Journal of Science, Computing and Engineering Research (JSCER)
Volume-3, Issue-1, Jan-Feb 2022.
DOI: *https://doi.org/10.46379/jscer.2022.030104*

A Comprehensive Review on Machine Learning Applications of Convolutional Neural Networks to Medical Image Analysis

[1]KOMMU NAVEEN [2]Dr. R.M.S PARVATHI
[1]Dept. Of ECE, Anna University, Chennai, Tamilnadu, India.
[2]Dept. Of CSE, Sri Ramakrishna Institute of Technology, Coimbatore, Tamilnadu, India.

Article Information

Received : January 08 2022
Revised : January 18 2022
Accepted : January 252022
Published : February 05 2022

Corresponding Author:
Kommu Naveen
Email: naveenkarunya@gmail.com

Abstract— Recently, a machine learning (ML) area called deep learning emerged in the computer-vision field and became very popular in many fields. When a deep-learning approach based on a convolutional neural network (CNN) won an overwhelming victory in the best-known worldwide computer vision competition, Image Net Classification. Since then, researchers in many fields, including medical image analysis, have started actively participating in the explosively growing field of deep learning. In this paper, deep learning techniques and their applications to medical image analysis are surveyed. This survey overviewed 1) standard ML techniques in the computer-vision field, 2) what has changed in ML before and after the introduction of deep learning, 3) ML models in deep learning, and 4) applications of deep learning to medical image analysis. The survey of deep learning also revealed that there is a long history of deep-learning techniques 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, and it is expected that deep learning will be the mainstream technology in medical image analysis in the next few decades.

Keywords: ***Computer-aided diagnosis, Convolutional neural network, Deep Learning, Medical Image, Medical image analysis.***

Citation: Kommu Naveen and Dr. R.M.S Parvathi "A Comprehensive Review on Machine Learning Applications of Convolutional Neural Networks to Medical Image Analysis", Journal of Science, Computing and Engineering Research, 3(1), 210-217, 2022.

I. INTRODUCTION

Machine learning (ML) is indispensable in the field of medical imaging [1-13], including medical image analysis, computer-aided diagnosis (CAD) [14-17], and radiomics, because objects of interest in medical images, such as lesions and organs, may be too complex to be represented accurately by a simple equation or model. For example, a polyp in the colon is model as a bulbous object, but there are also colorectal lesions that have a flat shape [18, 19]. Modelling of such complex objects needs a complex model with a large number of parameters. Determining such a large number of parameters cannot be accomplished manually. Thus, tasks in medical imaging essentially require "learning from data (or examples)" for determination of a large number of parameters in a complex model. Therefore, ML plays an essential role in the medical imaging field.

Fig. 1 Standard ML for classifying lesions (i.e., ML with feature input or feature-based ML) in the field of computer vision before the introduction of deep learning. Features (e.g., circularity, contrast, 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.

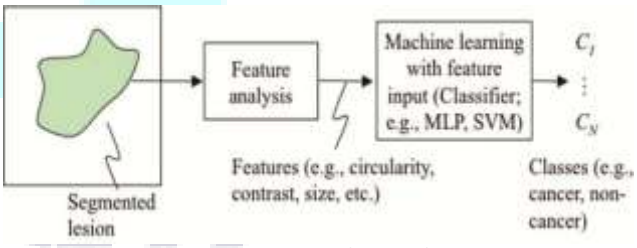

Figure: 1 Standard ML for classifying lesions

One of the most popular uses of ML in computer-aided diagnosis (CAD) 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].

Recently, an ML area called deep learning emerged in the computer vision field. 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 learning became very popular in the computer vision field after late 2012, when a deep-learning approach based on a convolutional neural network (CNN) [22] 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% [23]. 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 [24].

This paper surveys the research area of deep learning and its applications to medical image analysis. The surveys includes

1) Standard ML techniques in the computer-vision field,
2) What has changed in ML before and after the introduction of deep learning,
3) ML models in deep learning,
4) Applications of deep learning to medical image analysis.

II. MACHINE LEARNING BEFORE AND AFTER DEEP LEARNING

A. "Standard" ML—ML with feature input

ML algorithms are often used for classification of objects in images, and they are usually called classifiers. A "standard" ML approach in the computer vision field is illustrated in Fig. Fig. 2 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. Fig. 3 "New" ML class, ML with image input (image-based ML) in the field of computer vision after the introduction of deep learning.

Pixel values from an image are directly entered as input to an ML model with image input 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 with image input 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.

1. First, objects such as lesions in an image are segmented by using a segmentation technique such as grey-level-based segmentation, edge-based segmentation, and an active contour (or shape) model. Next, features such as circularity, contrast, and size of the segmented lesion are calculated in a feature analysis step. Then, the calculated features are inputted to an ML model such as linear or quadratic discriminant analysis (LDA or QDA) [25], a multilayer perceptron (MLP) [26], a support vector machine (SVM), [27], and random forests [28]. The ML model is trained with sets of input features of lesions and N known class labels from C1 to CN for the lesions. The training is performed to determine "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 simply a classifier.

B. "Deep Learning"—Ml With Image Input

A term, deep learning, was created by Hinton in 2007 for ML models for a high-level representation of objects, but it was not recognized widely until late 2012. Deep learning became very popular in the computer vision field after late 2012, when a deep-learning approach based on a convolutional neural network (CNN) [22] won an overwhelming victory in the best-known computer-vision competition, ImageNet [23]. The architecture of a general CNN is illustrated in Fig. 2. 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). Deep learning models, such as a deep CNN and a deep belief net (DBN) [29] which is a generative graphical model with multiple layers, use 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. 2. Deep learning has multiple layers (>4) of nonlinear or quasi-nonlinear processing to acquire a high-level representation of objects or features in images.

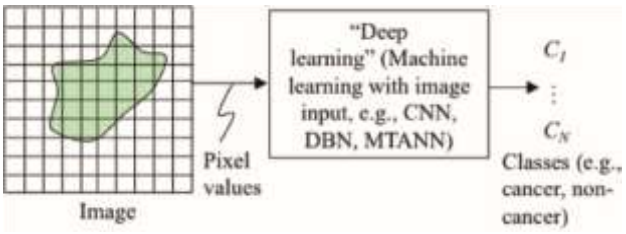

Figure: 2. Deep learning multiple layers

Compared to ML with feature input (also referred to as feature-based ML, 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. 4. Deep learning is also called an end-to-end ML paradigm or approach, because it enables the entire process to map from raw input images to the final classification, eliminating the need for hand-crafted features. 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).

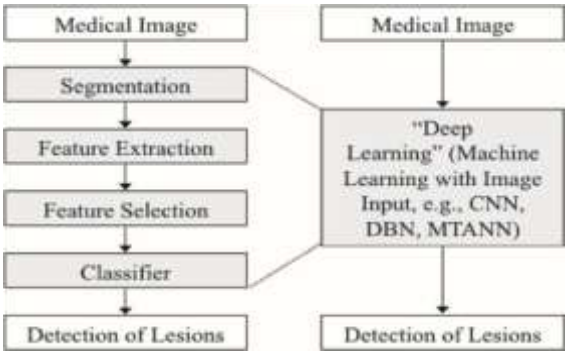

Figure: 3 Changes in ML approaches before and after the introduction of ML with image input (image-based ML) including "deep learning".

Compared to ML with feature input, 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.

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 deep learning, which is the evidence that people are confused by the terminology, deep learning. As shown in Figs. 1 and 3, the major and essential difference between ML with feature input (feature-based ML) and "deep learning" 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 the "deep learning" that people use right now would be ML with image input or image-based ML. 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 featureinput (feature-based ML).

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 deep learning, which is the evidence that people are confused by the terminology, deep learning. As shown in Figs. 1 and 3, the major and essential difference between ML with feature input (feature-based ML) and "deep learning" 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 lables on separate page at the end of manuscript with their lables.

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 the "deep learning" that people use right now would be ML with image input or image-based ML. 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 featureinput (feature-based ML).

A class of ML with image input or image-based ML was proposed and developed in the field of medical image analysis before the introduction of the term "deep learning". Suzuki et al. invented and developed massive-training artificial neural networks (MTANNs) for classification between lesions and non-lesions in medical images in 2003 before the introduction of "deep learning". MTANNs use images as input, as opposed to features extracted from a segmented lesion, and they are capable of deep layers. MTANNs are an end-to-end ML paradigm that does the entire process from input images to the final classification.

C. History Of Ml In Computer Vision And Medical Image Analysis

Figure 5 summarizes the history of ML in the fields of computer vision and medical image analysis. Before the popularity of "deep learning" starting in 2013, ML with feature input (feature-based ML) was dominant in the fields. Before 1980, even when the term "machine learning" did not exist, classical classifiers such as LDA, QDA, and a k-nearest neighbour classifier (k-NN) were used for classification. In 1986, MLP was proposed by Rumelhart and Hinton [26]. The introduction of the MLP created the 2nd neural network (NN) research boom (by the way, the 1st one was in 1960's). In 1995, Vapnik proposed an SVM

[27] and became the most popular classifier for a while, partially because of publicly available code on the Internet in the Internet age. Various ML methods were pro- posed, including random forests by Ho et al. in 1995 [28], and dictionary learning by Mairal et al. in 2009. We suggest that you use border for graphic (ideally 300 dpi), with all fonts embedded) and try to reduce the size of figure to be adjust in one column. Figure and Table Labels: Use 8 point Times New Roman for Figure and Table labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader.

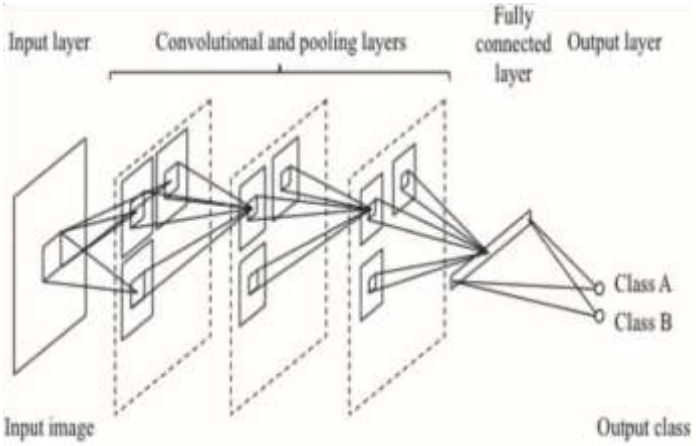

Figure: 4. Deep learning is also called an end-to-end ML paradigm or approach

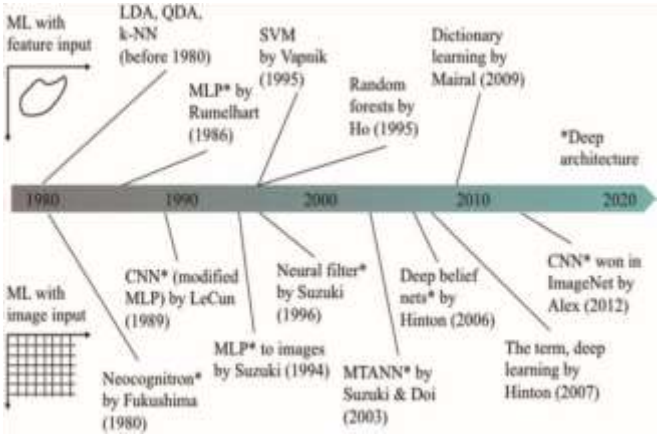

Figure:5 The history of ML in the fields of computer vision and medical imaging.

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 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 Neocognition by Fukushima in 1980. In 1989, Le Cun et al. simplified the Neocognition and proposed a CNN, 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. Two years later, Suzuki et al. proposed neural filters based on a modified MLP to reduce noise in images, and in 2000, they proposed neural edge enhancers. Suzuki et al. proposed MTANN for classification of patterns in 2003, detection of objects in 2009, separation of specific patters from other patterns in x- ray images in 2006, and reduction of noise and artifacts on CT images in 2013. Hinton et al. proposed a deep brief network (DBN) in 2006 [29], 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 [23]. Among them, Neocognitron, MLP, CNN, neural filters, neural edge enhancers, MTANNs, and DBN are capable of deep architecture. Thus, "deep learning", which is ML with image input (image-based ML) with deep architecture, to be accurate, has a long history. "Deep learning" 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.

We suggest that you use border for graphic (ideally 300 dpi), with all fonts embedded) and try to reduce the size of figure to be adjust in one column. Figure and Table Labels: Use 8 point Times New Roman for Figure and Table labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader.

III. MACHINE LEARNING MODELS IN "DEEP LEARNING"

A. 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. LeCun et al. developed a CNN called LeNet for handwritten ZIP-code recognition. The LeNet has 5 layers (1 input, 3 hidden, and 1 output layer). The input layer has a small 16×16 pixel image. The 3 hidden layers consist of 2 convolutional layers and one fully connected layer.

The architecture of a general CNN is illustrated in Fig. 2. 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 BP algorithm [26] 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 [26] 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.

Massive-Training Artificial Neural Network (MTANN) Predecessors of MTANNs that belong to the class of ML with image input or "deep learning" were proposed in the field of signal and image processing. Suzuki et al. proposed supervised nonlinear filters based on an MLP model (or a multilayer NN), called neural filters. The neural filter employs a linear-output-layer NN regression model as a convolution kernel of a filter. The inputs to the neural filter are pixels in a sub region (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, edge enhancement from noisy images, and enhancement of subjective edges traced by a physician ("semantic segmentation") in left ventriculograms .

An MTANN was developed by extending of neural filters to accommodate various pattern-recognition tasks, including classification, pattern enhancement and suppression, and object detection. 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. 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, and in chest radiographs (chest x-ray; CXR) , the separation of bones from soft tissue in CXR, and the distinction between benign and malignant lung nodules on 2D CT slices [45]. 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. Various MTANN architectures were developed, including multiple MTANNs, a mixture of expert MTANNs, a multi-resolution MTANN, a Laplacian Eigen function MTANN, as well as a massive-training support vector regression (MTSVR) and a massive-training nonlinear Gaussian process regression [50].

The general architecture of an MTANN for classification is illustrated in Fig. 6. An MTANN consists of an ML model such as linear-output-layer artificial NN (ANN) regression, support vector regression, and nonlinear Gaussian process regression, which is capable of operating on pixel data directly. The core part of the MTANN consists of an input layer, multiple hidden layers, and an output layer. The linear-output-layer ANN regression model employs a linear function instead of a sigmoid function as the activation function of the unit in the output layer because the characteristics of an ANN were improved significantly with a linear function when it was applied to the continuous mapping of values in image processing. 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 sub region (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 sub region, represented by

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

where x, y, and z are the coordinate indices, ML(•) is the output of the ML model, I(x, y, z) is a pixel value of the

Fig. 6 Architecture of an MTANN consisting of an ML model (e.g., linear-output-layer ANN regression) with sub region (or image patch, local kernel) input and single-pixel output. The entire output image representing a likelihood map is obtained by scanning with the input sub region of the MTANN in a convolutional manner on the entire input image. 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.

input image, and (i, j, k) is a coordinate in a sub region, R. 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. The entire output image is obtained by scanning with the input sub region of the MTANN in a convolutional manner on the entire input image, as illustrated in Fig. 6. This convolutional operation offers a shift-invariant property that is desirable for image classification. The input sub region 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. 6. A score for a given image is defined as a product of the image and a weighting function. The weighting function combines 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.

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. For enrichment of training samples, a training region, extracted from the input images is divided pixel by pixel into a large number of overlapping sub regions. 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 sub regions together with each of the corresponding teaching single pixels; hence the term "massive-training ANN". The MTANN is trained by a linear-output-layer back propagation (BP) algorithm which was derived for the linear-output-layer ANN model by use of the generalized delta rule [26]. After training, the MTANN is expected to output the highest value when an object of interest is located at the center of the sub region of the MTANN, a lower value as the distance from the sub region center increases, and zero when the input sub region contains other patterns.

D. Similarities And Differences Between The Two "Deep Learning" Models

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 in an end-to-end ML paradigm 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. 2 and 6. (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 ; and 4 images for separation of bone components from soft-tissue components in CXR).

Performance: The performance of well-known CNNs (including the AlexNet, the LeNet, a relatively deep CNN, a shallow CNN, and a fine-tuned AlexNet (FineTunedAlexNet) which used transfer learning from a computer-vision-trained AlexNet) and MTANNs was compared extensively in focal lesion (i.e., lung nodule) detection and classification problems in medical imaging. 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 100% sensitivity, which was significantly ($p<0.05$) lower than that for the best-performing CNN model (Fine Tuned Alex Net), 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.

IV. APPLICATIONS OF MACHINELEARNING" TO ANNO TATION IMAGING

A. Classification between lesions and non-lesions

Before the introduction of the term, deep learning, deep CNNs had been used for false positive (FP) reduction in CAD for lung nodule detection in CXRs. A convolution NN was trained with 28 CXRs for classifying between 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 maintained. CNNs have been applied to FP reduction in CAD for detection of microcalcifications and masses in mammograms. A CNN was trained with 34 mammograms for classifying between microcalcifications and FP detections (i.e., non-microcalcifications). The trained CNN reduced 90% of the FPs, which resulted in 0.5 FPs per image, whereas a true-positive detection rate of 87% was maintained. Shift-invariant NNs which are almost identical to CNNs, have been used for FP reduction in CAD for detection of microcalcifications. 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.

After the introduction of the term, deep learning, a CNN was used for classification of masses and nonmasses in digital breast tomosynthesis images. 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 the classification between 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. 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 overall 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. There are a growing number of papers for applications of CNNs in this area, and all the papers are not reviewed in this paper.

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 CT and CXR, and FP reduction in a CAD scheme for polyp detection in CT colonography. For enhancement of lesions and suppression of non-lesions, the teaching image contains 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. After training, a scoring method is used for combining of output voxels from the trained MTANNs. With the trained MTANN, nodules such as a solid nodule, a part-solid (mixed-ground- glass) nodule, and a non-solid (ground-glass) nodule were enhanced, whereas non-nodules such as different-sized lung vessels and soft-tissue opacity were suppressed. In this way, 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." In the application of the MTANN to FP reduction in CAD for lung CT, 83% of the false positives that had not been removed by ML with feature input were removed with a reduction of one true positive nodule (i.e., a drop in a sensitivity of 1.7%). With the MTANN, the FP rate of a CAD scheme was improved from 27.4 to 4.8 FPs per patient at an overall sensitivity of 80.3% [31]. In the MTANN application to FP reduction in CAD for CXR, the MTANN eliminated 68.3% of FPs that had not been removed by ML with feature input with a reduction of one true-positive result. The FP rate of the original CAD scheme was improved from 4.5 to 1.4 FPs per image at an overall sensitivity of 81.3%. In the MTANN application to CT colonography, the MTANN removed 63% of the FPs that had not been removed by ML with feature input without the loss of any true positive; thus, the FP rate of the original CAD scheme was improved to 1.1 FPs per patient while the original sensitivity of 96.4% was maintained. Classification of lesion types

After the introduction of the term "deep learning", a CNN was used for classification between perifissural nodules and non-perifissural nodules in CT. 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. The CNN achieved an AUC value of 0.80 in the classification between benign solitary cysts and malignant masses. CNN was used for classification of plaque compositions in carotid ultrasound. 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. The CNN achieved a classification accuracy of 88.8% in classification of 7 tooth types in ROIs. A CNN (AlexNet) with transfer learning was applied to classification among diffuse lung diseases in CT . With transfer learning, the weights of the AlexNet that had been trained with natural images were used as autoencoder (or feature extractor), and the classifier part in the final stage was fine-tuned to fit medical image data. A lung tissue classification system was developed with a convolutional classification restricted Boltzmann machine that learns feature descriptions directly from training data. 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 [45]. The MTANNs achieved an AUC value of 0.882 in the classification between 76 malignant and 413 benign lung nodules, whereas an AUC value for chest radiologists for the same task with a subset of the database was 0.56.

B. Segmentation Of Lesions Or Organs

Before the introduction of the term, deep learning, shift-invariant NNs were used for detection of the boundaries of

the human corneal endothelium in photomicrographs. After the introduction of the term, deep learning, a CNN was used for segmentation of the bladder in CT urography. 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. The CNN achieved average Dice coefficients of 0.820.91 in five different datasets. An automatic organ segmentation method for pancreas in CT images was developed with multi-level CNNs. The method achieved a Dice similarity coefficient of 71.87% in segmentation of pancreas. A method for segmentation of organs in CT images was developed with CNNs, which accomplishes an end to-end, voxel-wise multiple-class prediction to map each voxel in a CT image directly to an anatomical label.

Neural edge enhancers (NEEs; predecessor of MTANNs) enhanced subjective edges traced by a physician ("semantic segmentation") in left ventriculograms [44]. The NEE that had been trained with the physician's subjective edges was able to enhance the left ventricle contour very well. 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.

C. Detection of lesions or clinically significant patterns

After the introduction of the term, deep learning, deep CNNs were used for detection of lymph nodes in CT. 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. Automatic detection of cerebral micro-bleeds in MR images was developed by means of 3D convolutional neural networks. A computer-aided detection system for pulmonary nodules was developed by means of multi-view convolutional networks, for which discriminative features are automatically learnt from the training data.

An MTANN-based "lesion-enhancement" filter was developed for enhancement of actual lesions in CAD for detection of lung nodules in CT [38]. 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. Nodules were enhanced in the output image of the trained MTANN filter, whereas normal structures such as lung vessels were suppressed. After large and small remaining regions were removed by use of area information obtained with connected-component labeling, accurate nodule detection was achieved with no FP, which means that one MTANN functions as a complete CAD scheme with high accuracy.

D. 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. To prevent such misses, MTANNs were developed for separation of bones from soft tissues in CXR.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. With a trained MTANN, the contrast of ribs was suppressed substantially, whereas the contrast of soft tissue such as lung vessels was maintained. A filter learning in the class of ML with image input (image- based ML) was developed for suppression of ribs in CXR.

E. 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. The receptive field of the trained NEE that was revealed by application of a method for designing the optimal structure of an NN to the trained NEE. The receptive field shows which input pixels were used for enhancement of edges from noisy images. Furthermore, they analyzed the units in the hidden layer of the trained NEE. They show the analysis results of the internal representation of the trained NEE, which indicate the operations for diagonal edge enhancement together with smoothing in a hidden unit, vertical edge enhancement together with horizontal smoothing in another hidden unit, and edge enhancement with smoothing for another diagonal orientation in other hidden unit. 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. They showed, for the first time, that the ML model was able to acquire the receptive fields of various simple cells, which had been discovered by Hubel and Wiesel in the cat and monkey cerebral cortex .

V. ADVANTAGES AND LIMITATIONS OF "MACHINE 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 use pixel data directly would generally differ from those of ordinary feature-based classifiers (ML with feature input). 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.

VI. CONCLUSION

In this paper, deep learning techniques and their applications to medical image analysis are surveyed. Frist, standard ML techniques in the computer-vision field, namely, ML with feature input (or feature-based ML), are reviewed to make clear what has changed in ML before and after the introduction of deep learning. The comparisons between MLs before and after deep learning revealed that ML with feature input was dominant before 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 depth of the model is an important attribute. The survey of deep learning also revealed that 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". "Deep learning" even before the term existed, namely, the class of ML with image input was applied to various problems in medical image analysis including classification between lesions and non-lesions, classification between lesion types, segmentation of lesions or organs, and detection of lesions. 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, and it is expected that deep learningwill be the mainstream technology in medical image analysis in the next few decades.

REFERENCES

[1]. Suzuki K: Machine Learning for Medical Imaging. A special issue of Algorithms, 2010.

[2]. Wang F, Yan P, Suzuki K, et al.: Machine Learning in Medical Imaging (MLMI). Lecture Notes in Computer Science Vol. 6357 , Springer-Verlag, Berlin, 2010.

[3]. Suzuki K, Wang F, Shen D, et al.: Machine Learning in Medical Imaging (MLMI). Lecture Notes in Computer Science Vol. 7009 , Springer-Verlag, Berlin, 2011.

[4]. Suzuki K: Machine Learning for Medical Imaging 2012. A special issue of Algorithms, 2012.

[5]. Suzuki K, Yan P, Wang F, et al.: Machine learning in medical imaging. Int J Biomed Imaging 2012: 123727, 2015

[6]. Suzuki K: Machine Learning in Computer-Aided Diagnosis: Medical Imaging Intelligence and Analysis. IGI Global, Hershey, PA, 2016.

[7]. Wang F, Shen D, Yan P, et al.: Machine Learning in Medical Imaging (MLMI). Lecture Notes in Computer Science Vol. 7588 , Springer-Verlag, Berlin, 2012.

[8]. Suzuki K: Machine learning in computer-aided diagnosis of the thorax and colon in CT: A survey. IEICE Trans InfSyst E96-D: 772-783 , 2013.

[9]. Wu G, Zhang D, Shen D, et al.: Machine Learning in Medical Imaging (MLMI). Lecture Notes in Computer Science Vol. 8184 , Springer-Verlag, Berlin, 262, 2013.

[10]. Yan P, Suzuki K, Wang F, et al.: Machine learning in medical imaging. Mach Vision Appl 24: 1327-1329 , 2013.

[11]. Shen D, Wu G, Zhang D, et al.: Machine learning in medical imaging. Comput Med Imaging Graph 41: 1-2 , 2015.

[12]. Suzuki K, Zhou L, Wang Q: Machine learning in medical imaging Pattern Recognit 63: 465-467 , 2017.

[13]. El-BazA, Gimel'farb G, Suzuki K: Machine learning applications in medical image analysis. Comput Math Meth Med 2017: 2361061, 2017.

[14]. Doi K: Overview on research and development of computer-aided diagnostic schemes. Semin Ultrasound CT MRI 25: 404410 , 2014 .

[15]. Doi K: Current status and future potential of computer-aided diagnosis in medical imaging. Br J Radiol 78 (Suppl. 1): S3-S19, 2015.