Radiological Physics and Technology (2020) 13:6–19
https://doi.org/10.1007/s12194-019-00552-4

AI-based computer-aided diagnosis (AI-CAD): the latest review to read first

Hiroshi Fujita[1]

Received: 11 December 2019 / Revised: 19 December 2019 / Accepted: 21 December 2019 / Published online: 2 January 2020

Abstract
The third artificial intelligence (AI) boom is coming, and there is an inkling that the speed of its evolution is quickly increasing. In games like chess, shogi, and go, AI has already defeated human champions, and the fact that it is able to achieve autonomous driving is also being realized. Under these circumstances, AI has evolved and diversified at a remarkable pace in medical diagnosis, especially in diagnostic imaging. Therefore, this commentary focuses on AI in medical diagnostic imaging and explains the recent development trends and practical applications of computer-aided detection/diagnosis using artificial intelligence, especially deep learning technology, as well as some topics surrounding it.

Keywords Artificial intelligence (AI) · Computer-aided detection/diagnosis (CAD) · Medical image recognition · Machine learning · Deep learning · Convolutional neural network

1 Introduction

As shown in Fig. 1, the third artificial intelligence (AI) boom is gaining momentum. In particular, the field of computer-aided detection/diagnosis (CAD) for medical images is starting to experience changes with the advent of an advanced artificial neural network called deep learning, which is a part of machine learning technology. Figure 2 depicts the interrelationship of the four terms related to AI and presents their basic meanings; these have recently started being used frequently.

The term "deep learning" comes from Geoffrey Hinton, Emeritus Professor of the University of Toronto, who is now known as the "Godfather" of AI. This is because, in his 2006 paper, he referred to a deep multi-layered neural network as a generic term for deep learning. In an image recognition contest called ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012, Hinton and his colleagues used a deep learning program to reduce the recognition error rate, or false recognition rate, of the winning record from the previous year by 40%, from 25.7 to 15.3%, which was a historical overwhelming victory. Hinton made a presentation at an international conference in autumn 2016, saying that "deep learning will reach the level of a specialist in radiology within five years."

In the following section, we will present an overview of what is happening in the medical diagnostic imaging field in AI and machine learning, particularly in deep learning. Readers interested in more in-depth content may refer to the references given as examples [1–9].

2 Traditional CAD

As mentioned above, the third AI boom has arrived (Fig. 1), and the CAD domain has a strong tailwind, which is provided by the deep learning technology that will be described in Sect. 4. First, the background of CAD will be described.

Basic research on CAD, including research and development aimed at automatic diagnosis, began in the 1960s. The first device to be approved by the U.S. Food and Drug Administration (FDA) in 1998 was a mammography CAD device made by R2 Technology, Inc. (now Hologic). Therefore, 1998 is sometimes called "CAD first year." This device, shown in Fig. 3, was basically developed at the University of Chicago [10]. The Centers for Medicare and Medicaid Services approved reimbursement for the use of mammography CADs in 2002 in the United States, and this was a significant

✉ Hiroshi Fujita
hiroshi.fujita.gifu@gmail.com

1 Department of Electrical, Electronic and Computer Engineering, Faculty of Engineering, Gifu University, 1-1, Yanagido, Gifu City, Gifu 501-1194, Japan

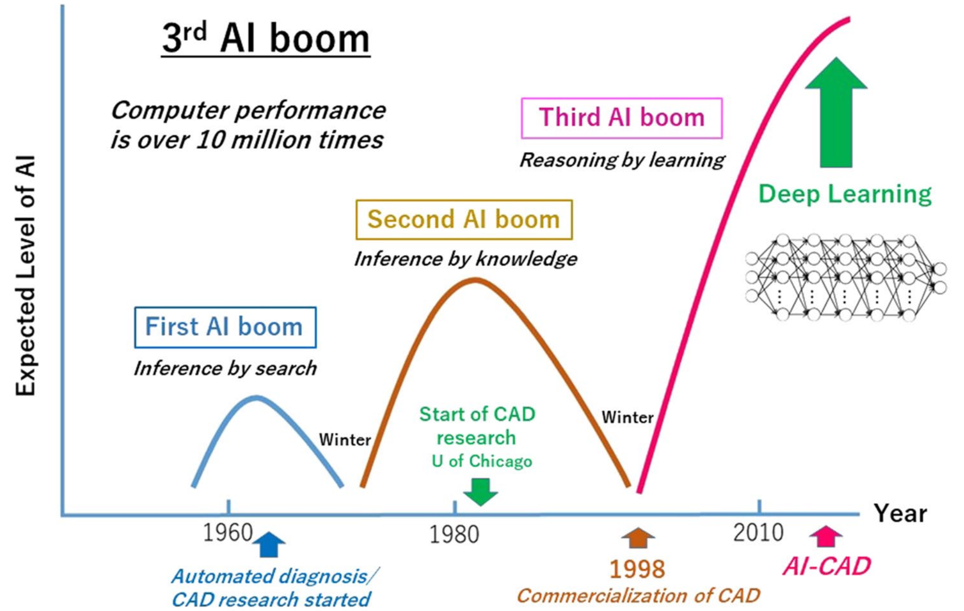

Fig. 1 Two AI booms in the past and the current third AI boom. CAD research progress is deeply related to AI technology progress

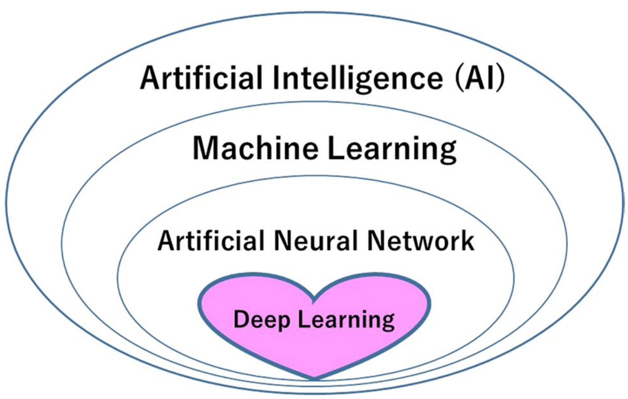

Fig. 2 Deep learning in AI. AI is an abbreviation of artificial intelligence, a research field that programs computers to have human-like intelligence. Machine learning is one of the important techniques in AI that makes machines (computers) smarter by giving them experience-based learning. An artificial neural network (ANN) is a machine learning technology that models information processing in the brain. Deep learning is leading the current AI boom, driven by machine learning methods, which have been developed by layering conventional ANNs and new learning techniques and by improving computer performance

factor that spurred the spread of CADs. Its reimbursement rate for CAD in 2003 was $19.13 per exam.

After that, commercialization has continued in the diagnostic imaging area to include chest radiographs, chest CT images, and CT colonoscopies. However, although it is already past the first 20 years of commercialization of CAD, it is no exaggeration to say that the most successful example of CAD used in clinical practice in the U.S. remains mammography CAD; CAD was used in about 92% of screening mammography readings in 2016 [11]. On the other hand, in Japan, until the end of 2018,[1] the only CAD system that has obtained regulatory approval was for mammography CAD.

While the need for CAD in the clinical setting has been sufficiently recognized, the limitations found in such conventional or traditional CADs are mainly owing to the following reasons:

- –High development cost.
- –A high rate of false-positive marks, possibly leading to an increased recall rate and unnecessary biopsies.
- –Not always effective in a clinical evaluation.
- –Workflow and cost-effectiveness issues.
- –Limited to specific lesions.

In terms of mammography cases, there are many published papers on clinical evaluation [12–15], some of which show the disadvantages of FDA-approved mammography systems assessed in clinical situations with the prospective procedure.

The next-generation CAD that solves these problems is expected to be AI-based and likely to use deep learning technology.

[1] On December 6, 2018, approval for the software to determine if it is a neoplastic polyp or not with possibility shown as a number in the super-magnifying endoscope was freshly obtained, based on the Pharmaceuticals and Medical Devices Law in Japan. Support vector machine (SVM) type machine learning instead of deep learning is used in this system.

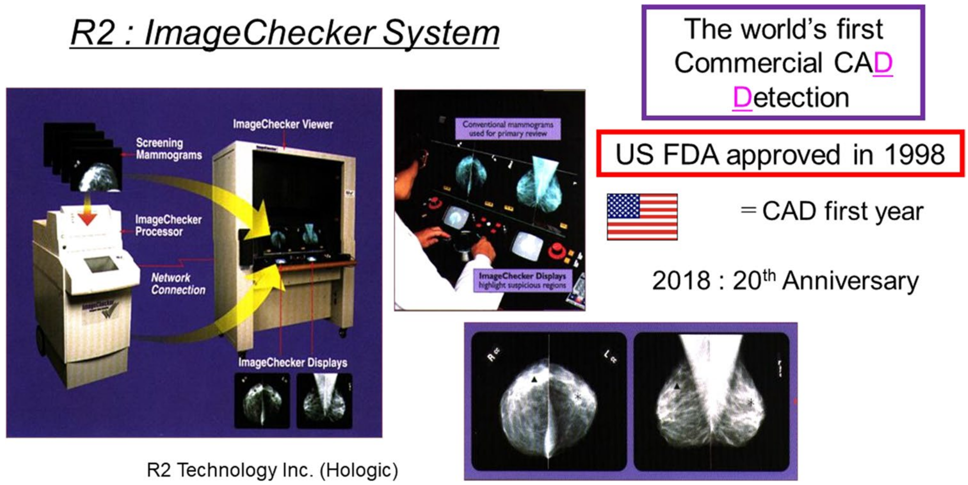

Fig. 3 World's first commercial CAD device (mammography CAD, ImageChecker M1000 system). The ImageChecker system is a computer-aided detection (CAD) system rather than a computer-aided diagnosis system (see 5.2.1). (Parts of this figure were taken from the homepage of former R2 Technology, Inc. Los Altos CA, USA, now Hologic)

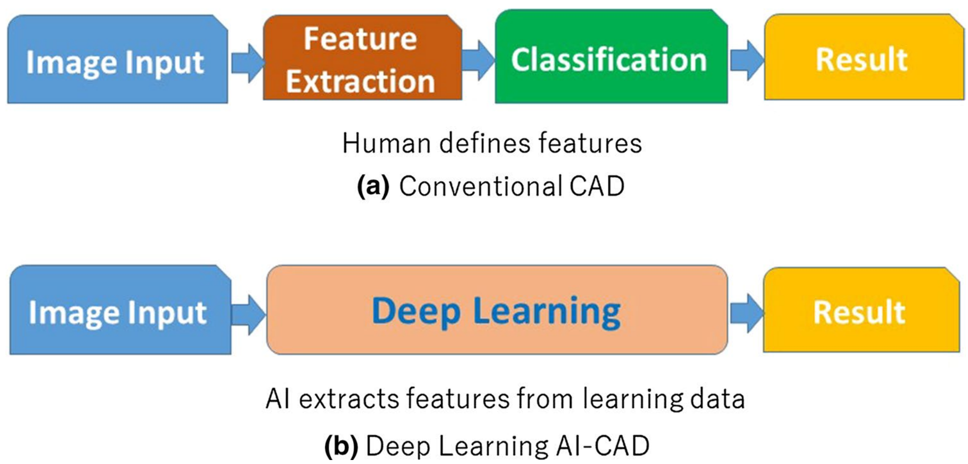

Fig. 4 Differences in the development process of **a** conventional CAD and **b** deep learning-based AI-CAD. The development process of CAD was changed significantly by deep learning

3 AI-CAD

In conventional CAD development, the human designer struggles to devise and create feature quantities that should be recognized in the image. Some such feature quantities are the shape and density information of the cancerous area. The advantage of deep learning is that it can create such features by itself through its learning process. That is, while conventional CAD follows the development process, as shown in Fig. 4a, AI-CAD using deep learning goes through the development process shown in Fig. 4b, saving much time and effort. In the past, development took years to complete, but with deep learning, development can be done in months. In this paper, deep learning CAD is referred to as AI-CAD. Although conventional CAD uses machine learning technology, like support vector machines, it is not included in AI-CAD in this paper.

4 Deep learning

4.1 AI-CAD propulsion engine

Deep learning is an evolutionary version of an artificial neural network that artificially models the neural network of the human brain with a computer. The reason for calling it the evolved version of a neural network is that there have already been two neural network booms in the past, in response to the past two AI booms. Depicting deep learning graphically, it looks much like Fig. 5. In the figure, circles represent individual neurons, and "arrow lines" connecting neurons represent weighting factors. The hierarchical structure is simulated from the bottom input layer to the top output layer, and it has a multilayer structure in which many middle layers are connected. It is called deep learning because there are many middle layers that are expressed as deep. During the second neural

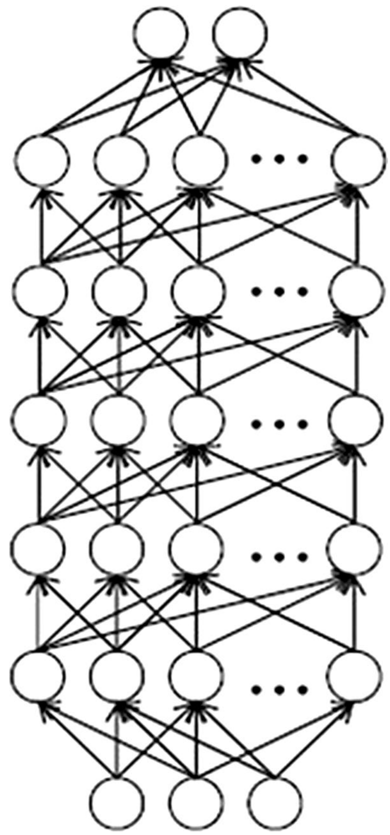

Fig. 5 Deep learning with a multi-layered structure. The symbol open circle represents one neuron. Information is propagated from the lowermost input layer to the uppermost output layer through multiple intermediate layers

network boom, a three-layer structure was common; this is no longer necessarily true.

Deep learning basically has a structure called a convolutional neural network (CNN). As shown in Fig. 6, this structure has three types of layer, convolution, pooling, and total connection, stacked in multiple layers. For example, if the entire chest radiograph is set up as shown in Fig. 6, the discrimination of abnormality or normality can be output. Alternatively, with respect to the input obtained by cutting out the periphery of the tumor, a classification of malignant or benign can be output. In addition, if the entire chest radiograph is input, the abnormal area can be enclosed in a rectangular frame in the output for detection, or the lesion area can be determined in pixel units for region extraction and segmentation. Further, it is even possible to point out each anatomical structure, e.g., to recognize the heart region or the upper lobe of the right lung. One representative example for multiple organ segmentation in torso CT scans is illustrated in Fig. 7, from the research of Zhou et al. at Gifu University [16]. Furthermore, another possible application of deep learning is in estimation or regression. For example, bone age estimation is possible from the radiograms. It is also applied to image formation, including image quality improvement, such as super-resolution, noise reduction, and image reconstruction in CT scans.

In deep learning, a system is completed by adjusting hyperparameters with a large amount of learning data. Therefore, deep learning AI-CAD is said to be data-driven, and the outcome that determines its performance depends on the power of image data collection, as demonstrated in

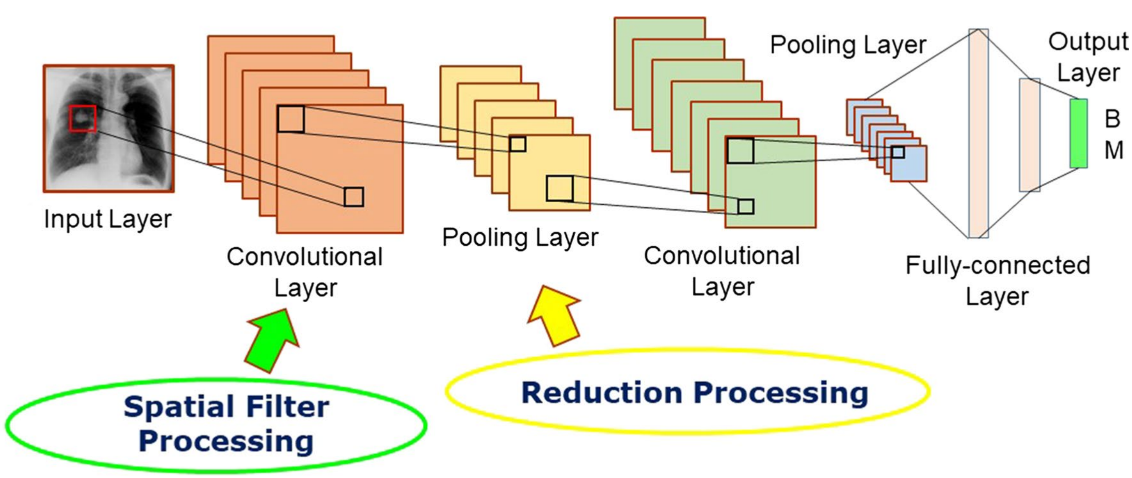

Fig. 6 Example of benign (B) and malignant (M) classification of tumors on chest radiographs by a CNN. Feature extraction processing is performed in a series of convolutional layers and pooling layers, which are intermediate layers, and discrimination and classification processing are performed in all subsequent fully connected layers. The convolutional layers work the same as spatial filtering in conventional image processing, and pooling layers are something like reduction processing. Note that Fig. 5 rotated 90 degrees to the right corresponds to this figure

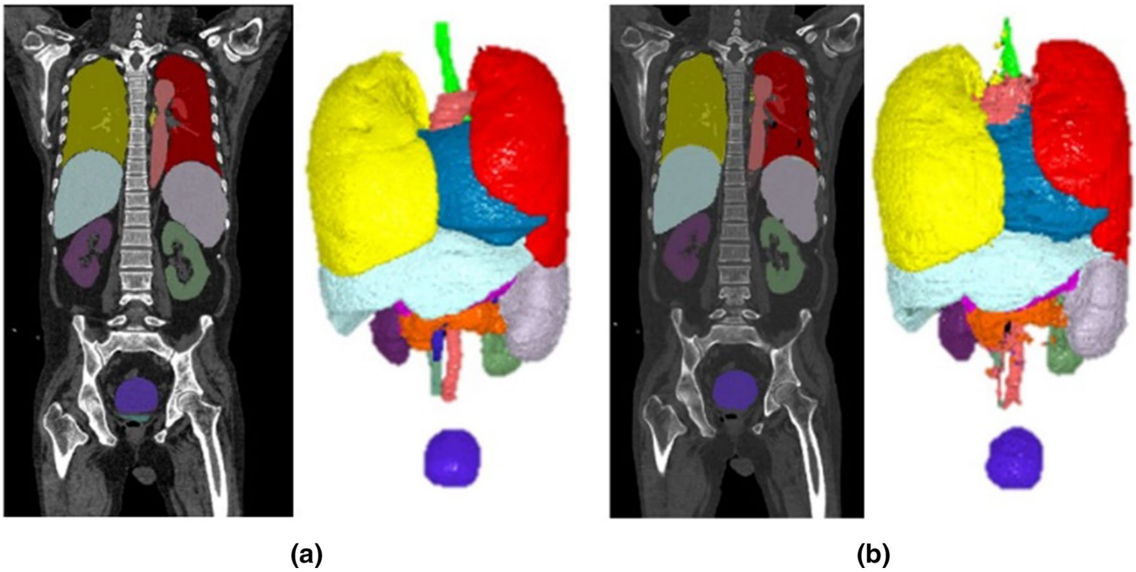

Fig. 7 Organ segmentation of torso CT scans using deep learning [16]. **a** Labeled image as ground truth made by medical doctors. **b** Results of automatically segmented organs based on deep learning technology. Image courtesy of Dr. Zhou of Gifu University

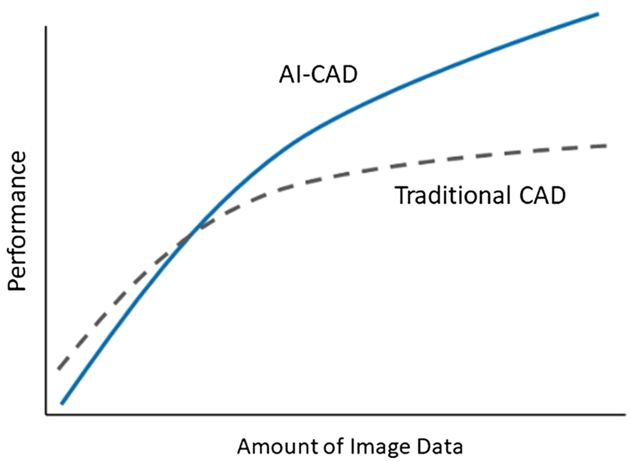

Fig. 8 Comparison of the performance of traditional CAD and deep learning-based AI-CAD on data volume. Deep learning is "Data-Driven and Data-Hungry!" Deep-learning AI-CAD is expected to show better performance than conventional CAD as the amount of image data increases

Fig. 8. Not only is the amount of data large, but also the quality of the data is important, and correct data are also required during learning. These data are called labeled data, e.g., marginal areas of a tumor, or tagged data, i.e., the presence or absence of a tumor or type of tumor. The tools needed for deep learning are available as open source so that software can be easily obtained as long as the computer as hardware is at hand (Fig. 9); this is an age of "AI democratization" in which anyone can try deep learning. A paper in [17] shows an interesting result, where even healthcare professionals with no coding or deep learning experience were able to develop useful models with comparable discriminative performance and diagnostic properties to state-of-the-art deep learning algorithms by AI engineers, by simply using four of the five publicly available open datasets.

4.2 Database

As mentioned above, although deep learning requires a large amount of data, the number of medical images is less than the number of general natural images. However, looking at reported cases of recent developments by academic societies and companies, there is a rough impression that a system is first built with 1000 cases, system performance is improved with 10,000 cases, and then practical use aims for 100,000 cases or even more. One wonders if the future development of AI means that only those with large-scale data will get the winner's seat.

4.2.1 Public dataset

Public medical image databases are increasing. For example, the U.S. National Institutes of Health (NIH) has a large-scale database of approximately 110,000 chest X-ray images, named ChestX-ray8 database [18]. There is also a nearly 10,600-body trunk CT database with more than 4400 unique patients at the NIH Clinical Center in Bethesda, named DeepLesion, which has been published to support the development and testing of AI algorithms for medical applications. The database is available for download as of October 18, 2019 [19].

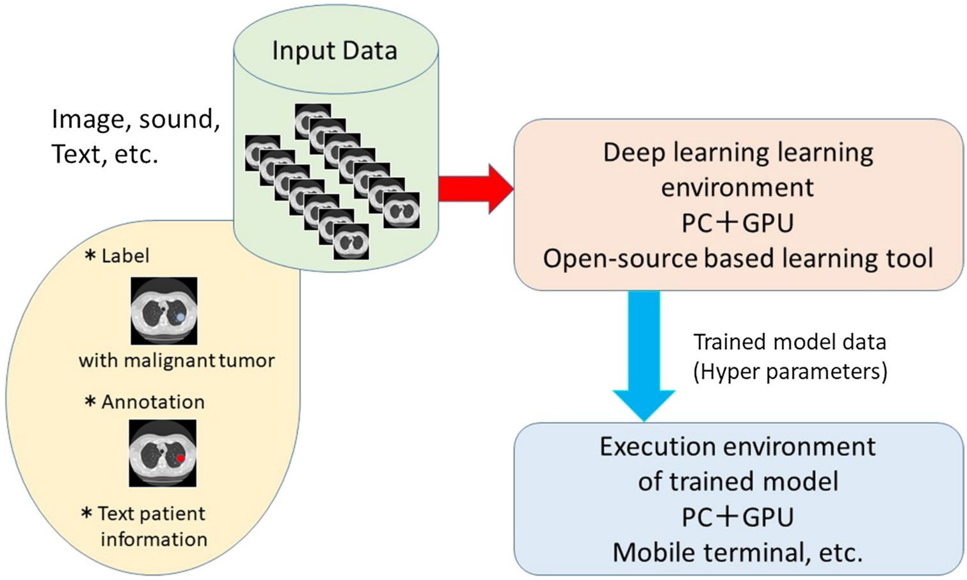

Fig. 9 Learning and recognition execution environment: tools needed for deep learning. *GPU* graphics processing unit

In addition, more than 224,000 chest radiographs called CheXpert have been released by Stanford University [20]. In a study of the same university group using this database, the number of images required for deep learning processing is reported to be 20,000 for the distinction of benign and malignant tumors in chest radiographs [21]. In mammography, there are DDSM, IRMA, INbreast, MIAS, BCDR, etc., but most of these are film-digitized mammogram databases; for details, see [22] and the CANCER IMAGING ARCHIVE site in [23]. Furthermore, many contests targeting medical images are also held, i.e., the "CHALLENGE," and it is possible to compare software performance with common image data.

In Japan, Japan Radiological Society, Japan Gastroenterological Endoscopy, Japanese Society of Pathology, Japanese Ophthalmological Society, Japanese Dermatological Association, and Japan Society of Ultrasonics in Medicine are currently collecting such images, aiming to develop a diagnostic support system, and constructing a database with the support of the Japan Medical Research and Development Organization. The National Institute of Informatics has also established the Medical Big Data Research Center. Furthermore, as a national measure for data use in the medical field, the Next-Generation Medical Infrastructure Law was enacted on May 11, 2018, and the defined system is being developed.

4.2.2 Augmentation and transfer learning

There are two common technologies for increasing the data in deep learning, data augmentation and transfer learning/fine tuning.

Data augmentation is a technique that increases the amount of data by pre-processing input image data such as rotation, contrast enhancement, and noise addition, and this technique is used frequently.

In deep learning medical image applications, a deep learning model that has already been trained on a large natural image database of over 14 million images of 20,000 types called an ImageNet is used. Further learning is performed on the target image, and this procedure is called fine tuning. Such a technique is overall called transfer learning and is a commonly used and very effective method. An effective example of transfer learning in classifying benign and malignant mammographic masses can be found in the paper by Huynh et al. [24], where deep learning itself was used as an image feature extractor. Also, Samala et al. conducted a study to transfer a model that had already been trained in mammography to detect mass shadows in tomosynthetic breast images, showing the effectiveness of transfer learning [25].

The ImageNet dataset consists of only 2D data of natural images, but there exist many 3D types of data such as CT and MRI in medical imaging. Therefore, another strong possibility is to use a highly trained model using medical data. The GrayNet deep learning model for predicting age, weight, and gender, and also segmenting each of 24 body parts, trained with CT images from 6 CT machines, is such an approach [26, 27], where the pre-trained model can be applied to improving the development of other deep learning applications, including organ segmentation. In addition, NVIDIA recently presented 13 different organ models pre-trained on public datasets for organ segmentation in CT images [28].

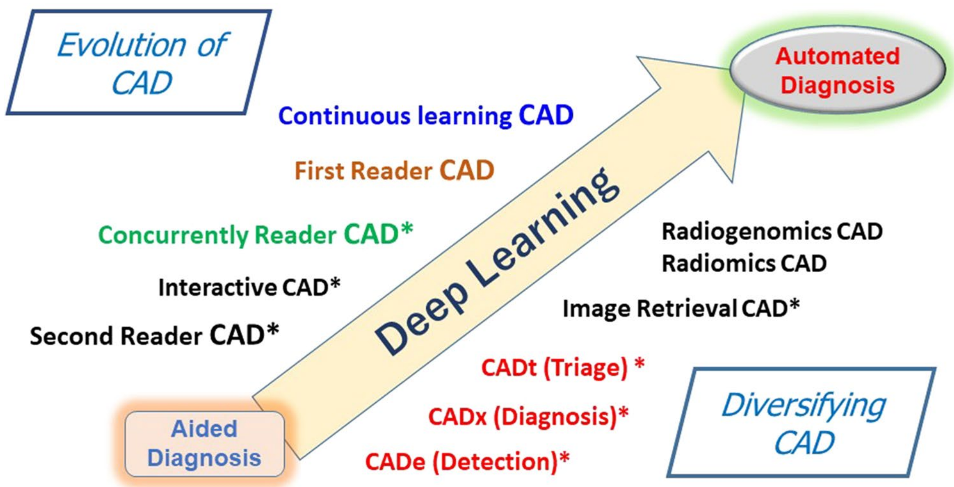

Fig. 10 CAD evolution and diversification. The emergence of deep learning brings about the evolution and diversification of CAD. *Commercialized

4.2.3 GAN

There is a deep learning method called the generative adversarial network (GAN), which is said to be the most interesting idea of the past 10 years (Director, LeCun, AI Research Institute, Facebook). As image data can be newly generated using GAN, research in that area is beginning to produce results [29].

As an effective example, Fris-Adar et al. showed an interesting result of GAN-based image data augmentation in the classification task of liver lesions on CT scans [30]. The performance by only classic augmentation and additional synthetic data augmentation yielded 78.6% and 85.7% in sensitivity and 88.4% and 92.4% in specificity, respectively. Teramoto et al. successfully employed the GAN technique for pulmonary nodule classification in CT scans, where they initially trained the CNN using GAN-generated images instead of ImageNet [31, 32]. Their results indicated that the proposed GAN method improved the classification accuracy by approximately 20% in comparison to training done using only the original images [31], and continuously demonstrated a further improved result when multiplanar images were used instead of single axial images [32].

Moreover, the GAN technique can also be applied to image quality improvements in terms of spatial resolution, image noise, and artifacts [29].

4.2.4 Federated learning

Federated learning, introduced by Google in 2017 [33], is defined as collaborative machine learning without centralized training data and updates common models with distributed data, often called distributed learning. This technology can be appropriate under medical imaging circumstances, when sharing of patient data between institutions often has limitations due to technical, legal, and/or ethical concerns [33–35]. In 2018, Intel began a collaboration with the Center for Biomedical Image Computing and Analytics at the University of Pennsylvania to show the first “real-world” proof-of-concept application of federated learning in medical imaging [35, 36]. Using this technology, only the model parameters, such as learning weights in the network model, are transferred among institutions. Hence, it is sometimes called privacy preserving.

5 CAD evolution and diversification

Using deep learning technology in the engine part of CAD, conventional CAD is now referred to as new-generation CAD, AI-CAD, defined as in Sect. 3; this technology has begun to evolve and diversify, as depicted in Fig. 10.

5.1 Evolution

We may categorize CADs based on how to use the circumstances surrounding the physician’s image interpretation:

- Second-reader CAD
- Concurrent-reader CAD
- First-reader CAD

The difference between these three types of CAD is clarified in Fig. 11 [37].

An additional type of CAD is an interactive CAD; this is discussed in Sect. 5.1.2.

5.1.1 Second-reader CAD

In 1998, second-reader CAD was the first type of CAD system approved by the FDA for mammography. In this system, a physician first interprets images without CAD, and then

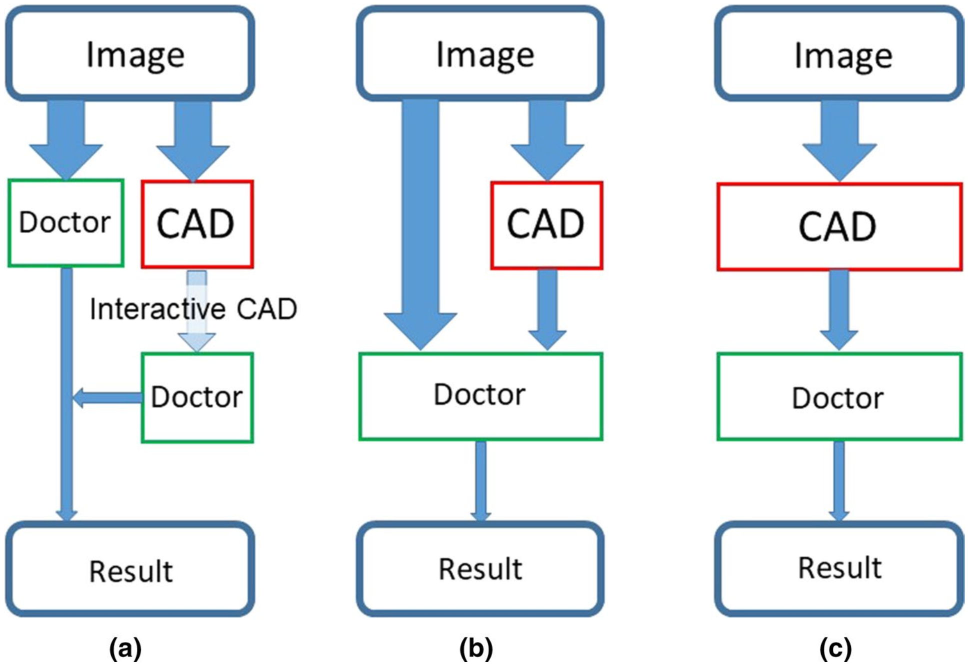

Fig. 11 Three types of CAD based on how they are used in image interpretation (modified from [37]). **a** Second-reader type, **b** concurrent-reader type, and **c** first-reader type. Note that interactive CAD is included as part of (**a**)

CAD is used in another step, as depicted in Fig. 11a; hence, the interpretation time tends to be longer. In addition, the system detects candidate positions of breast cancer such as masses and clustered microcalcifications and is called computer-aided detection (hereinafter termed CADe, as it is referred to in Fig. 10).

Subsequently, mammography CADe from several companies, CADe for chest radiographs and CTs for detecting lung nodules, and CT colonography CADe for detecting polyps were successfully commercialized. The commercialization of so-called computer-aided diagnosis, referred to as CADx in Fig. 10, which has a function to support differential diagnosis of benign and malignant tumors, has been long delayed. However, the recent emergence of deep learning has caused some major changes in the CAD world.

5.1.2 Interactive CAD

Interactive CAD is a CAD system developed by Dutch researchers in which CAD markers are not shown at the lesion site. If there is a suspicious area, an annotation explaining the analysis results with a computer-estimated malignancy score is available if the doctor clicks on it [38], as noted in Fig. 11a. In a recent paper in radiology [39], the authors showed that radiologists improved their cancer detection in mammography when using a deep learning-based interactive CAD system for support that did not require additional reading time. They also found that the breast cancer detection performance of the system was similar to the radiologists' average performance. In December 2018, ScreenPoint Medical, a Dutch venture company, received FDA approval for such a style of mammography CAD.

5.1.3 Concurrent-reader CAD

Simultaneous reader-type CAD, also called concurrent-reader CAD, was developed and commercialized as a method of referring to the results of CAD from the beginning of interpretation and is shown in Fig. 11b.

The first concurrent-reader CAD system, ClearRead CT for chest CTs, was developed by Riverin Technologies and approved by the FDA in September 2016. The system consists of two parts, a suppression processing unit for blood vessel shadows in the chest image and a detection processing unit for different types of nodules.

In November 2016, QView Medical's CADe system for 3D breast ultrasound imaging also received FDA approval. Furthermore, in March 2017, the CADe system for digital breast tomosynthesis (DBT) by iCAD, Inc. received FDA approval. As these are three-dimensional images, it is necessary to interpret a large number of images per patient; the expectation for simultaneous reader-type CAD development is large, and experimental results are effective in shortening the interpretation time. In December 2018, iCAD received another FDA clearance for DBT. This is a concurrent-reader-type CADe, which also functions as a CADx, showing a score indicating the possibility of cancer.

5.1.4 First-reader CAD

The first-reader CAD is a type of CAD in which the CAD first performs an interpretation process alone, and the physician's interpretation is restricted to only the CAD-marked images, as shown in Fig. 11c; hence, reading time may be significantly reduced. It is strongly desirable to use it from the physician's point of view, especially in the case of mass screenings, such as in mammography.

The study by Kyono et al. investigated whether deep learning could reduce the number of mammograms that a radiologist must interpret using a deep learning classifier to identify normal mammograms correctly and to select the uncertain and abnormal examinations for radiological interpretation [40]. They found that while maintaining an acceptable negative predictive value of 99%, their proposed model was able to identify 34% and 91% of the negative mammograms for test sets with a cancer prevalence of 15% and 1% by assuming a theoretical diagnostic setting and assuming a theoretical screening setting, respectively. This result implies that deep learning could be used to reduce the number of normal mammograms that radiologists need to read without degrading diagnostic accuracy.

However, it may take some time for such products to appear on the market.

5.1.5 Continuous learning CAD

Expectations are rising with the advent of deep learning with learning functions, and there is CAD with a learning function after marketing. This is a CAD algorithm that assumes that learning will be continued with new data and become smarter even after the introduction of CAD. In contrast to such adaptive algorithms, all the traditional CAD systems approved by the FDA were based on so-called "locked" algorithms.

The world's first cloud-based heart disease diagnosis support AI system for cardiac MR images was made by Arterys Inc., a U.S. venture company, and received FDA approval in January 2017. After its introduction, it was said that the program update of the system learned after marketing was implemented on a scale of five times a year by the mechanism that is re-trained by new data stored on the cloud. Although this system falls into the category of radiation image analysis and processing support systems, such full-scale CADs with post-marketing learning functions will appear in the near future.

However, to realize such a system, several issues must be solved. Some such issues are whether post-marketing learning by companies is possible and whether on-site post-marketing learning can be done by the users themselves. If it is possible, how often it is better to learn later, who will manage it when the facility's software version is born, who will deal with it if the performance should fall, how to deal with it when it has fallen, and so on. To discuss these issues, guidelines concerning regulatory review and development of the latest AI-CAD system corresponding to such post-marketing learning functions have been and are currently being discussed by some related organizations [41, 42].

The FAD's approval criteria have also evolved and diversified in line with the AI era, and the number of AI-CADs approved by FDA has increased dramatically over the last few years.[2]

5.2 Diversification

5.2.1 Three types of CAD

There are three types of CAD classified based on their purposes.

– CADe (detection).
– CADx (diagnosis).
– CADt (triage).

CADe and CADx are used for the detection of diseases on the image and classification of lesions into benign and malignant and have been explained in previous sections.

All the CADs commercialized from 1998 until very recently were only CADe. However, in July 2017, a second-reader CADx for breast MRIs was developed by Quantitative Insights, and the FDA finally approved it as the first CADx. CADx for breast MRIs presents an index indicating the grade of QI score for the candidate lesion site and has a similar image presentation function.

Triage refers to selection based on the severity of a patient's illness or trauma. It is an important viewpoint, especially in the field, when examining many injured persons during disasters. There is a system that applies and extends CAD technology, analyzes the image immediately after radiographing and before the radiologist interprets it, and presents or warns the specialist of the urgency of treatment for this patient; this system is called CADt. This type of software is termed radiological computer-assisted prioritization software for lesions. In CADt, fully automatic and initial computer analysis is performed by CAD technology, and a simple output of positive or negative or serious/minor/normal is output from the computer [43]. This means that, in this system, the questions of

[2] The American College of Radiology (ACR) Data Science Institute (DSI) has created a new resource for radiology researchers. A complete list of AI algorithms cleared by the FDA related to medical imaging, published on the ACR DSI website, is said to be updated regularly. https://www.acrdsi.org/DSI-Services/FDA-Cleared-AI-Algorithms

regarding whether the condition of the patient is urgent or not and whether the patient's life is in imminent danger are answered. CADt also may have a function to evaluate and judge the image quality, and this system requires higher performance than conventional CAD.

In February 2018, Viz.ai, Inc.'s CAD using deep learning received the FDA's first approval for CADt. This is a major vessel occlusion (LVO) stroke platform that identifies LVO in CT systems of emergency room (ER) patients. It is also connected to a CT device, which warns a stroke specialist (brain neurologist) that a suspicion of a stroke has been confirmed, and it can transmit an image directly to the smartphone of the relevant doctor. In more than 95% of cases, automatic notification to specialists reduced the notification time by an average of 52 min. It is also referred to as "clinical decision support software." In addition, a triage-like system, a system that uses techniques such as CAD to prioritize the interpretation of cases according to the degree of urgency of the patient's disease content determined from the images, is being developed by a company and is going to be incorporated into image interpretation systems.

In May 2019, the FDA gave 510(k) clearance to an AI alert for an urgent finding of a collapsed lung (pneumothorax) in chest X-rays developed by Zebra Medical Vision. The approval is a first for an AI-based chest X-ray solution that can help doctors make quicker diagnoses from one of the world's most used imaging modalities. This company also announced in June 2019, that its AI solution for CT intracranial hemorrhage (ICH) alerts has also gained FDA approval. Moreover, Zebra company received a fourth FDA clearance for the identification and triaging of pleural effusion in chest X-rays in Nov. 2019.

In September 2019, GE Healthcare received FDA 510(k) clearance for the first AI algorithms embedded on a device to prioritize critical chest X-ray reviews, which helps radiologists prioritize critical cases with a suspected pneumothorax by immediately flagging critical cases for triage. This AI offers first-of-its-kind automated AI quality check features that detect acquisition errors, flagging images for technologists' review and allowing them to make corrections before they go to radiologists for review.

5.2.2 Similar case retrieval

A similar image presentation retrieved using a content-based image retrieval technique would help assist the decision of malignancy by physicians [44]. In this manner, CADx can be used as a supplemental tool for the computerized likelihood of malignancy in differential diagnosis. In this area of image retrieval, the deep learning methodology is applied; for details, refer to [45].

5.2.3 Radiomics/radiogenomics

Radiomics refers to the field in which patient scans (i.e., images) are converted into higher-dimensional quantitative data, and subsequently, these data are mined for improved decision support, while radiogenomics is a specific application in which imaging features, radiomic or otherwise, are linked to genomic profiles [46–48]. For radiomics, CAD includes support by prognosis and prediction; radiogenomic CAD for diagnostic support combines genetic test results and imaging tests. The former is a category of CAD based on conventional image analysis and would be limited even in its expanded application, while the latter is a CAD that combines image information and gene information, thereby leading to significant expectations for the future.

Bodalal et al. described the outline of the two kinds of radiomic pipelines [47]. One is the classical/conventional radiomics model in which after image acquisition, areas of interest are delineated, and handcrafted features are extracted; thereafter, a model is built around these predefined features using either statistical or machine learning techniques. The other is a deep learning radiomics model, where after image acquisition, deep learning networks automatically and directly perform feature extraction, selection, and classification. Yoon et al. designed deep neural networks that combine both radiomic and genomic features to predict pathological stage and molecular receptor status of invasive breast cancer patients. They found that the results suggest superior performance on CNNs leveraging radiogenomics in comparison with CNNs trained on single modality data sources [49].

6 Remarkable AI-CAD examples

Many successful cases using AI technology, especially deep learning, have already been reported in the field of medical imaging diagnosis. In the following sections, several interesting examples in which AI-CAD shows equal or even better performance than human doctors will be introduced.

6.1 Fundus photographs

The mass screening of fundus photographs can be used to diagnose not only eye diseases such as glaucoma but also diabetic retinopathy and hypertensive retinopathy, both for hypertension and arteriosclerosis degree determination.

Google collected about 130,000 fundus images and analyzed them using deep learning [50]. As a result, Google published an interesting paper in 2016 that had a detection sensitivity of about 98%, which is comparable to that of an ophthalmologist; this paper received great attention when it was published.

Following this, in March 2018, Google published an even more interesting scientific paper [51], in which they collected fundus photographs together with other medical care data from about 300,000 patients and predicted heart disease from the fundus photographs using deep learning processing. The factors considered in this study were age, gender, smoking habits, blood pressure levels, the likelihood of diabetes, and body mass index (BMI) indicating obesity. It can be said that this has already exceeded human work. A technology called heat map was used to indicate the part of the fundus image that deep learning focuses on to make a decision.

6.2 Dermatology images

AI for skin cancer diagnosis has also reached the same level as or higher than that accomplished by doctors.

In January 2017, a research group at Stanford University used AI to diagnose skin cancer [52]. In their study, images of approximately 130,000 skin lesions were collected from the internet, and “skin cancer (melanoma),” “benign tumor,” etc. were learned by deep learning, and as a result, AI was able to diagnose skin cancer with accuracy equivalent to a dermatologist.

In addition, another researcher has recently evaluated AI and dermatologists to differentiate melanoma from over 100,000 developed images; the accuracy was that dermatologist vs. AI = 87% vs. 95%. It may be noted that the main title of this paper is “Human vs. Machine” [53].

6.3 Pathological images

Even in the field of pathological imaging, AI’s abilities can be overwhelming against doctors. According to a paper published at the end of 2017 [54], in a study of pathological images on detection of lymph node metastasis of breast cancer, the authors made a comparison of judgment results concerning the presence or absence of lymph node metastasis for deep learning AI versus 11 pathologists with an average experience of 16 years. As a result, the average value of the AUC (the area under the ROC curve) of AI was 0.994, and the average value of 11 doctors was 0.810, with a reading time limit. Also, even if the time was unlimited for the pathologist to review the image, it is still below the AUC of AI (AUC = 0.966).

6.4 Breast imaging

As described above, some issues exist in traditional CAD systems. To determine whether an AI-based CAD system like cmAssist (prototype, CureMetrix, La Jolla, CA) can be used to reduce false-positive (FP) markings on mammograms as compared to an FDA-approved traditional CAD-like ImageChecker CAD (version 10.0, Hologic, Inc., Sunnyvale, CA), Mayo et al. investigated employing the same dataset retrospectively [55]. They found that there is an overall 69% reduction in FPs using AI-CAD and mentioned that this decrease could result in a 17% decrease in radiologist reading time per case. Almost half of the cases showed no AI-CAD markings, while only 17% show no conventional CAD marks. Aboutalib et al. investigated the deep learning method to distinguish recalled but benign mammography images from negative exams and those with malignancy and showed its possibility to help reduce false recalls [56].

Wu et al. [57] investigated a CNN-based model for breast cancer screening exam classification using over 1,000,000 images obtained from over 200,000 exams and found that their model is as accurate as experienced radiologists, and the hybrid model is more accurate. Their best models are publicly available at GITHUB [57].

6.5 Chest imaging

Hwang et al. developed a deep learning-based algorithm that can classify normal and abnormal major thoracic diseases using nearly 90,000 chest radiographs and evaluated them for external validation using over 1000 chest radiographs [58]. They showed that the algorithms consistently outperformed physicians, including thoracic radiologists. Google researchers reported that an AI model using deep three-dimensional learning could assess cancer risk on low-dose CT lung cancer screening studies as well as, or even better than, experienced radiologists [59]

6.6 The door to automatic diagnosis opens

A recent topic of particular interest is the IDX-DR, a system for detecting diabetes from fundus photographs, developed by IDx Technologies, Inc., a startup company of ophthalmologist Abràmoff et al. in the U.S., which obtained FDA approval in April 2018. This system now performs beyond the level of existing CADs and is called a system for autonomous diagnosis or an AI doctor. It analyzes a fundus photograph taken with a non-mydriatic eye fundus camera, TRC-NW400, manufactured by Topcon Healthcare Solutions, Inc., and when the picture is input to the system, it outputs as “Diabetes detection: Recommend a doctor’s visit,” “Not detected: Recommend re-examination within 12 months,” or “Indicate a defect in image quality: Re-shooting.” The detection sensitivity of the clinical test at performance evaluation is 87%, and the specificity at this time is 90% [60]. This device is a new-genre AI medical device that can be used by primary care doctors (health care providers) without a specialist’s diagnosis of photographs or analysis results.

6.7 Issues to be considered

As mentioned above, there are many cases and much evidence indicating that deep learning-based AI or AI-CAD has already surpassed professional medical doctors in terms of the ability to interpret medical images. However, can this be possibly realized in clinical circumstances? To reach such a conclusion, some issues must be considered:

1) Data issues in terms of content, quantity, and quality.
   Most of the applications were examined for one possible facility; hence, if it is applied to another facility, we have no evidence that it will work in the same way. New technologies, such as GAN and unsupervised learning methods, may solve these kinds of issues. Image labeling and annotation for large datasets are also remaining issues.
2) Whether the data shown in the paper shows differences with statistical significance.
3) No evidence in large-scale clinical practice.
4) Black box issues of deep learning in medicine.
   Explainable AI (XAI), including the heat map-type method, is now under investigation by AI-related researchers, including knowledge engineers.

Although algorithms with a performance that exceeds the level of physicians have already appeared, we need to validate these academic reports carefully. Liu et al. reviewed over 30,000 articles originally but found that only a few of these presented algorithms that were sufficiently robust in their design to succeed [61]. Although they concluded in their systematic review and meta-analysis that the diagnostic performance of some of the deep learning models was equivalent to that of healthcare professionals, they caution that the diagnostic capabilities of AI technology remain uncertain due to a lack of research that directly compares AI and physician performance or validates AI performance in a real clinical environment.

In a recent paper in radiology [62], a roadmap for basic research on AI in medical imaging is summarized and reported from the 2018 NIH/RSNA/ACR/The Academy Workshop, and it might be helpful to consider the important tasks for future research initiatives.

7 Conclusion

In this review article, the topic was limited to diagnostic imaging and the application of AI in this area. The support of image interpretation was mainly described; however, AI is also making inroads into areas such as patient positioning, image reconstruction, image formation (image quality improvement and radiation dose reduction), image processing, automatic dose estimation, maintenance of imaging equipment, and even development of an automated radiology reporting and analytics tool [63].

We may keep in mind that "radiologists who use AI will replace radiologists who do not," as mentioned by Curtis Langlotz of Stanford University (RSNA2017) [64]. Finally, we should note the statement of Nan Wu, of the NTU Center for Data Science, "the transition to AI support in diagnostic radiology should proceed like the adoption of self-driving cars—slowly and carefully, building trust, and improving systems along the way with a focus on safety" [65].

Funding This research was supported in part by a Grant-in-Aid for Scientific Research (B) (No. 19H03599) and (C) (No. 19K10347), by the Japan Society for the Promotion of Science, and also in part by a Collaborative Research Grant from EyeTech Co., Ltd., Japan.

Compliance with ethical standards

Conflict of interest The author declares that he has no conflict of interest.

References

1. von Ginneken B. Fifty years of computer analysis in chest imaging: rule-based, machine learning, deep learning. Radiol Phys Technol. 2017;10(1):23–322.
2. Suzuki K. Overview of deep learning in medical imaging. Radiol Phys Technol. 2017;10(3):257–73.
3. Shen D, Wu G, Suk H. Deep learning in medical image analysis. Annu Rev Biomed Eng. 2017;19:221–48.
4. Ker J, Wang L, Rao J, Lim T. Deep learning applications in medical image analysis. IEEE Access. 2018;6:9375–89.
5. Sahiner B, Pezeshk A, Hadjiiski LM, Wang X, Drukker K, Cha KH, Summers RM, Giger ML. Deep learning in medical imaging and radiation therapy. Med Phys. 2019;46(1):e1–e36.
6. Soffer S, Ben-Cohen A, Shimon O, Amitai MM, Greenspan H, Klang E. Convolutional neural networks for radiologic images: a radiologist's guide. Radiology. 2019;290(3):590–606.
7. Kaji S, Kida S. Overview of image-to-image translation by use of deep neural networks: denoising, super-resolution, modality conversion, and reconstruction in medical imaging. Radiol Phys Technol. 2019;12(3):235–48.
8. Hwang JJ, Jung YH, Cho BH, Heo MS. An overview of deep learning in the field of dentistry. Imaging Sci Dent. 2019;49(1):1–7.
9. Lee G, Fujita H. Deep learning in medical image analysis: challenges and applications. Cham: Springer; 2020 (**in press**).
10. Doi K. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. Comput Med Imaging Graph. 2007;31(4–5):198–21111.
11. Gao Y, Geras KJ, Lewin AA, Moy L. New frontiers: an update on computer-aided diagnosis for breast imaging in the age of artificial intelligence. AJR Am J Roentgenol. 2019;212:300–7.
12. Freer TW, Ulissey MJ. Screening mammography with computer aided detection: prospective study of 12,860 patients in a community breast center. Radiology. 2001;220(3):781–6.
13. Fenton JJ, Taplin SH, Carney PA, Abraham L, Sickles EA, D'Orsi C, Berns EA, Cutter G, Hendrick RE, Barlow WE, Elmore JG.

Influence of computer-aided detection on performance of screening mammography. N Engl J Med. 2007;356(14):1399–409.
14. Lehman CD, Wellman RD, Buist DS, Kerlikowske K, Tosteson AN, Miglioretti DL. Breast cancer surveillance consortium. Diagnostic accuracy of digital screening mammography with and without computer-aided detection. JAMA Intern Med. 2015;175(11):1828–37.
15. Kobli A, Jha S. Why CAD failed in mammography. J Am Coll Radiol. 2018;15(3):535–7.
16. Zhou X, Takayama R, Wang S, Hara T, Fujita H. Deep learning of the sectional appearances of 3D CT images for anatomical structure segmentation based on an FCN voting method. Med Phys. 2017;44(10):5221–333.
17. Faes L, Wagner SK, Fu DJ, Liu X, Korot E, Ledsam JR, Back T, Chopra R, Pontikos N, Kern C, Moraes G, Schmid MK, Sim D, Balaskas K, Bachmann LM, Denniston AK, Keane PA. Automated deep learning design for medical image classification by health-care professionals with no coding experience: a feasibility study. Lancet Digit Health. 2019;1(5):e232–e242242.
18. Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. ChestX-ray8: hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: 2017 IEEE conference on computer vision and pattern recognition (CVPR). Honolulu: IEEE; 2017. p. 2097–106. https://doi.org/10.1109/CVPR.2017.369, http://openaccess.thecvf.com/content_cvpr_2017/papers/Wang_ChestX-ray8_Hospital-Scale_Chest_CVPR_2017_paper.pdf.
19. https://nihcc.app.box.com/v/DeepLesion. Accessed 28 Dec 2019.
20. https://stanfordmlgroup.github.io/competitions/chexpert/. Accessed 28 Dec 2019.
21. Dunnmon J, Yi D, Langlots CP, Ré C, Rubin DL, Lungren MP. Assessment of convolutional neural networks for automated classification of chest radiographs. Radiology. 2019;290(2):537–44.
22. Abdelhafiz D, Yang C, Ammar R, Nabavi S. Deep convolutional neural networks for mammography: advances, challenges and applications. BMC Bioinform. 2019;20(Suppl 11):281.
23. https://www.cancerimagingarchive.net/. Accessed 28 Dec 2019.
24. Huynh BQ, Li H, Giger ML. Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. J Med Imaging. 2016;3(3):034501.
25. Samala RK, Chan HP, Hadjiiski L, Helvie MA, Wei J, Cha K. Mass detection in digital breast tomosynthesis: deep convolutional neural network with transfer learning from mammography. Med Phys. 2016;43:6654–66.
26. Kim M, Lee H, Song K, Sehyo Y, Ramaraj P, Lee C, Baik J, Do S. GrayNet: a versatile base model for practical deep learning CT applications. In: Proc. of Conference on Machine Intelligence in Medical Imaging (C-MIMI), 2-page extended abstract is available from C-MIMI 2019 website, Austin, TX, Sep. 22–23, 2019.
27. https://github.com/MGH-LMIC/graynet_keras. Accessed 28 Dec 2019.
28. https://devblogs.nvidia.com/annotation-transfer-learning-clara-train/. Accessed 28 Dec 2019.
29. Yi X, Walia E, Babyn P. Generative adversarial network in medical imaging: a review. Med Image Anal. 2019;58:101552.
30. Frid-Adar M, Diamant I, Klang E, Amitai M, Goldberger J, Greenspan H. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. Neurocomputing. 2018;321:321–31.
31. Onishi Y, Teramoto A, Tsujimoto M, Tsukamoto T, Saito K, Toyama H, Imaizumi K, Fujita H. Automated pulmonary nodule classification in computed tomography images using a deep convolutional neural network trained by generative adversarial networks. Biomed Res Int. 2019;9:6051939.
32. Onishi Y, Teramoto A, Tsujimoto M, Tsukamoto T, Saito K, Toyama H, Imaizumi K, Fujita H. Multiplanar analysis for pulmonary nodule classification in CT images using deep convolutional neural network and generative adversarial networks. Int J Comput Assist Radiol Surg. 2019. https://doi.org/10.1007/s11548-019-02092-z.
33. https://ai.googleblog.com/2017/04/federated-learning-collaborative.html. Accessed 28 Dec 2019.
34. Chang K, Balachandar N, Lam C, Yi D, Brown J, Beers A, Rosen B, Rubin DL, Kalpathy-Cramer J. Distributed deep learning networks among institutions for medical imaging. J Am Med Inform Assoc. 2018;25(8):945–54.
35. https://www.intel.ai/federated-learning-for-medical-imaging/#gs.9988iz. Accessed 28 Dec 2019.
36. Sheller MJ, Reina GA, Edwards B, Martin J, Bakas S. Multi-institutional deep learning modeling without sharing patient data: a feasibility study on brain tumor segmentation, Lecture notes in computer science book series (Volume 11383). Brainlesion. 2019;11383:92–104.
37. Beyer F, Zierott L, Fallenberg EM, Juergens KU, Stoeckel J, Heindel W, Wormanns D. Comparison of sensitivity and reading time for the use of computer-aided detection (CAD) of pulmonary nodules at MDCT as concurrent or second reader. Eur Radiol. 2007;17(11):2941–7.
38. Samulski M, Hupse R, Boetes C, Mus RDM, den Heeten GJ, Karssemeijer N. Using computer-aided detection in mammography as a decision support. Eur Radiol. 2010;20(10):2323–30.
39. Rodríguez-Ruiz A, Krupinski E, Mordang J-J, Schilling K, Heywang-Köbrunner SH, Sechopoulos I, Ritse Mann RM. Detection of breast cancer with mammography: effect of an artificial intelligence support system. Radiology. 2019;290(2):305–14.
40. Kyono T, Gilbert FJ, van der Schaar M. Improving workflow efficiency for mammography using machine learning. J Am Coll Radiol. 2019. https://doi.org/10.1016/j.jacr.2019.05.012.
41. PSEHB/MDED (Yakuseikisinn). Notification 0523 No. 2, May 23, 2019, Japan. Appendix 4. Guidance for evaluation of artificial intelligence-assisted medical imaging systems for clinical diagnosis. 2019. http://dmd.nihs.go.jp/jisedai/tsuuchi/index.html. Accessed 31 Dec 2019.
42. US Food and Drug Administration. Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD)—discussion paper and request for feedback. 2019. https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device. Accessed 29 Dec 2019.
43. Goldenberg R, Peled N. Computer-aided simple triage. Int J Comput Assist Radiol Surg. 2011;6(5):705–11.
44. Muramatsu C. Overview of subjective similarity of images for content-based medical image retrieval. Radiol Phys Technol. 2018;11(2):109–24.
45. Owais M, Arsalan M, Choi J, Park KR. Effective diagnosis and treatment through content-based medical image retrieval (CBMIR) by using artificial intelligence. J Clin Med. 2019;8(4):462.
46. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology. 2016;278(2):563–77.
47. Bodalal Z, Trebeschi S, Nguyen-Kim TDL, Schats W, Beets-Tan R. Radiogenomics: bridging imaging and genomics. Abdom Radiol. 2019;44(6):1960–84.
48. Kai C, Uchiyama Y, Shiraishi J, Fujita H, Doi K. Computer-aided diagnosis with radiogenomics: analysis of the relationship between genotype and morphological changes of the brain magnetic resonance images. Radiol Phys Technol. 2018;11(3):265–73.
49. Yoon H-J, Ramanathan A, Alamudun F, Tourassi G. Deep radiogenomics for predicting clinical phenotypes in invasive breast cancer. In: Proc. SPIE 10718, 14th International Workshop on Breast Imaging (IWBI 2018), 2018. p. 107181H.

50. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, Venugopalan S, Widner K, Madams T, Cuadros J, Kim R, Raman R, Nelson PC, Mega JL, Webster DR. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA. 2016;316(22):2402–10.
51. Poplin R, Varadarajan AV, Blumer K, Liu Y, McConnell MV, Corrado GS, Peng L, Webster DR. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. Nat Biomed Eng. 2018;2(3):158–64.
52. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542:115–8.
53. Haenssle HA, Fink C, Schneiderbauer R, Toberer F, Buhl T, Blum A, Kalloo A, Hassen ABH, Thomas L, Enk A, Uhlmann L. Reader study level-I and level-II Groups. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Ann Oncol. 2018;29(8):1836–42.
54. Bejnordi BE, Veta M, van Diest PJ, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA. 2017;318(22):2199–210.
55. Mayo RC, Kent D, Sen LC, Kapoor M, Leung JWT, Watanabe AT. Reduction of false-positive markings on mammograms: a retrospective comparison study using artificial intelligence-base CAD. J Dig Imag. 2019;32(4):618–24.
56. Aboutalib SS, Mohamed AA, Berg WA, Zuley ML, Sumkin JH, Wu S. Deep learning to distinguish recalled but benign mammography images in breast cancer screening. Clin Cancer Res. 2018;24(23):5902–9.
57. Wu N, Phang J, Park J, Shen Y, Huang Z, Zorin M, Jastrzebski S, Fevry T, Katsnelson J, Kim E, Wolfson S, Parikh U, Gaddam S, Lin LLY, Ho K, Weinstein JD, Reig B, Gao Y, Pysarenko HTK, Lewin A, Lee J, Airola K, Mema E, Chung S, Hwang E, Samreen N, Kim SG,Heacock L, Moy L, Cho K, Geras KJ. Deep neural networks improve radiologists' performance in breast cancer screening. IEEE Trans Med Imaging. 2019. https://doi.org/10.1109/TMI.2019.2945514.
58. Hwang EJ, Park S, Jin KN. Development and validation of a deep learning–based automated detection algorithm for major thoracic diseases on chest radiographs. JAMA Netw Open. 2019;2(3):e191095.
59. Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, Tse D, Etemadi M, Ye W, Corrado G, Naidich DP, Shetty S. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nat Med. 2019;25:954–61.
60. Abràmoff MD, Lavin PT, Birch M, Shah NA, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. Npj Digit Med. 2018;1:39.
61. Liu X, Faes L, Kale AU, Wagner SK, Fu DJ, Bruynseels A, Mahendiran T, Moraes G, Shamdas M, Kern C, Ledsam JR, Schmid MK, Balaskas K, Topol EJ, Bachmann LM, Keane PA, Denniston AK. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. Lancet Digit Health. 2019;1(6):e271–e297297.
62. Langlotz C, Allen B, Erickson BJ, Kalpathy-Cramer J, Bigelow K, Cook TS, Flanders AE, Lungren MP, Mendelson DS, Rudie JD, Wang G, Kandarpa K. A roadmap for foundational research on artificial intelligence in medical imaging: from the 2018 NIH/RSNA/ACR/the academy workshop. Radiology. 2019;291(3):781–91.
63. Choy G, Khalilzadeh O, Michalski M, Do S, Samir AE, Pianykh OS, Geis JR, Pandharipandle PV, Brink JA, Dreyer KJ. Current applications and future impact of machine learning in radiology. Radiology. 2018;288(2):318–28.
64. https://aimi.stanford.edu/news/rsna-2017-rads-who-use-ai-will-replace-rads-who-don-t. Accessed 28 Dec 2019.
65. Walter M. Radiologists, AI an accurate combination for detecting breast cancer. AI in Healthcare. 2019. https://www.aiin.healthcare/topics/medical-imaging/radiologists-ai-accurate-breast-cancer-mammography. Accessed 29 Dec 2019.