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., \( C_1, C_2, \ldots, C_N \) 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.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.](image1)

![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).](image2)

![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.](image3)
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 patterns 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

---

Medical Imaging and Information Sciences
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 \( i \) 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 (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

\[
O(x, y, z) = ML(I(x-i, y-j, z-k)(i, j, k) \in R),
\]

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.
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 patient 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
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 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].

![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.](image)

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 dimensionality 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


[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


[80] Hasegawa A, Itoh K, Ichikawa Y: Generalization of shift invariant neural networks: image processing of corneal...
endothelium, Neural Networks, 9(2), 345-356, 1996.


