



## Review Article

# Deep Learning in Medical Imaging

### Corresponding Author

Namkug Kim

<https://orcid.org/0000-0002-3438-2217>

Department of Convergence Medicine,  
Asan Medical Institute of Convergence  
Science and Technology, University of  
Ulsan College of Medicine, Asan Medical  
Center, 88 Olympic-ro 43-gil, Songpa-gu,  
Seoul 05505, Korea  
Tel: +82-2-3010-6573  
E-mail: namkugkim@gmail.com

Received: November 21, 2019

Revised: December 10, 2019

Accepted: December 12, 2019



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © 2019 by the Korean Spinal  
Neurosurgery Society

## INTRODUCTION

Artificial intelligence (AI) technology, powered by advanced computing power, a large amount of data, and new algorithms, becomes more and more popular. It has been applied to various kinds of fields such as healthcare, manufacturing and convenient living life, so on. AI, in general, has 3 categories. One is a symbolic approach that outputs answers using a rule-based search engine. Another is the Bayesian theorem-based approach. The other is the connectionism approach based on deep neural networks (DNNs). While each approach has its strengths and weaknesses, the connectionism approach is recently gaining a lot of attention to solve complex problems.

Machine learning (ML) is a subset of AI that learns data itself with minimum human intervention to classify categories or predict future or uncertain conditions.<sup>1</sup> Since ML is data-driven

Mingyu Kim<sup>1</sup>, Jihye Yun<sup>1</sup>, Yongwon Cho<sup>1</sup>, Keewon Shin<sup>1</sup>, Ryoungwoo Jang<sup>1</sup>,  
Hyun-jin Bae<sup>1</sup>, Namkug Kim<sup>1,2</sup>

<sup>1</sup>Department of Convergence Medicine, Asan Medical Institute of Convergence Science and Technology, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Korea

<sup>2</sup>Department of Radiology, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Korea

The artificial neural network (ANN), one of the machine learning (ML) algorithms, inspired by the human brain system, was developed by connecting layers with artificial neurons. However, due to the low computing power and insufficient learnable data, ANN has suffered from overfitting and vanishing gradient problems for training deep networks. The advancement of computing power with graphics processing units and the availability of large data acquisition, deep neural network outperforms human or other ML capabilities in computer vision and speech recognition tasks. These potentials are recently applied to health-care problems, including computer-aided detection/diagnosis, disease prediction, image segmentation, image generation, etc. In this review article, we will explain the history, development, and applications in medical imaging

**Keywords:** Artificial intelligence, Deep learning, Machine learning, Precision medicine, Radiology

learning, it is categorized into nonsymbolic AI and can predict from unseen data. Here, ML tasks include regression, classification, detection, segmentation, etc. Generally, data sets of ML consist of exclusive training, validation, and test sets. It learns characteristics of data from the training data set and validates the learned characteristics from the validation data set. Finally, one can confirm the accuracy of ML by using the test data set.

As a part of ML, an artificial neural network (ANN) is a brain-inspired algorithm that consists of layers with connected nodes. It consists of input and output layers with hidden layers. Here, the first layer has input values and the last layer has corresponding labeled values. During training, the value of each node is determined by parameterizing weights through learning algorithms such as back propagation. Weights for each node are optimized towards the direction to reduce losses and thus increase accuracy. By iterating the backpropagation, optimized weights

can be obtained. However, ANN has limitations that sometimes the training ends up in a local minimum or optimized only for trained data which results in overfitting problems. Recently researchers progress a deep learning to expand ANN into DNN by stacking multi-hidden layers with connected nodes between input and output layers. The multilayer can deal with more complex problems by composing simple decisions between layers. DNN generally shows better performance than the shallow layered network in prediction tasks such as classification and regression.<sup>2</sup> Each layer of DNN optimized its weights based on the unsupervised restricted Boltzmann machine<sup>3</sup> to prevent learning converges at local minimum or overcome overfitting problems. Recently, residual neural networks is also known to avoid vanishing gradient problem using skip connections.<sup>2</sup> Besides, the advent of big data and graphics processing units could solve complex problems and shorten the computation time.

Accordingly, the deep learning algorithm gets a lot of attention these days to solve various problems in medical imaging fields. One example is to detect disease or abnormalities from X-ray images and classify them into several disease types or severities in radiology.<sup>4,5</sup> This kind of task has been executed based on the various ML algorithms with proper optimization, theoretical or empirical approaches. One example is computer-aided detection (CAD) systems which were developed and applied to the clinical system since the 1980s. However, the CAD system generates more false positives than physicians and thus led to the increment of assessment time and unnecessary biopsies.<sup>6,7</sup> Thanks to the deep learning technology, these problems could have been overcome with great answering accuracy and allow humans to spend time on other productive tasks. However, the advent of this technology does not mean the ultimate replacement of physicians, especially radiologists. Instead, it helps radiologists to diagnose patients more accurately.

### 1. Supervised and Unsupervised Learning

Primary ML methods are categorized into supervised learning, unsupervised learning and reinforcement learning (RL). RL is not adequate to medical application, because the decision of a RL system affects both the patient's future health and future treatment options. As a result, long-term effects are harder to estimate.<sup>8</sup> The main difference between supervised and unsupervised learnings is whether the training data set has labeled outputs corresponding to input data. The supervised learning infers a mathematical relationship between the inputs and the labeled outputs while the unsupervised learning infers a function that expresses hidden characteristics reside in input data.

In supervised learning, output data can have categorical value or numerical continuous value depending on its task. It becomes a classification or a pattern recognition problem when the output data is in categorical value while it becomes a regression problem when the output data is in continuous numerical value. Here, the classification task can be binary, multiclass or multilabeled where the multilabeled means more than one class exists in each input data. On the other hand, unsupervised learning includes cluster analysis, principal component analysis, and generative adversarial networks (GANs). In addition, semisupervised learning is also widely used when one has a small amount of labeled data set. Since acquiring the labeled data is generally very difficult or very expensive, semisupervised learning could be cost-effective.

In supervised learning, K-nearest neighbors (KNN) is the simplest ML algorithm for classification or regression tasks.<sup>9</sup> It finds K numbers of nearest data points from input data and votes to decide its class in a classification task. In a regression task, a value of input point is decided by averaging out K numbers of nearest data points. However, prediction through KNN gets slower when the number of training data becomes higher.<sup>10</sup> On the other hand, linear regression eliminates this problem.<sup>11</sup> Linear regression parameterizes a linear model with given training data. Once the parameter of a linear model is optimized, the prediction of a given data is just an output from the best-fit formula. Support vector regression and ANN are widely used nowadays since they show better performances in various regression problems.<sup>12</sup> Similarly, logistic regression, random forest and support vector machine are widely used for classification.<sup>13</sup> Logistic regression parameterizes the logistic model to predict binary classification. Recently, ensemble learning, by combining various classification algorithms for more accurate prediction, is commonly used.<sup>14</sup>

### 2. Convolutional Neural Network

As a part of deep learning, a convolutional neural network (CNN) is recently spotlighted in computer vision for both supervised and unsupervised learning tasks.<sup>15</sup> The CNN has broken the all-time records from traditional vision tasks.<sup>16</sup> The compositions of CNN are convolutional, pooling and fully connected layers. The primary role of the convolutional layer is to identify patterns, lines, and edges so on. Each hidden layer of CNN consists of convolutional layers that convolve input array with weight-parameterized convolution kernels. The multiple kernels generate multiple feature images and made succeed in various vision tasks such as segmentation and classification. Between

the convolutional layers, feature maps are locally progressively and spatially pooled pooling layers. The pooling layer transfers the maximum or average value and thus reduces the size of feature maps. This process catches features of an image with robust to the position and shape. Empirically, max pooling process is generally used. The CNN architecture is composed of alternating these convolutional and pooling layers repeatedly. For the classification or regression tasks, the fully connected layers are attached at the end of the CNN architecture and provide a final decision. During training, a loss is estimated by differencing labeled value and predicted value. On the other hand, in the segmentation task, convolutional layers and up-sampling layers are attached at the end of pooling layers to reconstruct the size of the input image. Thus, training loss is evaluated by differencing labeled mask image and reconstructed output image through CNN. Since CNN architecture is composed of many layers, a number of parameters for training can reach millions. This means a lot of data is needed for training needs to acquire competent accuracy. The number of data depends on task purpose and image characteristics. For instance, at least 1,000 images per class are necessary to get a competent result in a classification task if one trains the data from scratch. However, data collection is usually very difficult and even more hard if one also needs labeled data. To overcome this problem, data augmentation is also tried in general task which generates images from a limited number of data using image transformation methods such as

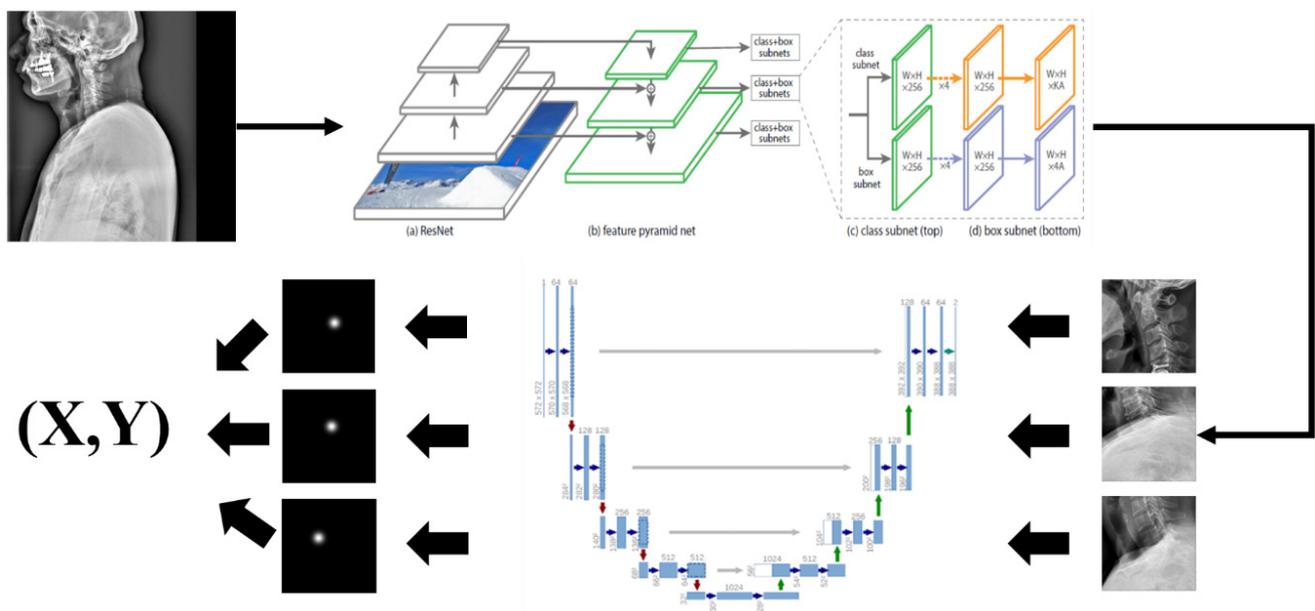
rotation, translation, scaling, flipping, so on.<sup>17</sup>

## RADIOLOGICAL APPLICATIONS

In this chapter, various kinds of radiologic applications in classification, object detection, image segmentation, image generation, and image transformation were discussed.

### 1. Image Classification

One key task for radiologists is an appropriate differential diagnosis for each patient's medical images, and this classification task includes a wide range of applications from determining the presence or absence of a disease to identifying the type of malignancy. Recently introduced DNN, especially CNN, has improved imaging-based classification performance in various medical applications, including the diagnosis of tuberculosis, diabetic retinopathy, and skin cancers.<sup>18-20</sup> Since medical images contain various sizes and types of complex disease patterns, it would be difficult for the CNN models to directly train complicated disease patterns. These complex problems could be solved by curriculum learning strategy that involves gradual training of more complex concepts.<sup>21</sup> This curriculum learning with weak labeling of high-scale chest X-ray scans performed well for classification of 5 disease patterns and required less preparation to train the model.<sup>22</sup> Deep learning requires a large amount of data to minimize overfitting and improve the performances, where-



**Fig. 1.** Network for detecting landmarks to plan spine surgery of patients in spine sagittal X-ray. From the detection of region of interest in images, landmarks were segmented with U-Net, and its corresponding coordinates (x and y) were evaluated.

as it is difficult to achieve these big datasets with medical images of low-incidence serious diseases in general practice. Thus, a better classification strategy is needed for these small datasets. Combination of radiomic features and multilayer perceptron network classifier served a high-performing and generalizable model for a small dataset with heterogeneous magnetic resonance imaging (MRI) protocols.<sup>23</sup> Furthermore, CNNs could

be incorporated into current radiomics model by extracting a large number of deep features from hidden layers of it.<sup>24,25</sup> These deep features which are evaluated not by feature engineering (handcrafted) but by feature learning could contain more abstract information of medical images and provide more predictive patterns compared with the handcrafted features.

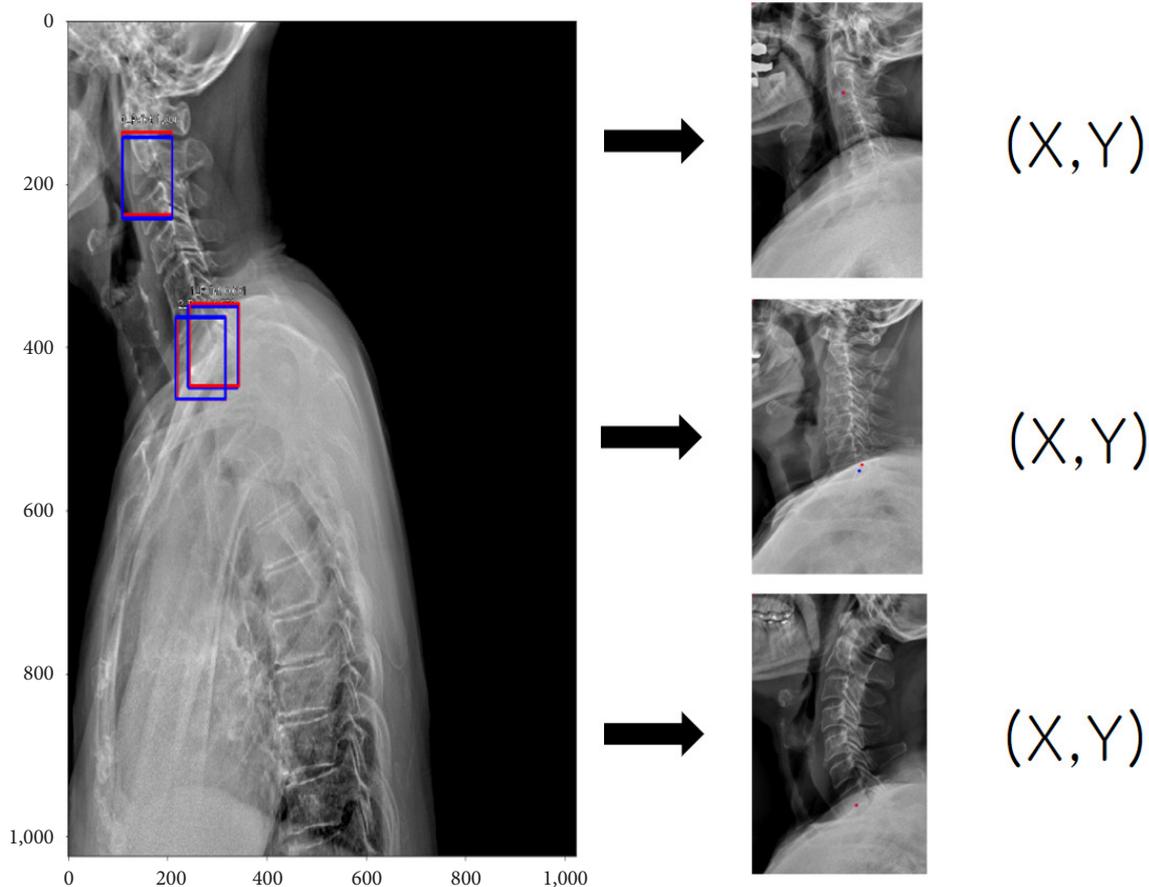
## 2. Object Detection

Object detection is finding and categorizing objects. In biomedical images, a detection technique is also performed to identify the areas where the patient's lesions are located as box coordinates. Deep learning-based object detection can be composed of 2 types. One is the region proposal-based algorithms.<sup>26-28</sup> This approach extracts various types of patches using a selective search algorithm from input images. Afterward, the trained model decides whether multiple objects exist in each patch and classifies objects based on region of interest (ROI). Specifically, the region proposal network was developed to increase the speed of

**Table 1.** The error of landmark prediction based on cascaded convolutional neural net

Landmark point	Mean $\pm$ SD (mm)	Range (mm)	Miss
C2 lower midpoint	0.56 $\pm$ 0.42	0–1.82	0
C7 lower dorsal point	1.82 $\pm$ 1.4	0–6.16	1
C7 lower midpoint	1.4 $\pm$ 1.68	0–5.32	1
S1 upper dorsal point	1.26 $\pm$ 1.68	0.14–4.62	3
S1 upper midpoint	1.96 $\pm$ 1.82	0.14–13.3	3

SD, standard deviation.



**Fig. 2.** A typical example for detecting of region of interest (ROI) images (left; red, gold standard; blue, prediction ROI) and their coordinates (x and y) of corresponding landmarks in spine sagittal X-ray.

the detection process.<sup>28</sup> The other techniques performed the object detection using the regression method as one stage network.<sup>29-32</sup> These approaches are directly finding and detecting bounding box coordinates and class probabilities from image pixels in whole images.<sup>29-31</sup> Although the region proposal approach as 2-stage network shows better performance in terms of accuracy, the regression-based model as one stage network is better in terms of speed as well as accuracy. Recently, RetinaNet<sup>32</sup> has been introduced to complement the disadvantage of 1 stage network. This network has applied focal loss<sup>32</sup> to solve the problem caused by extreme foreground-background data or class imbalance. Various object detection algorithms proposed for biomedical images are based on strong labels of per-pixel or the coordinated of bounding box coordinates. To acquire strong labels for detecting disease patterns or conditions is expensive and inevitable in medical environments. To overcome the cost of annotation data, we should exploit a transfer learning with the pretrained weight of the model learned from general natural images or a large number of medical images like national institutes of health dataset and fine-tune the model with a small number of medical images.

For example, we introduce deep learning techniques through a spine sagittal X-ray. The spine sagittal X-ray plays an important role in clinical diagnosis and operation plans in spine patients. We developed deep learning algorithms to acquire sagittal parameter data of the whole spine X-ray. The training procedure to detect the parameter of the sagittal X-ray consists of 2 steps in Fig. 1. First, the regional patterns of spine X-ray images were identified by RetinaNet with spine X-ray images. Second, the coordinates of x and y as landmarks to diagnose spine patients were inferred by U-Net with the detected ROI images. Table 1 shows the error of landmark prediction. Fig. 2 shows the results of the detection of ROI images in the spine sagittal X-ray using RetinaNet. Fig. 3 shows the results of point detection in the spine sagittal X-ray with RetinaNet and U-Net. Each is the best, mean, and worst results. Fig. 4 shows the results of point detection in spine sagittal X-ray with RetinaNet and U-Net. Each result is the best, mean, and worst results. All results are totally able to detect and find point information to diagnose patients in spine sagittal X-ray.

### 3. Image Segmentation and Registration

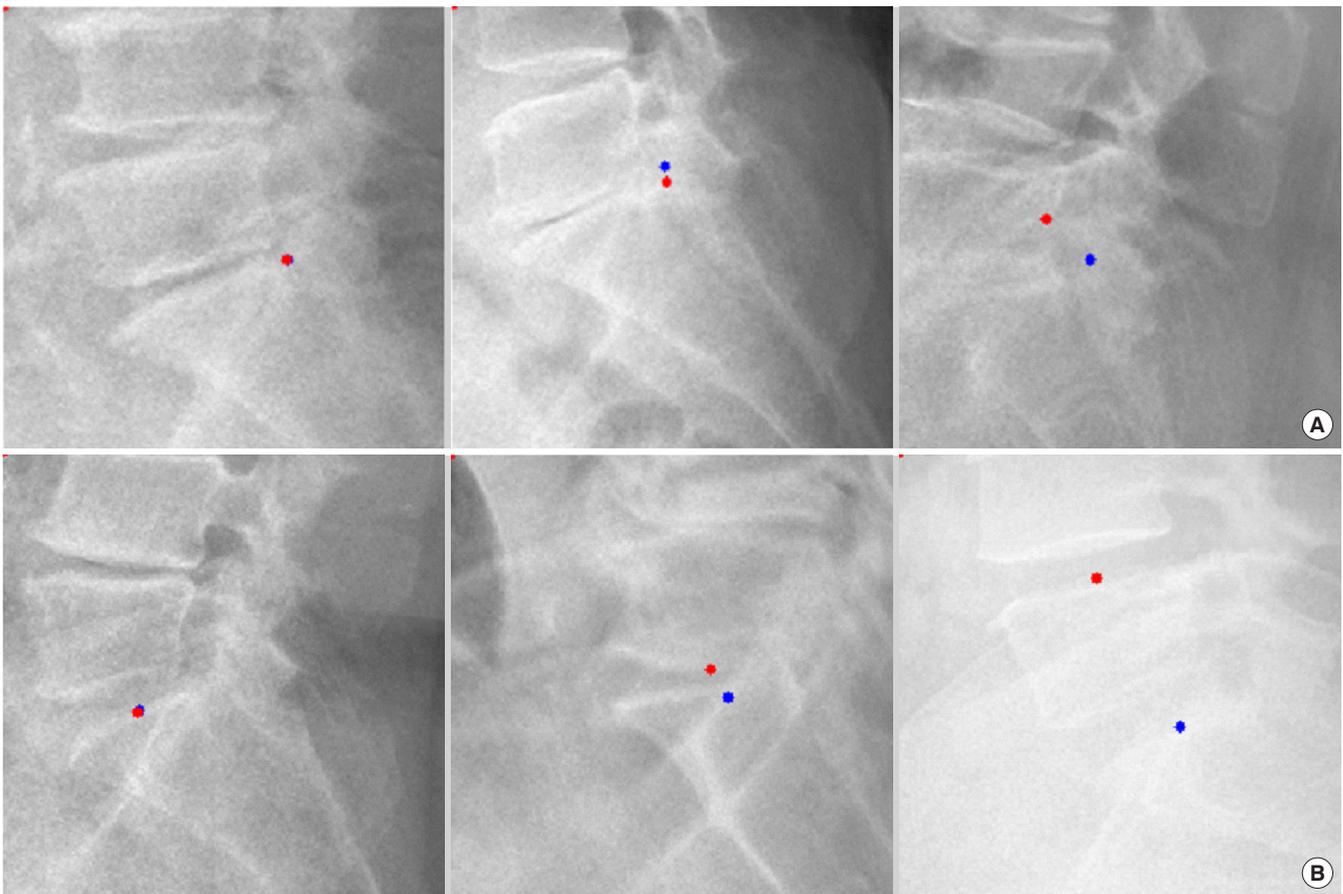
As medical images provide a lot of information, various automatic segmentation and registration algorithms have been studied and proposed for use in clinical settings. In recent years, deep learning technology has been used for analysing medical imag-



**Fig. 3.** Examples of landmark detections of c-spine region in spine sagittal X-ray (best case in left column, mean case in middle column, and the worst case in right column; red, gold standard; blue, prediction). (A) Landmark of C2 lower midpoint, (B) landmark of C7 lower dorsal point, and (C) landmark of C7 lower midpoint.

es in various fields, and it shows excellent performance in various applications such as segmentation and registration.

The classical method of image segmentation is based on edge detection filters and several mathematical algorithms. Using several techniques to improve targeted segmentation performance such as dependent thresholding and close-contour methods.<sup>33</sup> Alternatively, registration was attempted for segmentation.<sup>34</sup> To improve segmentation performance associated with medical images, DNNs, especially CNNs, have been gradually introduced. Attempts have been made for the segmentation of the tumors and other structures in the brain, lungs, biological cells, and membranes.<sup>35-37</sup> These approaches used patch-based



**Fig. 4.** Examples of landmark detections of l-spine region in spine sagittal X-ray (best case in left column, mean case in middle column, and the worst case in right column; red, gold standard; blue, prediction). (A) Landmark of S1 upper dorsal point and (B) landmark of S1 upper midpoint.

2-dimensional CNN techniques and postprocessing in the same way as classical ML. However, training a patch-based method can take a long time, and depending on the number of patches, learning might not be possible.

Several CNN architectures have been proposed that feed through entire images with better image resolution.<sup>37-39</sup> Long et al.<sup>40</sup> introduced the fully CNN (fCNN) for the segmentation of images, however, fCNNs produce segmentations of lower resolution as compared to input images. That was due to the successive use of convolutional and pooling layers, both of which reduce the dimensionality. To predict segmentation of the same resolution as the input images, Brosh et al.<sup>38,39</sup> proposed the use of a 3-layer convolutional encoder network for multiple sclerosis lesion segmentation. The combination of convolutional and deconvolutional layers allows the network to produce segments that are of the same resolution as the input images. Ronneberger et al.<sup>41</sup> was suggested novel architecture named U-NET using convolutional and deconvolutional layers with skip

connections which enables to obtain highly accurate segmentation probability maps in fully convolutional layers.

As 2-dimensional segmentation performance increases well enough, research has been conducted to segment the multiple slices of MRI and computed tomography (CT). The 2.5-dimensional (2.5D) approaches were inspired that 2.5D has richer spatial information of neighboring pixels with less computational costs than 3-dimensional (3D).<sup>42,43</sup> Yet, there were still limitations using 2D kernels, not applying 3D filters which can extract abundant volumetric information. The 3D extended U-Net model is developed to segment kidney.<sup>44</sup> The suggested model demonstrated 0.863 averaged intersection over union of a kidney from 3D volume. However, 3D U-Net has the disadvantage of not being able to put the whole image due to memory limitations and reducing the image input. To optimize this, there has been a lot of researches focusing on performance optimization while reducing computation.<sup>45-47</sup>

Recently, deep learning networks for improving segmenta-

tion performance in medical imaging have been continuously proposed. Performing multitasks with segmentation and classification, regression or registration has synergy to gain more precise segmentation performance.<sup>48,49</sup> As segmentation performance increases, studies have been conducted to consider the uncertainty of labels.<sup>50</sup> Furthermore, due to the high cost of medical labels, Semisupervised/unsupervised learning approaches were suggested using unlabelled data.<sup>51-53</sup> Since these studies have not yet surpassed the segmentation performance of supervised learning, it is considered future value as a technology that can overcome the severe imbalances in medical imaging.

#### 4. Image Generation

While many applications using CNN were introduced in medical imaging, it is often challenging to obtain high quality, balanced datasets with labels in the medical domain.<sup>54</sup> Medical images are mostly imbalanced, and time-consuming to obtain their labels. In addition, medical images are hard to obtain due to their privacy issues.<sup>55</sup> To overcome the issues, several studies exploit GAN to make realistic synthetic images of whole X-ray or CT or ROIs of specific lesions, such as liver cancer.<sup>56,57</sup>

GAN is a combination of 2 different neural networks that can generate realistic synthetic images.<sup>58</sup> Since GAN was introduced in 2014, many applications using GANs were introduced in medical imaging. In many studies, GANs were primarily used to generate various imaging modalities such as X-ray, CT, magnetic resonance, positron emission tomography, histopathology images, retinal images, and surgical videos.<sup>56-71</sup> Generated images in the studies were mainly used for data augmentation to have a more balanced dataset for training neural networks of classification or segmentation. With the synthetic images, classification or segmentation accuracies were significant increases than those with the imbalanced dataset.<sup>72</sup>

Another interesting application using GAN is anomaly detection in medical imaging. To generate realistic synthetic images, the model is trying to mimic the distribution of source images in the latent space during the training. If the model learned distribution of normal images (i.e., images without disease), one may use the model as a tool in anomaly detection. By exploiting the DCGAN (deep convolutional generative adversarial network) model,<sup>73</sup> Schlegl et al.<sup>74</sup> introduced an unsupervised anomaly detection method to find a guide marker in OCT images. The method showed a high performance in marker detection (area under the curve = 0.89) but the iteration process was time-consuming. It was further improved for real-time anomaly detection in their recent study by adopting the encoder-decoder

scheme in the model architecture.<sup>75</sup> The studies demonstrated potential of GANs in unsupervised anomaly detection in medical imaging.

#### 5. Image Transformation

History of image to image translation goes back to Hertzmann et al.<sup>76</sup> In this study, a nonparametric model was developed for texture analysis. However, more recent studies focus on using CNN. These studies can be classified into 2 categories including studies with or without GAN.

### IMAGE TO IMAGE TRANSLATION WITHOUT USING GENERATIVE ADVERSARIAL NETWORK

Rise of an image to image translation cannot be separated from style transfer. Gatys et al.<sup>77</sup> used CNN for artistic style transfer. Gu et al.<sup>78</sup> changed loss and reshuffled feature vectors to transfer style. They argue that feature reshuffling can be a complementary solution for parametric and nonparametric neural network style transfer. Though their success of transferring style, overabstraction of features made these algorithms unrealistic. To overcome this hurdle, Li et al.<sup>79</sup> used wavelet transformation as well as multilevel stylization. Following this research, Yoo et al.<sup>80</sup> devised the wavelet pooling layer to enable photorealistic style transfer.

CNN can be used in image denoising. Jain and Seung<sup>81</sup> show frontiers of denoising technique using CNN architecture. They compared the performance of the Markov random field method to that of CNN and showed the CNN network can be used in denoising. Not only CNN but also autoencoder (AE) can be used in denoising. Vincent et al.<sup>82</sup> developed denoising AE, and they also developed stacked denoising AE as well.<sup>83</sup> Batson and Royer<sup>84</sup> used the concept of J-invariant and designed the Noise-2Self concept. Interestingly, this is a single image-level denoising concept. Modality transfer can be also performed with the CNN network. Han<sup>85</sup> used an encoder-decoder network for MRI to CT to transfer modality.

#### 1. Image to Image Translation With Using GAN

Isola et al.<sup>86</sup> used conditional GAN to perform image to image translation with pixel to pixel correspondence. This model is called a pix2pix network. To overcome the limitation that requires pixel to pixel correspondence, Zhu et al.<sup>87</sup> designed CycleGAN architecture which does not require pixel to pixel correspondence. Though CycleGAN can be applied to unmatched

images, style transfer between more than 3 domains requires too many generators – about half of square of the number of domains. Choi et al.<sup>88</sup> solved this issue with one general generator and named this architecture as StarGAN. Chen et al.<sup>89</sup> used GAN to denoising pipeline. They used GAN to estimate noise distribution, and another CNN architecture subtracts estimated noise distribution from the original image. Kudo et al.<sup>90</sup> chose conditional GAN to make thicker CT to thinner CT. They used a 3-dimensional patch to make CT slice thinner. Wolterink et al.<sup>91</sup> borrowed concept of CycleGAN to generate CT from MRI. In short, similar image to image tasks can be performed with or without using GAN. However, there are no consensus or proven facts that which one shows more satisfactory results between using or without using GAN.

## DISCUSSION AND CONCLUSIONS

From the ANN inspired by the human neuronal synapse system in the 1950s to deep learning technology, AI suggests its potential to perform better than humans in some visual and auditory recognition tasks, which may indicate its applications in medicine and healthcare, especially in medical imaging. There could be many kinds of applications of deep learning technology in medical imaging to enhance the burden of medical doctors, quality of healthcare system and patient outcomes. Besides, these kinds of intelligent technology could be applied to precision medicine, which involves the prevention and treatment strategies that consider individual variability.<sup>92</sup> The success of precision medicine is largely dependent on robust quantitative imaging biomarkers, which can be accomplished by deep learning. In particular, imaging is noninvasively and routinely performed for clinical practice and can be used to compute quantitative imaging biomarkers. Many radiomics studies have correlated imaging biomarkers with genomic expression or clinical outcome.<sup>93</sup>

Even with many promising results from previous studies, there are several issues to be resolved before the introduction of deep learning in medical imaging as follows: Firstly, the high dependency on the quality and amount of training dataset, and the tendency of overfitting and bias should be considered. Considering the differences in disease prevalence, imaging modality and protocols in clinical settings across the world, a generalization of deep learning methods should be declared. The evaluation methods to test the performance of each technique, therefore, requires development. Secondly, there could be legal and ethical issues about the use of clinical imaging data for the com-

mercialization, since the performance will be highly dependent on the quality of the data. Thirdly, the black-box nature of the current deep learning technique should be taken into account. Even when the deep learning-based method shows excellent results, in many cases, it is difficult or almost impossible to explain the logical bases of the decision. Lastly, the legal liability issues would be raised if we used a deep learning system in a specific process of clinical practice, independent from the supervision of a physician.

At present, the physician experiences an increasing number of complex readings. This makes it difficult to finish reading in time and provide appropriate reports. However, the deep learning is expected to help radiologists provide a more exact diagnosis, by delivering a quantitative analysis of suspicious lesions, and may also enable a shorter time in the clinical workflow.

Deep learning has already shown comparable performance to humans in recognition and computer vision tasks. These technological changes make it reasonable to think that there might be some major changes in clinical practices. When we consider the use of AI in medical imaging, we anticipate this technological innovation to serve as a collaborative medium in decreasing the burden and distraction from many repetitive and humdrum tasks, rather than replacing physicians. The use of deep learning and AI in radiology is currently in the stages of infancy. One of the key factors for the development and its proper clinical adoption in medicine would be a good mutual understanding of the AI technology, and the most appropriate form of clinical practice and workflow by both clinicians and computer scientists/engineers. Furthermore, there are various other issues including ethical, regulatory, and legal issues to solve and overcome, which should be carefully considered for the development of AI in the use of clinical image data.

## CONFLICT OF INTEREST

The authors have nothing to disclose.

## REFERENCES

1. Murphy KP. Machine learning: a probabilistic perspective. Cambridge (MA): MIT Press; 2012.
2. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In: Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016); 2016 Jun 27-30; Las Vegas (NV), USA. Piscataway (NJ): Institute of Electrical and Electronics Engineers; 2016. p. 770-8.

3. Hinton GE, Osindero S, Teh YW. A fast learning algorithm for deep belief nets. *Neural Comput* 2006;18:1527-54.
4. Rajpurkar P, Irvin J, Zhu K, et al. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv preprint arXiv:1711.05225v3 [cs.CV] 25 Dec 2017.
5. Qin C, Yao D, Shi Y, et al. Computer-aided detection in chest radiography based on artificial intelligence: a survey. *Biomed Eng Online* 2018;17:113.
6. Fenton JJ, Taplin SH, Carney PA, et al. Influence of computer-aided detection on performance of screening mammography. *N Engl J Med* 2007;356:1399-409.
7. Lehman CD, Wellman RD, Buist DS, et al. Diagnostic accuracy of digital screening mammography with and without computer-aided detection. *JAMA Intern Med* 2015;175:1828-37.
8. Gottesman O, Johansson F, Komorowski M, et al. Guidelines for reinforcement learning in healthcare. *Nat Med* 2019;25:16-8.
9. Altman NS. An introduction to kernel and nearest-neighbor nonparametric regression. *Ame Stat* 1992;46:175-85.
10. Hassanat ABA. Two-point-based binary search trees for accelerating big data classification using KNN. *PLoS One* 2018;13:e0207772.
11. Schneider A, Hommel G, Blettner M. Linear regression analysis: part 14 of a series on evaluation of scientific publications. *Dtsch Arztebl Int* 2010;107:776-82.
12. Drucker H, Burges CJ, Kaufman L, et al. Support vector regression machines. In: *Advances in neural information processing systems*. Cambridge (MA): Massachusetts Institute of Technology Press; 1997. p. 155-61.
13. Byvatov E, Fechner U, Sadowski J, et al. Comparison of support vector machine and artificial neural network systems for drug/nondrug classification. *J Chem Inf Comput Sci* 2003;43:1882-9.
14. Amari S. *The handbook of brain theory and neural networks*. Cambridge (MA): MIT press; 2003.
15. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*. Cambridge (MA): Massachusetts Institute of Technology Press; 2012. p. 1097-105.
16. Wang H, Zhou Z, Li Y, et al. Comparison of machine learning methods for classifying mediastinal lymph node metastasis of non-small cell lung cancer from 18F-FDG PET/CT images. *EJNMMI Res* 2017;7:11.
17. Mikołajczyk A, Grochowski M. Data augmentation for improving deep learning in image classification problem. In: *2018 international interdisciplinary PhD workshop (IIPh-DW)*. Piscataway (NJ): Institute of Electrical and Electronics Engineers; 2018. p. 117-22.
18. Lakhani P, Sundaram B. deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology* 2017;284:574-82.
19. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 2016;316:2402-10.
20. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017;542:115-8.
21. Bengio Y, Louradour J, Collobert R, et al. Curriculum learning. In: *Proceedings of the 26th annual international conference on machine learning*; 2009 Jun 14-18; Montreal, Canada. New York: ACM; 2009. p. 41-8.
22. Park B, Cho Y, Lee G, et al. A curriculum learning strategy to enhance the accuracy of classification of various lesions in chest-PA X-ray screening for pulmonary abnormalities. *Sci Rep* 2019;9:15352.
23. Yun J, Park JE, Lee H, et al. Radiomic features and multilayer perceptron network classifier: a robust MRI classification strategy for distinguishing glioblastoma from primary central nervous system lymphoma. *Sci Rep* 2019;9:5746.
24. Lao J, Chen Y, Li ZC, et al. A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme. *Sci Rep* 2017;7:10353.
25. Li Z, Wang Y, Yu J, et al. Deep learning based radiomics (DLR) and its usage in noninvasive IDH1 prediction for low grade glioma. *Sci Rep* 2017;7:5467.
26. Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation. In: *Proceedings of the 2014 IEEE conference on computer vision and pattern recognition*; 2014 Jun 23-28; Washington, DC, USA. Piscataway (NJ): Institute of Electrical and Electronics Engineers; 2014. p. 580-7.
27. Girshick R. Fast r-cnn. In: *Proceedings of the 2015 IEEE international conference on computer vision*; 2015 Jun 7-12; Boston (MA), USA. Piscataway (NJ): Institute of Electrical and Electronics Engineers; 2015. p. 1440-8.
28. Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks. In: *Advances in neural information processing systems*. Cambridge (MA): Massachusetts Institute of Technology Press;

2015. p. 91-9.
29. Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection. In: Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016); 2016 Jun 27-30; Las Vegas (NV), USA. Piscataway (NJ): Institute of Electrical and Electronics Engineers; 2016. p. 779-88.
  30. Redmon J, Farhadi A. YOLO9000: better, faster, stronger. In: Proceedings of the 2017 IEEE conference on computer vision and pattern recognition; 2017 Jul 21-23; Honolulu (HI), USA. Piscataway (NJ): Institute of Electrical and Electronics Engineers; 2017. p. 7263-71.
  31. Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector. In: 2016 European conference on computer vision; 2016 Oct 8-16; Amsterdam, Netherland. Springer. 2016; p. 21-37.
  32. Lin TY, Goyal P, Girshick R, et al. Focal loss for dense object detection. In: Proceedings of the 2017 IEEE conference on computer vision and pattern recognition; 2017 Jul 21-23; Honolulu (HI), USA. 2017. p. 2980-8.
  33. Aslam A, Khan E, Beg MS. Improved edge detection algorithm for brain tumor segmentation. *Proced Comput Sci* 2015;58:430-7.
  34. Hao L. Registration-based segmentation of medical images. Singapore: School of Computing National University of Singapore; 2006.
  35. Pereira S, Pinto A, Alves V, et al. Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Trans Med Imaging* 2016;35:1240-51.
  36. Middleton I, Damber RI. Segmentation of magnetic resonance images using a combination of neural networks and active contour models. *Med Eng Phys* 2004;26:71-86.
  37. Ning F, Delhomme D, LeCun Y, et al. Toward automatic phenotyping of developing embryos from videos. *IEEE Trans Image Process* 2005;14:1360-71.
  38. Brosch T, Yoo Y, Tang LY, et al. Deep convolutional encoder networks for multiple sclerosis lesion segmentation. In: Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention; 2015 Oct 5-9; Munich, Germany. Berlin: Springer; 2015. p. 3-11.
  39. Brosch T, Tang LY, Youngjin Yoo, et al. Deep 3D convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation. *IEEE Trans Med Imaging* 2016;35:1229-39.
  40. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition; 1994 Jun 21-23; Seattle (WA), USA. Piscataway (NJ): Institute of Electrical and Electronics Engineers; 2015. p. 3431-40.
  41. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention; 2015 Oct 5-9; Munich, Germany. Berlin: Springer; 2015. p. 234-41.
  42. Moeskops P, Wolterink JM, van der Velden BH, et al. Deep learning for multi-task medical image segmentation in multiple modalities. In: Proceedings of the 19th International Conference on Medical Image Computing and Computer-Assisted Intervention; 2016 Oct 17-21; Athens, Greece. 2016: Springer. p. 478-86.
  43. Prasoon A, Petersen K, Igel C, et al. Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. *Med Image Comput Comput Assist Interv* 2013;16(Pt 2):246-53.
  44. Çiçek Ö, Abdulkadir A, Lienkamp SS, et al. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Proceedings of the 19th International Conference on Medical Image Computing and Computer-Assisted Intervention; 2016 Oct 17-21; Athens, Greece. 2016: Springer. p. 424-32.
  45. Milletari F, Navab N, Ahmadi SA. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV); 2016 Oct 25-28; Stanford (CA), USA. Piscataway (NJ): Institute of Electrical and Electronics Engineers; 2016. p. 565-71.
  46. Chen W, Liu B, Peng S, et al. S3D-UNet: separable 3D U-Net for brain tumor segmentation. In: International MIC-CAI Brainlesion Workshop; 2018 Sep 16-20; Granada, Spain. Berlin: Springer; 2018. p. 358-68.
  47. Oktay O, Schlemper J, Folgoc LL, et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:180403999* 2018.
  48. Le TLT, Thome N, Bernard S, et al. Multitask classification and segmentation for cancer diagnosis in mammography. *arXiv preprint arXiv:190905397* 2019.
  49. Cao L, Li L, Zheng J, et al. Multi-task neural networks for joint hippocampus segmentation and clinical score regression. *Multimed Tool Appl* 2018;77:29669-86.
  50. Kohl S, Romera-Paredes B, Meyer C, et al. A probabilistic u-net for segmentation of ambiguous images. In: Advances in

- neural information processing systems. Cambridge (MA): Massachusetts Institute of Technology Press; 2018. p. 6965-75.
51. Anirudh R, Thiagarajan JJ, Bremer T, et al. Lung nodule detection using 3D convolutional neural networks trained on weakly labeled data. In: Medical Imaging 2016: Computer-Aided Diagnosis; 2016: International Society for Optics and Photonics. p. 978532.
  52. Hwang S, Kim HE. Self-transfer learning for weakly supervised lesion localization. In: Proceedings of the 19th International Conference on Medical Image Computing and Computer-Assisted Intervention; 2016 Oct 17-21; Athens, Greece. 2016: Springer. p. 239-46.
  53. Feng X, Yang J, Laine AF, et al. Discriminative localization in CNNs for weakly-supervised segmentation of pulmonary nodules. *Med Image Comput Comput Assist Interv* 2017; 10435:568-76.
  54. Ker J, Wang L, Rao J, et al. Deep learning applications in medical image analysis. *IEEE Access* 2017;6:9375-89.
  55. Price WN 2nd, Cohen IG. Privacy in the age of medical big data. *Nat Med* 2019;25:37-43.
  56. Salehinejad H, Valaee S, Dowdell T, et al. Generalization of deep neural networks for chest pathology classification in x-rays using generative adversarial networks. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2018 Apr 15-20; Calgary, Canada. Piscataway (NJ): Institute of Electrical and Electronics Engineers; 2018. p. 990-4.
  57. Frid-Adar M, Diamant I, Klang E, et al. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing* 2018; 321:321-31.
  58. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. In: Advances in neural information processing systems. Cambridge (MA): Massachusetts Institute of Technology Press; 2014. p. 2672-80.
  59. Bozorgtabar B, Mahapatra D, von Teng H, et al. Informative sample generation using class aware generative adversarial networks for classification of chest Xrays. *Comput Vision Image Underst* 2019;184:57-65.
  60. Mirsky Y, Mahler T, Shelef I, et al. CT-GAN: malicious tampering of 3D medical imagery using deep learning. arXiv preprint arXiv:190103597 2019.
  61. Han C, Hayashi H, Rundo L, et al. GAN-based synthetic brain MR image generation. 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018); Piscataway (NJ): Institute of Electrical and Electronics Engineers; 2018. p. 734-8.
  62. Dar SU, Yurt M, Karacan L, et al. Image synthesis in multi-contrast MRI with conditional generative adversarial networks. *IEEE Trans Med Imaging* 2019;38:2375-88.
  63. Lee D, Kim J, Moon WJ, et al. CollaGAN: Collaborative GAN for missing image data imputation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 1994 Jun 21-23; Seattle (WA), USA. Piscataway (NJ): Institute of Electrical and Electronics Engineers; 2019. p. 2487-96.
  64. Bi L, Kim J, Kumar A, et al. Synthesis of positron emission tomography (PET) images via multi-channel generative adversarial networks (GANs). In: Cardoso MJ, Arbel T, Gao F, et al. Molecular imaging, reconstruction and analysis of moving body organs, and stroke imaging and treatment. Berlin: Springer; 2017:43-51.
  65. Wang Y, Yu B, Wang L, et al. 3D conditional generative adversarial networks for high-quality PET image estimation at low dose. *Neuroimage* 2018;174:550-62.
  66. Senaras C, Niazi MKK, Sahiner B, et al. Optimized generation of high-resolution phantom images using cGAN: Application to quantification of Ki67 breast cancer images. *PLoS One* 2018;13:e0196846.
  67. Mahmood F, Borders D, Chen R, et al. Deep adversarial training for multi-organ nuclei segmentation in histopathology images. *IEEE Trans Med Imaging* 2019 Jul 5 [Epub]. <https://doi.org/10.1109/TMI.2019.2927182>.
  68. Zhao H, Li H, Maurer-Stroh S, et al. Synthesizing retinal and neuronal images with generative adversarial nets. *Med Image Anal* 2018;49:14-26.
  69. Andreini P, Bonechi S, Bianchini M, et al. A two stage GAN for high resolution retinal image generation and segmentation. arXiv preprint arXiv:190712296 2019.
  70. Chen Y, Zhong K, Wang F, et al. Surgical workflow image generation based on generative adversarial networks. In: 2018 International Conference on Artificial Intelligence and Big Data (ICAIBD); 2018 May 26-28; Chengdu, China. Piscataway (NJ): Institute of Electrical and Electronics Engineers; 2018. p. 82-6.
  71. Engelhardt S, De Simone R, Full PM, et al. Improving surgical training phantoms by hyperrealism: deep unpaired image-to-image translation from real surgeries. In: Proceedings of the 21st International Conference on Medical Image Computing and Computer Assisted Intervention ; Granada, Spain. 2018: Springer. p. 747-55.

72. Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. *J Big Data* 2019;6:60.
73. Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434 2015.
74. Schlegl T, Seeböck P, Waldstein SM, et al. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In: *The 25th biennial international conference on Information Processing in Medical Imaging (IPMI 2017)*; 2017 Jun 25-30; Boone (NC), USA. Berlin: Springer; 2017. p. 146-57.
75. Schlegl T, Seeböck P, Waldstein SM, et al. f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks. *Med Image Anal* 2019;54:30-44.
76. Hertzmann A, Jacobs CE, Oliver N, et al. Image analogies. In: *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*; 2001 Aug 12-17; Los Angeles (CA), USA New York: ACM; 2001. p. 327-40.
77. Gatys LA, Ecker AS, Bethge M. Image style transfer using convolutional neural networks. In: *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016)*; 2016 Jun 27-30; Las Vegas (NV), USA. Piscataway (NJ): Institute of Electrical and Electronics Engineers; 2016. p. 2414-23.
78. Gu S, Chen C, Liao J, et al. Arbitrary style transfer with deep feature reshuffle. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; 1994 Jun 21-23; Seattle (WA), USA. Piscataway (NJ): Institute of Electrical and Electronics Engineers; 2018. p. 8222-31.
79. Li Y, Fang C, Yang J, et al. Universal style transfer via feature transforms. In: *Advances in neural information processing systems*. Cambridge (MA): Massachusetts Institute of Technology Press; 2017. p. 386-96.
80. Yoo J, Uh Y, Chun S, et al. Photorealistic style transfer via wavelet transforms. arXiv preprint arXiv:1903.09760 2019.
81. Jain V, Seung S. Natural image denoising with convolutional networks. In: *Advances in neural information processing systems*. Cambridge (MA): Massachusetts Institute of Technology Press; 2009. p. 769-76.
82. Vincent P, Larochelle H, Bengio Y, et al. Extracting and composing robust features with denoising autoencoders. In: *Proceedings of the 25th international conference on Machine learning*; 2008 Jul 5-9; Helsinki, Finland. New York: ACM; 2008. p. 1096-103.
83. Vincent P, Larochelle H, Lajoie I, et al. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *J Mach Learn Res* 2010;11:3371-408.
84. Batson J, Royer L. Noise2self: Blind denoising by self-supervision. arXiv preprint arXiv:1901.11365 2019.
85. Han X. MR-based synthetic CT generation using a deep convolutional neural network method. *Med Phys* 2017;44:1408-19.
86. Isola P, Zhu JY, Zhou T, et al. Image-to-image translation with conditional adversarial networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*; 1994 Jun 21-23; Seattle (WA), USA. Piscataway (NJ): Institute of Electrical and Electronics Engineers; 2017. p. 1125-34.
87. Zhu JY, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: *Proceedings of the 2017 IEEE international conference on computer vision*; 2017 Jul 21-23; Honolulu (HI), USA. Piscataway (NJ): Institute of Electrical and Electronics Engineers; 2017. p. 2223-32.
88. Choi Y, Choi M, Kim M, et al. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; 1994 Jun 21-23; Seattle (WA), USA. Piscataway (NJ): Institute of Electrical and Electronics Engineers; 2018. p. 8789-97.
89. Chen J, Chen J, Chao H, et al. Image blind denoising with generative adversarial network based noise modeling. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; 1994 Jun 21-23; Seattle (WA), USA. Piscataway (NJ): Institute of Electrical and Electronics Engineers; 2018. p. 3155-64.
90. Kudo A, Kitamura Y, Li Y, et al. Virtual thin slice: 3D conditional GAN-based Super-resolution for CT slice interval. In: *International Workshop on Machine Learning for Medical Image Reconstruction*; 2019 Oct 17; Shenzhen, China. Berlin: Springer; 2019. p. 91-100.
91. Wolterink JM, Dinkla AM, Savenije MH, et al. Deep MR to CT synthesis using unpaired data. In: *International Workshop on Simulation and Synthesis in Medical Imaging*; 2017 Sep 10; Québec City, Canada. Berlin: Springer; 2017. p. 14-23.
92. Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med* 2015;372:793-5.
93. Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 2014;5:4006.