708, 2002.
[49] Aoyama M, Li Q, Katsuragawa S, et al.: Computerized scheme for determination of the likelihood measure of malignancy for pulmonary nodules on low-dose CT

images, Med. Phys., 30(3), 387-394, 2003.
[50] 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., 12(10), 1310-1319, 2005.
[51] Wu Y, Doi K, Giger ML, et al.: Computerized detection of clustered microcalcifications in digital mammograms: applications of artificial neural networks, Med Phys, 19(3), 555-560, 1992.
[52] El-Naqa I, Yang Y, Wernick MN, et al.: A support vector machine approach for detection of microcalcifications, IEEE Trans Med Imaging, 21(12), 1552-1563, 2002.
[53] Yu SN, Li KY, Huang YK: Detection of microcalcifications in digital mammograms using wavelet filter and Markov random field model, Comput Med Imaging Graph, 30(3), 163-173, 2006.
[54] Ge J, Sahiner B, Hadjiiski LM, et al.: Computer aided detection of clusters of microcalcifications on full field digital mammograms, Med Phys, 33(8), 2975-2988, 2006.
[55] Wu YT, Wei J, Hadjiiski LM, et al.: Bilateral analysis based false positive reduction for computer-aided mass detection, Med Phys, 34(8), 3334-3344, 2007.
[56] Huo Z, Giger ML, Vyborny CJ, et al.: Automated computerized classification of malignant and benign masses on digitized mammograms, Acad Radiol, 5(3), 155-168, 1998.
[57] Delogu P, Evelina Fantacci M, Kasae P, et al.: Characterization of mammographic masses using a gradient-based segmentation algorithm and a neural classifier, Comput Biol Med, 37(10), 1479-1491, 2007.
[58] 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, 35(1), 280-290, 2008.
[59] Yoshida H, Nappi J: Three-dimensional computer-aided diagnosis scheme for detection of colonic polyps, IEEE Trans. Med. Imaging, 20(12), 1261-1274, 2001.
[60] Jerebko AK, Summers RM, Malley JD, et al.: Computer-assisted detection of colonic polyps with CT colonography using neural networks and binary classification trees, Medical Physics, 30(1), 52-60, 2003.
[61] Wang S, Yao J, Summers RM: Improved classifier for computer-aided polyp detection in CT colonography by nonlinear dimensionality reduction, Med Phys, 35(4), 1377-1386, 2008.
[62] 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, 33(2), 394-401, 2006.
[63] 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., 34(7), 2890-2895, 2007.
[64] 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., 33(9), 3460-3468, 2006.
[65] Muramatsu C, Li Q, Suzuki K, et al.: Investigation of psychophysical measure for evaluation of similar images for mammographic masses: preliminary results, Med. Phys., 32(7), 2295-2304, 2005.
[66] Lo SB, Lou SA, Lin JS, et al.: Artificial convolution neural network techniques and applications for lung nodule detection, IEEE Trans Med Imaging, 14(4), 711-718, 1995.
[67] Lo SCB, Chan HP, Lin JS, et al.: Artificial convolution neural network for medical image pattern recognition, Neural Networks, 8(7-8), 1201-1214, 1995.
[68] Lin JS, Lo SB, Hasegawa A, et al.: Reduction of false positives in lung nodule detection using a two-level neural classification, IEEE Trans Med Imaging, 15(2), 206-217, 1996.
[69] Lo SC, Li H, Wang Y, et al.: A multiple circular path convolution neural network system for detection of mammographic masses, IEEE Trans Med Imaging, 21(2), 150-158, 2002.
[70] 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, 15(5), 598-610, 1996.
[71] Zhang W, Doi K, Giger ML, et al.: An improved shift-invariant artificial neural network for computerized detection of clustered microcalcifications in digital mammograms, Med Phys, 23(4), 595-601, 1996.
[72] Zhang W, Doi K, Giger ML, et al.: Computerized detection of clustered microcalcifications in digital mammograms using a shift-invariant artificial neural network, Med Phys, 21(4), 517-524, 1994.
[73] Samala RK, Chan HP, Hadjiiski L, et al.: Mass detection in digital breast tomosynthesis: Deep convolutional neural network with transfer learning from mammography, Med Phys, 43(12), 6654, 2016.
[74] Teramoto A, Fujita H, Yamamuro O, et al.: Automated detection of pulmonary nodules in PET/CT images: Ensemble false-positive reduction using a convolutional neural network technique, Med Phys, 43(6), 2821, 2016.
[75] 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, 26(1), 195-202, 2015.
[76] Kooi T, van Ginneken B, Karssemeijer N, et al.: Discriminating solitary cysts from soft tissue lesions in mammography using a pretrained deep convolutional neural network, Med Phys, 44(3), 1017-1027, 2017.
[77] 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, 21(1), 48-55, 2017.
[78] Miki Y, Muramatsu C, Hayashi T, et al.: Classification of teeth in cone-beam CT using deep convolutional neural network, Comput Biol Med, 80, 24-29, 2017.
[79] 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, 17(Pt 1), 520-527, 2014.
[80] Hasegawa A, Itoh K, Ichioka Y: Generalization of shift invariant neural networks: image processing of corneal

endothelium, Neural Networks, 9(2), 345-356, 1996.
[81] Cha KH, Hadjiiski L, Samala RK, et al.: Urinary bladder segmentation in CT urography using deep-learning convolutional neural network and level sets, Med Phys, 43(4), 1882, 2016.
[82] Moeskops P, Viergever MA, Mendrik AM, et al.: Automatic Segmentation of MR Brain Images With a Convolutional Neural Network, IEEE Trans Med Imaging, 35(5), 1252-1261, 2016.
[83] Austin JH, Romney BM, Goldsmith LS: Missed bronchogenic carcinoma: radiographic findings in 27 patients with a potentially resectable lesion evident in retrospect, Radiology, 182(1), 115-122, 1992.
[84] 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, 226(1), 235-241, 2003.
[85] Ishigaki T, Sakuma S, Horikawa Y, et al.: One-shot dual-energy subtraction imaging, Radiology, 161(1), 271-273, 1986.
[86] Loog M, van Ginneken B: Segmentation of the posterior ribs in chest radiographs using iterated contextual pixel classification, IEEE Trans. Med. Imaging, 25(5), 602-611, 2006.