REGULAR PAPER

A holistic overview of deep learning approach in medical imaging

Rammah Yousef[1] · Gaurav Gupta[1] · Nabhan Yousef[2] · Manju Khari[3]

Abstract

Medical images are a rich source of invaluable necessary information used by clinicians. Recent technologies have introduced many advancements for exploiting the most of this information and use it to generate better analysis. Deep learning (DL) techniques have been empowered in medical images analysis using computer-assisted imaging contexts and presenting a lot of solutions and improvements while analyzing these images by radiologists and other specialists. In this paper, we present a survey of DL techniques used for variety of tasks along with the different medical image's modalities to provide critical review of the recent developments in this direction. We have organized our paper to provide significant contribution of deep leaning traits and learn its concepts, which is in turn helpful for non-expert in medical society. Then, we present several applications of deep learning (e.g., segmentation, classification, detection, etc.) which are commonly used for clinical purposes for different anatomical site, and we also present the main key terms for DL attributes like basic architecture, data augmentation, transfer learning, and feature selection methods. Medical images as inputs to deep learning architectures will be the mainstream in the coming years, and novel DL techniques are predicted to be the core of medical images analysis. We conclude our paper by addressing some research challenges and the suggested solutions for them found in literature, and also future promises and directions for further developments.

Keywords Medical imaging · Deep learning (DL) · Medical data augmentation · Transfer learning

1 Introduction

Health no doubt is on the top of concerns hierarchy in our life. Through the lifetime, human has struggled of diseases which cause death; in our life scope, we are fighting against enormous number of diseases, moreover, improving life expectancy and health status significantly. Historically medicine could not find the cure of numerous diseases due to a lot of reasons starting from clinical equipment and sensors to the analytical tools of the collected medical data. The fields of big data, AI, and cloud computing have played a missive role at each aspect of handling these data. Across

the worldwide, Artificial Intelligence (AI) has been widely common and well known enough to most of the people due to the rapid progress achieved in almost every domain in our life. The importance of AI comes from the remarkable progress within the last 2 decades only, and it is still growing and specialists from different fields are investing. AI's algorithms were attributed to the availability of big data and the efficiency of modern computing criteria that is provided lately.

This paper aims to give a holistic overview in the field of healthcare as an application of AI and deep learning particularly. The paper starts by giving an overview of medical imaging as an application of deep learning and then moving to why do we need AI in healthcare; in this section, we will give the key terms of how AI is used in both the main medical data types which are medical imaging and medical signals. To provide a moderate and rich general perspective, we will mention the well-known data which are widely used for generalization and the main pathologies, as well. Starting from classification and detection of a disease to segmentation and treatment and finally survival rate and prognostics. We will talk in detail about each pathology with the relevant

key features and the significant results found in literature. In the last section, we will discuss about the challenges of deep learning and the future scope of AI in healthcare. Generally, AI is being a fundamental path in nowadays medicine which is in short a software that can learn from data like human being and it can develop an experience systematically and finally deliver a solution or diagnostic even faster than humans. AI has become an assistive tool in medicine with benefits like error reduction, improving accuracy, fast computing, and better diagnosis were introduced to help doctors efficiently. From clinical perspective, AI is used now to help the doctors in decision-making due to faster pattern recognition from the medical data which also in turn are registered more precisely in computers than humans; moreover, AI has the ability to manage and monitor the patients' data and creating a personalized medical plan for future treatments. Ultimately, AI has proved to be helpful in medical field with different levels, such as telemedicine diagnosis diseases, decision-making assistant, and drug discovery and development. Machine learning (ML) and deep learning (DL) have tremendous usages in healthcare such as clinical decision support (CDS) system which incorporate human's knowledge or large datasets to provide clinical

recommendations. Another application is to analyze large historical data and get the insights which can predict the future cases of a patient using pattern identification. In this paper, we will highlight the top deep learning advancement and applications in medical imaging. Figure 1 shows the workflow chart of paper highlights.

2 Background concepts

2.1 Medical imaging

Deep learning in medical imaging [1] is the contemporary scope of AI which has the top breakthroughs in numerous scientific domains including computer vision [2], Natural Language Processing (NLP) [3] and chemical structure analysis, where deep learning is specialized with highly complicated processes. Lately due to deep learning robustness while dealing with images, it has attracted big interest in medical imaging, and it holds big promising future for this field. The main idea that DL is preferable is that medical data are large and it has different varieties such as medical images, medical signals, and medical logs' data of patients
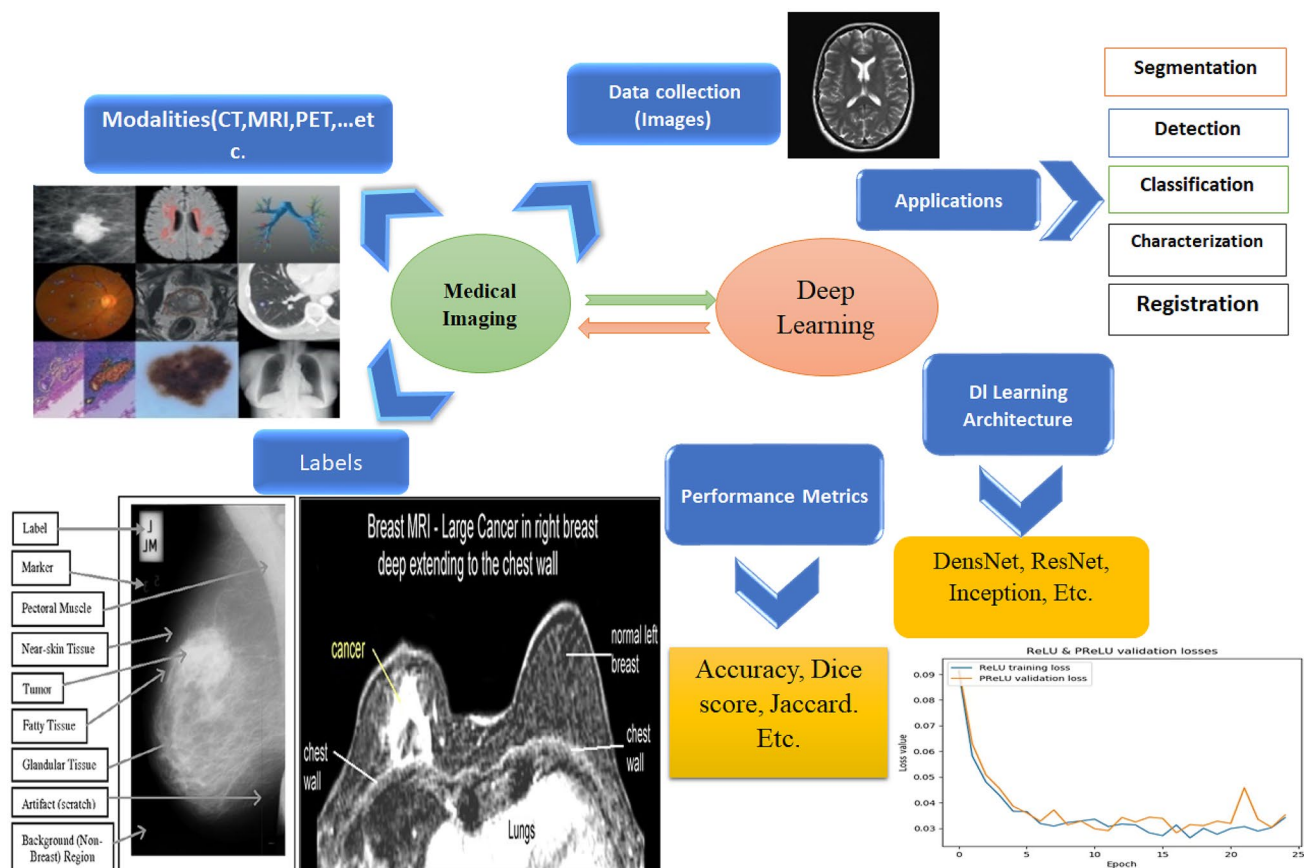

Fig. 1 Deep learning implementation and traits for medical imaging application

monitoring of body sensed information. Analyzing these data especially historical data by learning very complex mathematical models and extracting meaningful information is the key feature where DL scheme outperformed humans. In other words, DL framework will not replace the doctors, but it will assist them in decision-making and it will enhance the accuracy of the final diagnosis analysis. Our workflow procedure is shown in Fig. 1.

2.1.1 Types of medical imaging

There are plenty of medical image types, and selecting the type depends on the usage, in a study which was held in US [4], it was found that there are some basic and widely used modalities of these medical images which also have increased, and these modalities are Magnetic Resonance Images (MRI), Computed Tomography (CT) scans, and Positron Emission Tomography (PET) to be on the top and some other common modalities like, X-ray, Ultrasound, and histology slides. Medical images are known to be so complicated, and in some cases, acquisition of these images is considered to be long process and it needs specific technical implications, e.g., an MRI which may need over 100 Mega Byte of memory storage.

Because of a lack of standardization while image acquisition and diversity in the scanning devices' settings, a phenomenon called "distribution drift" might arise and cause non-standard acquisition. From a clinical need perspective, medical images are the key part of diagnosis of a disease and then the treatment too. In traditional diagnosis, a radiologist reviews the image, and then, he provides the doctors with a report of his findings. Images are an important part of the invasive process to be used in further treatment, e.g., surgical operations or radiology therapies for example [5, 6].

2.2 DL frameworks

Conceptually, Artificial Neural Networks (ANN) are a mimic of the human neuro system in the structure and work. Medical imaging [7] is a field by which is specialized in observing and analyzing the physical status of the human body by generating visual representations like images of internal tissues or some organs of the body through either invasive or non-invasive procedure.

2.2.1 Key technologies and deep learning

Historically, AI scheme has been proposed in 1970s and it has mainly the two major subcategories, such as Machine Learning (ML) and Deep Learning (DL). The earlier AI used heuristics-based techniques for extracting features from data, and further developments started using handcrafted features' extraction and finally to supervised learning. Where basically Convolutional Neural Networks (CNN) [8] is used in images and specifically in medical images. CNN is known to be hungry for data, so it is the most suitable methodology for images, and the recent developments in hardware specifications and GPUs have helped a lot in performing CNN algorithms for medical image analysis. The generalized formulation of how CNN work was proposed by Lecun et al. [9], where they have used the error backpropagation for the first example of digits hand written recognition. Ultimately, CNNs have been the predominant architecture among all other algorithms which belong to AI, and the number of research of CNN has increased especially in medical images analysis and many new modalities have been proposed. In this section, we explain the fundamentals of DL and its algorithmic path in medical imaging. The commonly known categories of deep learning and their subcategories are discussed in this section and are shown in Fig. 2.

Fig. 2 DL basic categories as per paper organization
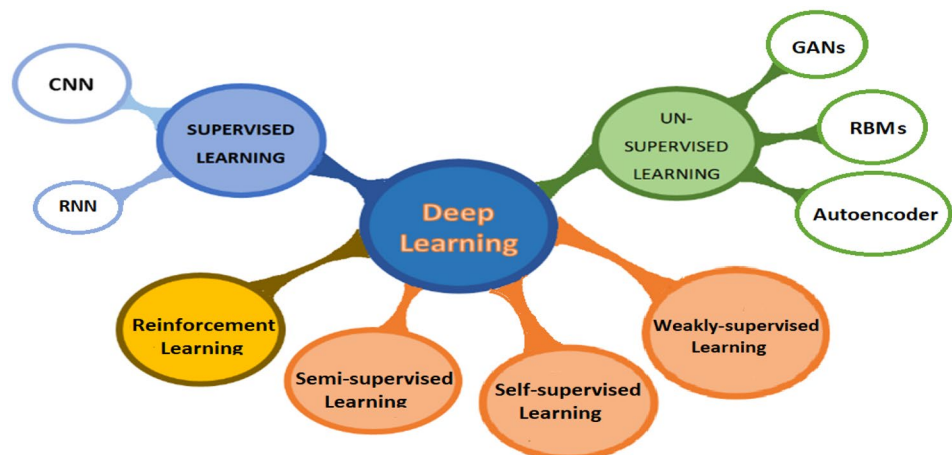
2.2.2 Supervised learning

Convolutional neural networks: CNN [10] have taken the major role in many aspects and have lead the work in image-based tasks, including image reconstruction, enhancement, classification, segmentation, registration, and localization. CNNs are considered to be the most deep learning algorithm regarding images and visual processing because of its robustness in image dimensionality reduction without losing image's important features; in this way, CNN algorithm deals with less parameters which mean increasing the computational efficiency. Another key term about CNN is that this architecture is suitable for hospitals use, because it can handle both 2D and 3D images, because some of medical images modalities like X-ray images are 2D-based images, while MRI and CT scan images are 3-dimensional images. In this section, we will explain the framework of CNN architecture as the heart of deep learning in medical imaging.

Convolutional layer: Before deep learning and CNN, in image processing, convolution terminology was used for extracting specific features from an image, such as corners, edges (e.g., sobel filter), and noise by applying a particular filters or kernels on the image. This operation is done by sliding the filter all over the image in a sliding window form until all the image is covered. In CNN, usually, the startup layers are designed to extract low-level features, such as lines and edges, and the progressive layers are built up for extracting higher features like full objects within an image. The goodness of using modern CNNs is that the filters could be 2D or 3D filters using multiple filters to form a volume and this depends on the application. The main discrimination in CNN is that this architecture obliges the elements in a filter to be the network weights. The idea behind CNN architecture is the convolution operation which is denoted by the symbol *. Equation (1) represents the convolution operation

$$s(t) = I(t) * K(a) = (I * K)(t), \tag{1}$$

where $s(t)$ is the output feature map and $I(t)$ is the original image to be convolved with the filter $K(a)$.

Activation function: Activation functions are the enable button of a neuron; in CNN, there are multiple popular activation functions which are widely used such as, sigmoid, tanh, ReLU, Leaky ReLU, and Randomized ReLU. Especially, in medical imaging, most papers found in literature uses ReLU activation function which is defined using the formula

$$f(x) = \max(0, x), \tag{2}$$

where x represents the input of a neuron.

There are other used activation functions used in CNN, such as sigmoid, tanh, and leaky-ReLu

$$\text{Sigmoid} = \frac{1}{1 + e^{-x}}. \tag{3}$$

Pooling layer: Mainly, this layer is used to reduce the parameters needed to be computed and it reduces the size of the image but not the number of channels. There are few pooling layers, such as Max-pooling, average- pooling, and L2-normalization pooling, where Max-pooling is the widely used pooling layer. Max-pooling means taking the maximum value of a position of the feature map after convolution operation.

Fully connecter layer: This layer is the same layer that is used in a casual ANN where usually in such network each neuron is connected to all other neurons in both the previous and next layer's neurons; this makes the computation very expensive. A CNN model can get the help of the stochastic gradient descent to learn significant associations from the existing examples used for training. Thus, the benefit of a CNN usage is that it gradually reduces the feature map size before finally is get flatten to feed the fully connected layer which in turn computes the probability scores of the targeted classes for the classification. Fc-connected layer is the last layer in a CNN model, Furthermore, this layer processes the strongly extracted features from an image due to the convolutional a pooling layer before and finally fc-layer indicate to which class is an image belong to.

Recurrent neural networks: RNN is a major part from supervised deep learning models, and this model is specific with analyzing sequential data and time series. We can imagine an RNN as a casual neural network, while each layer of it represents the observations at a particular time (t). In [11], RNN was used for text generating which further connected to speech recognition and text prediction and other applications too. RNN are recurrent, because same work is done for every element in a sequence and the output depends on the previous output computation of the previous element in that sequence general, the output of a layer is fed as an input to the new input of the same layer as it is shown in Fig. 3. Moreover, since the backpropagation of the output will suffer of vanishing gradient with time, so commonly a network is evolved which is Long Short-Term Memory (LSTM).

In network and three bidirectional gated recurrent units is (BGRU) to help the RNN to hold long-term dependencies.

There were few papers found in the literature of RNN in medical imaging and particularly in segmentation, in [12], Chen et al. have used RNN along with CNN for segmenting fungal and neuronal structures from 3D images. Another application of RNN is in image caption generation [13], where these models can be used for annotating medical images like X-ray with text captions extracted and trained from radiologists' reports [14]. RuoxuanCui et al. [15] have used a combination of CNN and RNN for diagnosing
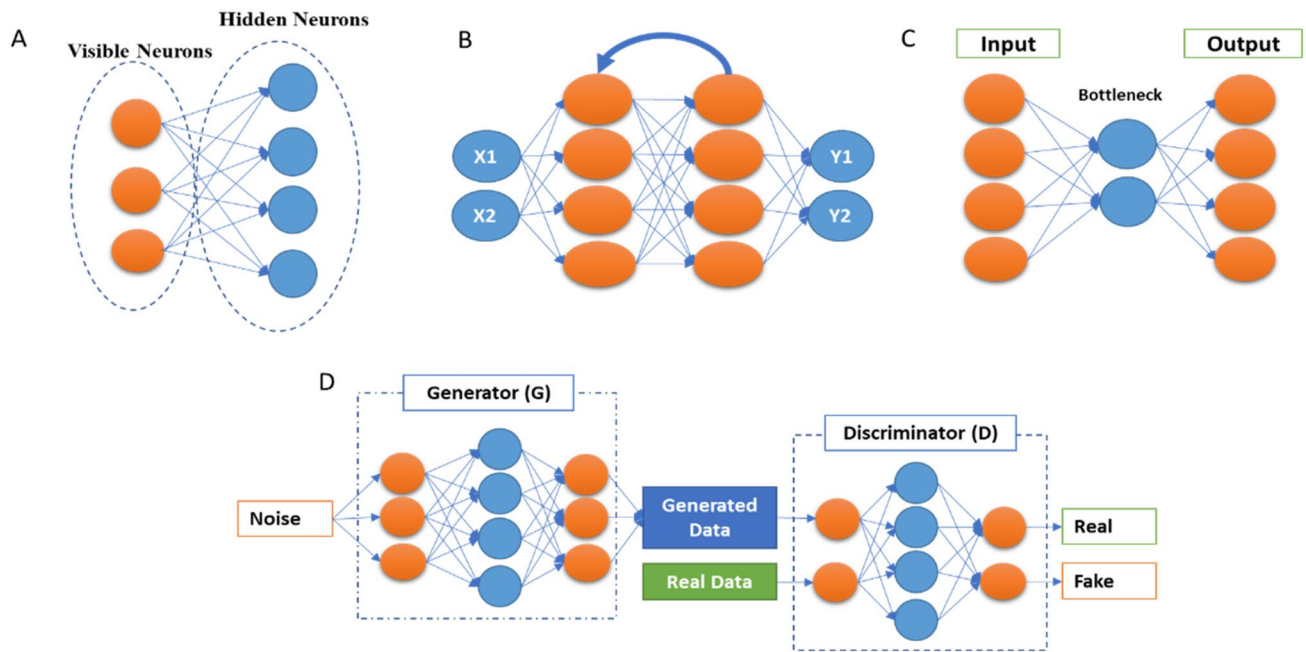
Fig. 3 Basic common deep learning architectures. **A** Restricted Boltzmann machine. **B** Recurrent Neural Network (RNN). **C** Autoencoders. **D** GANs

Alzheimer disease where their CNN model was used for classification task, after that the CNN model's output is fed to an RNN model with cascaded bidirectional gated recurrent units (BGRU) layers to extract the longitudinal features of the disease. In summary, RNN is commonly used with a CNN model in medical imaging. In [16], authors have developed a novel RNN for speeding up an iterative MAP estimation algorithm.

2.2.3 Unsupervised deep learning

Beside the CNN as a supervised machine leaning algorithm in medical imaging, there are a few unsupervised learning algorithms for this purpose as well, such as Deep Belief Networks (DBNs), Autoencoders, and Generative Adversarial Networks (GANs), where the last has been used for not only performing the image-based tasks but as a data synthesis and augmentation too. Unsupervised learning models have been used for different medical imaging applications, such as motion tracking [17] general modeling, classification improvement [18], artifact reduction [19], and medical image registration [20]. In this section, we will list the mostly used unsupervised learning structures.

2.2.3.1 Autoencoders Autoencoders [21, 22] are an unsupervised deep learning algorithm by which this model refers to the important features of an input data and dismisses the other data. These important representations of features are called 'codings' where it is commonly called represen-

tation learning. The basic architecture is shown in Fig. 3. The robustness of autoencoders stems from the ability to reconstruct output data, which is similar to the input data, because it has cost function which applies penalties to the model when the output and input data are different. Moreover, autoencoders are considered as an automatic features detector, because they do not need labeled data to learn from due to the unsupervised manner. Autoencoders architecture is similar to a formal CNN model, but with the feature is that the number of input neurons must be equal to the number in the output layer. Reducing dimensionality of the raw input data is one of the features of autoencoders, and in some cases, autoencoders are used for denoising purpose [23], where this autoencoders are called denoising autoencoders. In general, there are few kinds of autoencoders used for different purposes, we mention here the common autoencoders, for example, *Sparse Autoencoders* [24] where the neurons in the hidden layer are deactivated through a threshold which means limiting the activated neurons to get a representation in the output similar to the input where for extracting most of the features from the input, most of the hidden layer neurons should be set to zero. *Variational autoencoders* (VAEs) [25] are generative model with two networks (Encoder and Decoder) where the encoder network projects the input into latent representation using Gaussian distribution approximation, and the decoder network maps the latent representations into the output data. *Contractive autoencoders* [26] and adversarial autoencoders are mostly similar to a Generative Adversarial Network (GAN).

2.2.3.2 Generative Adversarial Networks GANs [27][28] were first introduced by Ian Goodfellow in 2014; it consists basically on a combination of two CNN networks: the first one is called Generative model and another is the discriminator model. For better understanding how GANs work, scientists describe the two networks as a two players who competing against each other, where the generator network tries to fool the discriminator network by generating near authentic data (e.g., artificial images), while the discriminator network tries to distinguish between the generator output and the real data, Fig. 3. The name of the network is inspired from the objective of the generator to overcome the discriminator. After the training process, both the generator and discriminator networks get better, where the first generates more real data, and the second learns how to differentiate between both previously mentioned data better until the end-point of the whole process where the discriminator network is unable to distinguish between real and artificial data (images). In fact, the criteria by which both networks learn from each other are using the backpropagation for the both, Markov chains, and dropout too. Recently, we have seen tremendous usage of GANs for different applications in medical imaging such as, synthetic images for generating new images and enhance the deep learning models efficiency by increasing the number of training images in the dataset [29], classification [30, 31], detection [32], segmentation [33, 34], image-to-image translation [35], and other application too. In a study by Kazeminia et al. [36], they have listed all the applications of GANs in medical imaging and the most two used applications of this unsupervised models are image synthesis and segmentation.

2.2.3.3 Restricted Boltzmann machines *Axkley* et al. were the first to introduce the Boltzmann machines in 1985 [37], Fig. 3, also known as Gibbs distribution, and further *Smolensky* has modified it to be known as Restricted Boltzmann Machines (RBMs) [38]. RBMs consist on two layers of neural networks with stochastic, generative, and probabilistic capabilities, and they can learn probability distributions and internal representations from the dataset. RBMs work using the backpropagation path of input data for generating and estimating the probability distribution of the original input data using gradient descent loss. These unsupervised models are used mostly for dimensionality reduction, filtering, classification, and features representation learning. In medical imaging, Tulder et al. [39] have modified the RBMs and introduced a novel convolutional RBMs for lung tissue classification using CT scan images; they have extracted the features using different methodologies (generative, discriminative, or mixed) to construct the filters; after that, Random Forest (RF) classifier was used for the classification objective. Ultimately, a stacked version of RBMs is called Deep Belief Networks (DBNs) [40]. Each RBM model performs

non-linear transformation which will again be the input for the next RBM model; performing this process progressively gives the network a lot of flexibility while expansion.

DBNs are generative models, which allow them to be used as a supervised or unsupervised settings. The feature learning is done through an unsupervised manner by doing the layer-by-layer pre-training. For the classification task, a backpropagation (gradient descent) through the RBM stacks is done for fine-tuning on the labeled dataset. In medical imaging applications, DBNs were used widely; for example, Khatami et al. [41] used this model for classification of X-ray images of anatomic regions and orientations; in [42], AVN Reddy et al. have proposed a hybrid deep belief networks (DBN) for glioblastoma tumor classification from MRI images. Another significant application of DBNs was reported in [43] where they have used a novel DBNs' framework for medical images' fusion.

2.2.4 Self-supervised learning

Self-supervised learning is basically a subtype of unsupervised Learning, by which it learns features' representations using a proxy task where the data contain supervisory signals. After representation learning, it is fine-tuned using annotated data. The benefit of self-supervised learning is that it eliminates the need of humans to label the data, where this system extracts the visibly natural relevant context from the data and assign metadata with the representations as supervisory signals. This system matches with unsupervised learning, because both systems learn representations without using explicitly provided labels, but the difference is that self-supervised learning does not learn inherent structure of data and it is not centered around clustering, anomaly detection, dimensionality reduction, and density estimation. The genesis model of this system can retrieve the original image from a distorted image (e.g., non-linear gray-value transformation, image inpainting, image out-painting, and pixels shuffle) using proxy task [44]. Zhu et al. [45] have used self-supervised learning and its proxy task to solve Rubik's cube which mainly contain three operations (rotating, masking, and ordering) the robustness of this model comes from that the network is robust to noise and it learns features that are invariant to rotation and translation. Shekoofeh et al. [46] have exploited the effectiveness of self-supervised learning in pre-training strategy used to classify medical images for tow tasks (dermatology skin condition classification, and multi-label chest X-ray classification). Their study has improved the classification accuracy after using two self-supervised learning systems: the first one is trained on ImageNet dataset and the second one is trained on unlabeled domain specific medical images.

2.2.5 Semi-supervised learning

Semi-supervised learning is a system by which it stands in between supervised learning and unsupervised learning systems, because for example, it is used for classification task (supervised learning) but without having all the data labeled (Unsupervised learning). Thus, this system is trained on small, labeled dataset, and then generates pseudo-labels to get larger dataset with labels, and the final model is trained by mixing up both the original dataset and the generated one of images. Nie et al. [47] have proposed semi-supervised learning-based deep network for image segmentation, the proposed method trains adversarially a segmentation model, from the confidence map is computed, and the semi-supervised learning strategy is used to generate labeled data. Another application of semi-supervised learning is used for cardiac MRI segmentation, [48]. Liu et al. [49] have presented a novel relation-driven semi-supervised model to classify medical images, they have introduced a novel Sample Relation Consistency (SRC) paradigm to use unlabeled data by generalizing and modeling the relationship information between different samples; in their experiment, they have applied the novel method on two benchmark medical images for classification, skin lesion diagnosis from ISIC 2018 challenge, and thorax disease classification from the publicly dataset ChestX-ray14, and the results have achieved the state-of-the-art criteria.

2.2.6 Weakly (partially) supervised learning

Weak supervision is basically a branch of machine learning used to label unlabeled data by exploiting noisy, limited sources to provide supervision signal that is responsible of labeling large amount of training data using supervised manner. In general, the new labeled data in "weakly-supervised learning" are imperfect, but it can be used to create a robust predictive model. The weakly supervised method uses image-level annotations and weak annotations (e.g., dots and scribbles) [50]. Weakly supervised multi-label disease system was used for classification task of chest X-ray [51], Also, it is used for multi-organ segmentation, [52] by learning single multi-class network from a combination of multiple datasets, where each one of these datasets contains partially organ labeled data and low sample size. Roth et al. [53] have used weakly supervised learning system for medical image segmentation and their results has speeded up the process of generating new training dataset used for the development purpose of deep learning in medical images analysis. Schleg et al. [54] have used this type of deep learning approach to detect abnormal regions from test images. Hu et al. [55] proposed an end-to-end CNN approach for displacement field prediction to align multiple labeled corresponding structures, and the proposed work was used for medical image registration of prostate cancer from T2-weighted MRI and 3D transrectal ultrasound images; the results reached 0.87 of Mean Dice score. Another application is applied in diabetic retinopathy detection in a retinal image dataset [56].

2.2.7 Reinforcement learning

Reinforcement learning (RL) is subtype of deep learning by which it takes the beneficial action toward maximizing the rewards of specific situation. The main difference between supervised learning and reinforcement learning is that in the first one, the training data have the answer within it, but in case of reinforcement learning, the agent decides how to act with the task where in the absence of the training dataset the model learn from its experience. Al Walid et al. [57] have used reinforcement learning for landmark localization in 3D medical images; they have introduced the partial policy-based RL, by learning optimal policy of smaller partial domains; in this paper, the proposed method was used on three different localization task in 3D-CT scans and MR images and proved that learning the optimal behavior requires significantly smaller number of trials. Also in [58], RL was used for object detection PET images. RL was also used for color image classification on neuromorphic system [59].

2.2.7.1 Transfer learning Transfer learning is one of the powerful enablers of deep learning [60], which involves training a deep leaning model by re-using of a an already trained model with related or un-related large dataset. It is known that medical data face the problem of lacking and insufficient for training deep learning models perfectly, so Transfer learning can provide the CNN models with large learned features from non-medical images which in turn can be useful for this case [61]. Furthermore, Transfer Learning is a key feature for time-consuming problem while training a deep neural network, because it uses the freeze weights and hyperparameters of another model. In usual using transfer learning the weights which is already trained on different data (images) are freezed to be used for another CNN model, and only in the few last layers, modifications are done and these few last layers are trained on the real data for tuning the hyperparameters and weights. For these reasons, transfer learning was widely used in medical imaging, for example a classification of the interstitial lung disease [61] and detecting the thoraco-abdominal lymph nodes from CT scans; it was found that transfer learning is efficient, even though the disparity between the medical images and natural images. Transfer learning as well could be used for different CNN models (e.g., VGG-16, Resnet-50, and Inception-V3), Xue

et al. [62], have developed transfer learning-based model for these models, and furthermore, they have proposed an Ensembled Transfer Learning (ETL) framework for classification enhancement of cervical histopathological images. Overall, in many computer vision tasks, tuning the last classification layers (fully connected layers) which is called "shallow tuning" is probably efficient, but in medical imaging, a deep tuning for more layers is needed [63], where they have studied the benefit of using transfer learning in four applications within three imaging modalities (polyp detection from colonoscopy videos, segmentation of the layers of carotid artery wall from ultrasound scans, and colonoscopy video frame classification), their study results found that training more CNN layers on the medical images is efficient more than training from the scratch.

2.3 Best deep learning models and practices

Convolutional Neural Networks (CNNs) based models are usually used in different ways with keeping in minds that CNNs remains the heart of any model; in general, CNN could be trained on the available dataset from the scratch when the available dataset is very large to perform a specific task (e.g., segmentation, classification, detection, etc.), or a pre-trained model with a large dataset (e.g., ImageNet) where this model could be used to train new datasets (e.g., CT scans) with fine-tuning some layers only; this approach is called transfer learning (TL) [60]. Moreover, CNN models could be used for feature extraction only from the input images with more representation power before proceeding to the next stage of processing these features. In the literature, there were commonly used CNN models which has proven their effectiveness, and based on these models, some developments have arisen; we will mention the most efficient and used models of deep learning in medical images analysis. First, it was **AlexNet** which was introduced by Alex Krizhevsky [64] and Siyuan Lu et al. [65], have used transfer learning with a pre-trained AlexNet with replacing the parameters of the last three layers with a random parameters for pathological brain detection. Another frequently used model is **Visual Geometry Group (VGG-16)** [66] where 16 refers to the number of layers; later on, some developments were proposed for VGG-16 like VGG-19; in [67], they have listed medical imaging applications using different VGGNet architectures. Inception Network [68] is one of the most common CNN architectures which aim to limit the resources consumption. And further modifications on this basic network were reported with new versions of it [69]. Gao et al. [70] have proposed a new architecture of Residual Inception Encoder–Decoder Neural Network (RIEDNet) for medical images synthesis. Later on, Inception network was called Google Net [71]. **ResNet** [72] is a powerful architecture for very deep architectures sometimes over than 100

layers, and it helps in limiting the loss of gradient in the deeper layers, because it adds residual connections between some convolutional layers Fig. 4. Some of ResNet models in medical imaging are mostly used for robust classification [73, 74], for pulmonary nodes and intracranial hemorrhage.

DenseNet exploits same aspect of residual CNN (ResNet) but in a compact mode for achieving good representations and feature extraction. Each layer of the network has in its input outputs from the previous layers, so comparing to a traditional CNN, DenseNet contains more connections (L) than CNN (L connections) where DenseNet has $[L(L-1)]/2$ connections. DenseNet is widely used with medical images, Mahmood et al.[78] have proposed a Multimodal DenseNet for fusing multimodal data to give the model the flexibility of combining information from multiple resources, and they have used this novel model for polyp characterization and landmark identification in endoscopy. Another application used transfer learning with DenseNet for fundus medical images [79].

U-net [80] is one of the most popular network architectures used mostly for segmentation, Fig. 4. The reason behind it is mostly used in medical images is that because it is able to localize and highlight the borders between classes (e.g., brain normal tissues and malignant tissues) by doing the classification for each pixel. It is called U-net, because the network architecture takes the shape of U alphabet and it contains concatenation connections; Fig. 4 shows the basic structure of the U-Net. Some developments of U-Net were U-Net + + [75], have proposed a new architecture U-Net + + for medical image segmentation, and in their experiments, U_Net + + has outperformed both U-Net and wide U-Net architectures for multiple medical image segmentation tasks, such as liver segmentation from CT scans, polyp segmentation in colonoscopy videos, and nuclei segmentation from microscopy images. From these popular and basic DL models, some other models were inspired and even some of these models were inspired and rely on the insights from others (e.g., inception and ResNet); Fig. 5 shows the timeline of the mentioned models and other popular models too.

3 Deep learning applications in medical imaging

For the purpose of studying the most applications of deep learning in medical imaging, we have organized a study based on the most-cited papers found in literature from 2015 to 2021; the number of surveyed literatures for segmentation, detection, classification, registration, and characterization are: 30, 20, 30, 10, and 10, respectively. Figure 6 shows the pie chart of these applications.
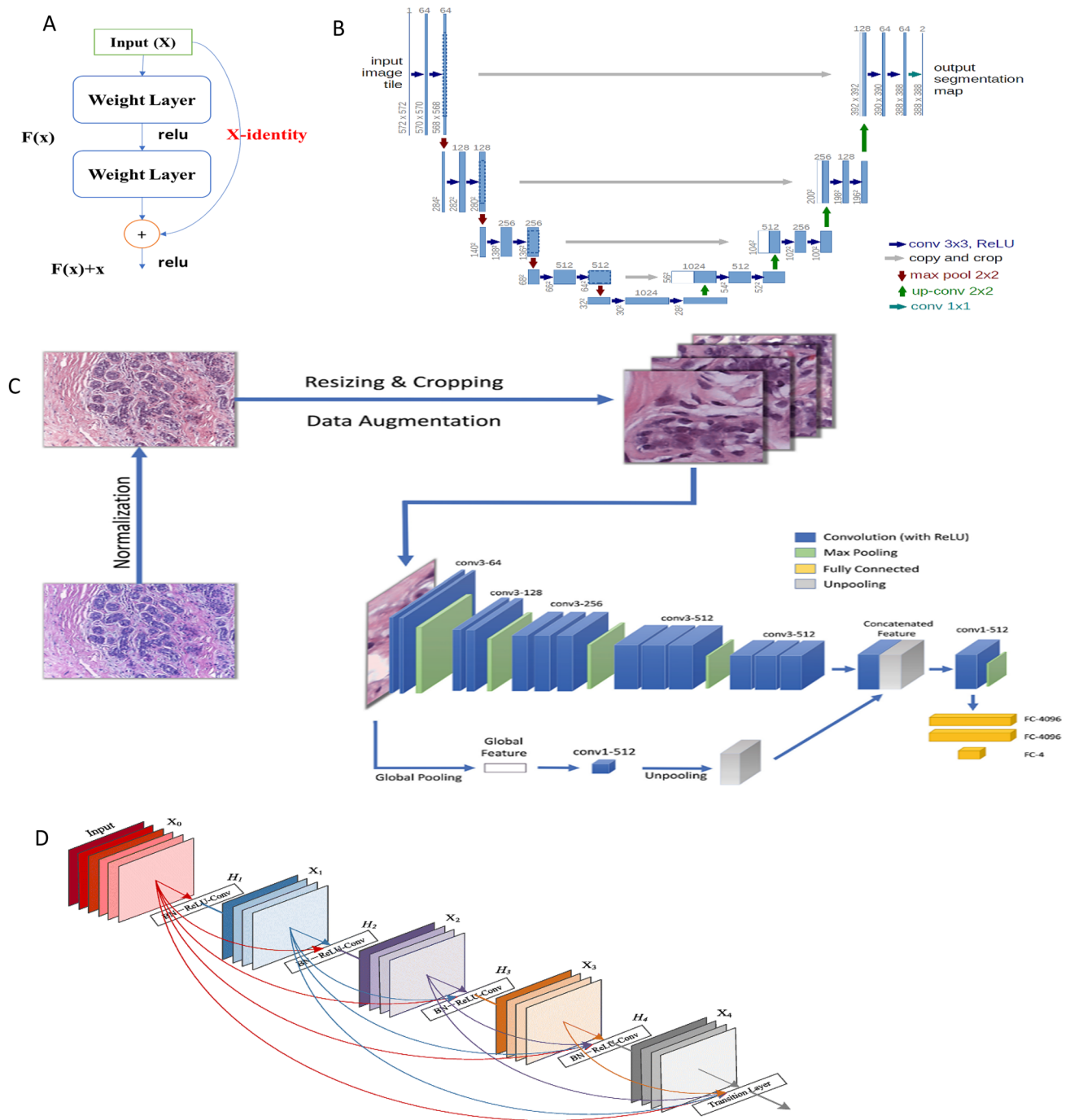
Fig. 4 The basic models used in medical imaging: **A** ResNet architecture, **B** U-Net architecture [75], **C** CNN AlexNet architecture for breast cancer [76], and **D** Dense Net architecture [77]

3.1 Image segmentation

Deep learning is used to segment different body structures from different imaging modalities such as, MRI, CT scans, PET, and ultrasound images. Segmentation means portioning an image into different segments where usually these segments belongs to specific classes (tissue classes, organ, or biological structure) [81]. In general overview, for CNN models, there are two main approaches for segmenting a medical image; the first is using the entire image as an input and the second is using patches from the image. Segmentation process of Liver tumor using CNN architecture is shown in Fig. 7 according to Li et al., and both the methods work well in generating an output map which provides
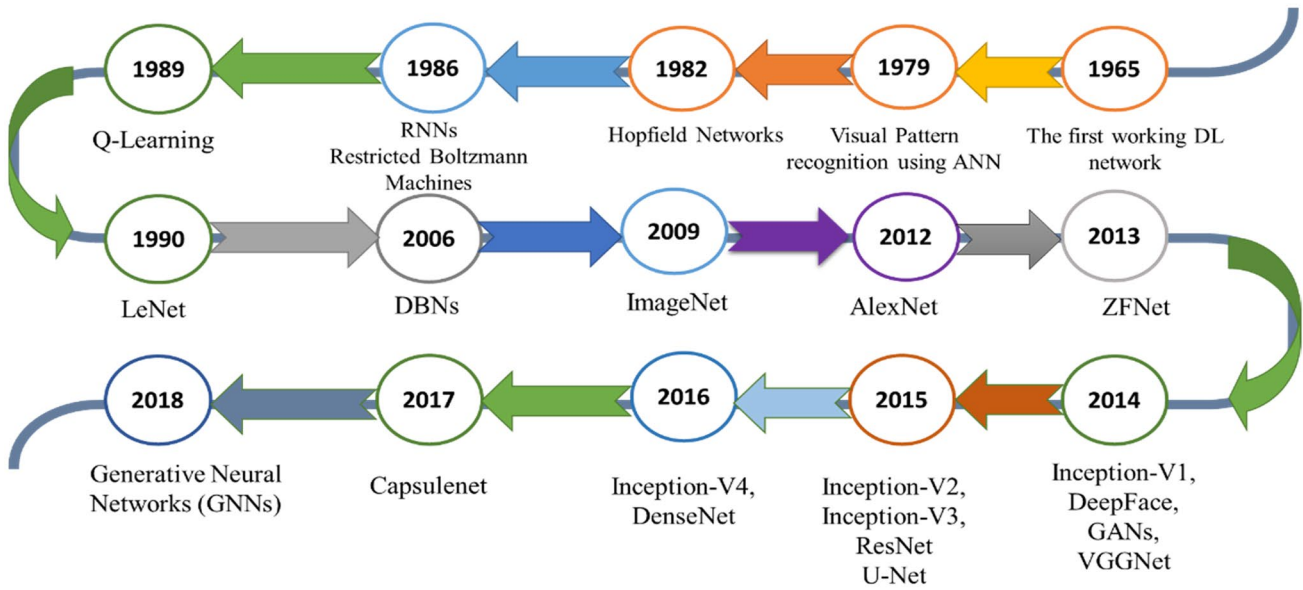
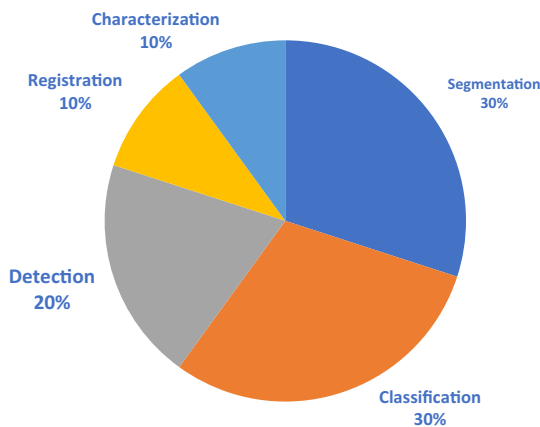**Fig. 5** Timeline of mostly used DL models in medical imaging

Fig. 6 Surveyed DL applications in medical imaging

the segmented output image. Segmentation is potential for surgical planning and determining the exact boundaries of sub-regions (e.g., tumor tissues) for better guidance during the direct surgery resection. Most likely segmentation is common in neuroimaging field and with brain segmentation more than other organs in the body. Akkus et al. [82] have reviewed different DL models for segmentation of different organs with their datasets. Since CNN architecture can handle both 2-dimensional and 3-dimensional images, it is considered suitable for MRI which is in 3D scheme; Milleteria et al. [83] have used 3D MRI images and applier 3D-CNN for segmenting prostate images. They have proposed new CNN architecture which is V-Net which relies on the insights of U-Net [80] and their output results have achieved 0.869 dice similarity coefficient score; this is
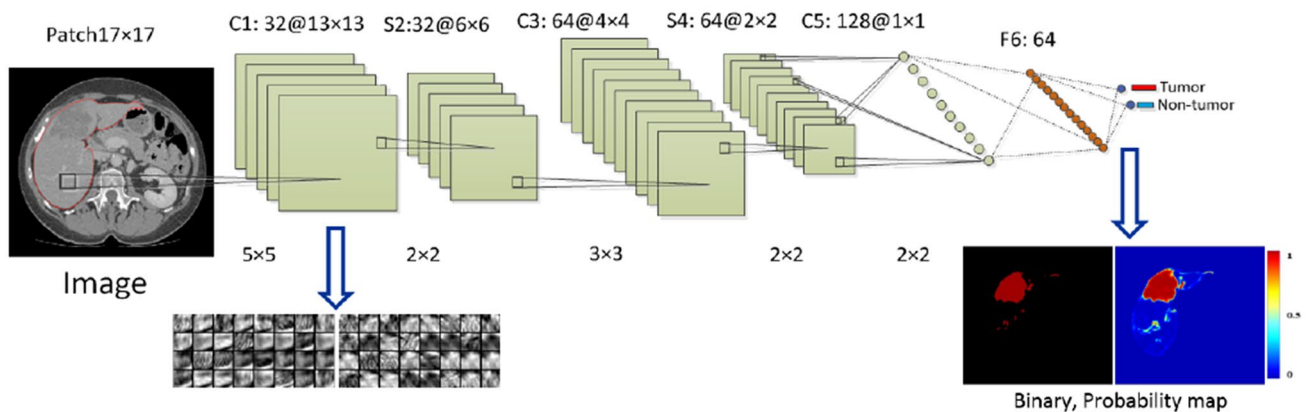

Fig. 7 Liver tumor segmentation using CNN architecture [86]

considered as efficient model regarding to the small dataset (50 MRI for training and 30 MRI for testing). Havaei et al. [84] have worked on Glioma segmentation from BRATS-2013 with 2D-CNN model and this model took only 3 min to run. From clinical point of view, segmentation of organs is used for calculating clinical parameters (e.g., volume) and improving the performance of Computer-Aided Detection (CAD) to define the regions accurately. Taghanaki et al. [85] have listed the segmentation challenges from 2007 to 2020 with different imaging modalities; Fig. 8 shows the number of these challenges. We have summarized Deep Learning models for segmentation for different organs in the body, based on the highly cited paper and variations in deep learning models shown in Table 1.

3.2 Image detection/localization

Detection simply means is to identify a specific region of interest in an image and finally to draw a bounding box around it. Localization is just another terminology of detection which means to determine the location of a particular structure in images. In deep learning for medical images, analysis detection is referred as Computer-Aided Detection (CAD), Fig. 9. CAD is divided commonly for anatomical structure detection or for lesions (abnormalities) detection. Anatomical structure detection is a crucial task in medical images analysis due to determining the locations of organs substructures and landmarks which in turn guide for better organ segmentation and radiotherapy planning for analysis and further surgical purposes. Deep learning for organ or lesion detection can be either classification-based or regression-based methods; the first one is used for discriminating body parts, while the second method is used for determining more detailed locations information. In fact, most of the deep learning pathologies are connected; for example, Yang et al. [114] have proposed a custom CNN classifier

for locating landmarks which is the initialization steps for the femur bone segmentation. In case of lesion detection which is considered to be clinically time-consuming for the radiologists and physicians and it may lead to errors due to the lack of data needed to find the abnormalities and also to the visual similarity of the normal and abnormal tissues in some cases (e.g., low contrast lesions in mammography). Thus, the potential of CAD systems comes from overcoming these cons, where it reduces the times needed, computational cost, providing alternative way for the people who live in areas that lacks specialists and improve the efficiency of thereby streamlining in the clinical workflow. Some CNN custom models were developed specifically for lesion detection [115, 116]. Both organ anatomical structures and lesion detection are applicable for mostly all body's organs (e.g., Brain, Eye, Chest, Abdominal, etc.), and CNN architectures are used for both 2D and 3D medical images. When using 3D volumes like MRI, it is better to use patches fashion, because it is more efficient than sliding window fashion, so in this way, the whole CNN architecture will be trained using patches before the fully connected layer, [117]. Table 2 shows top-cited papers with different deep learning models for both structure and lesion detection within different organs.

3.3 Image classification

This task is the fundamental task for the computer-aided diagnosis (CAD), and it aims to discover the presence of disease indicators. Commonly in medical images, the deep learning classification model's output is a number that represents the disease presence or absence. A subtype of classification is called lesion classification and is used in a segmented images from the body [136]. Traditionally, classification used to rely on the color, shape, and texture, etc. but in medical images, features are more complicated to be categorized as these low-level features which lead to poor model generalization due to the high-level features for medical image. Recently, deep learning has provided an efficient way of building an n end-to-end model which produce classification labels-based from different medical images' modalities. Because of the high resolution of medical images, expensive computational costs arise and limitations in the number of deep model layers and channels; Lai Zhifei et al. [137] have proposed the Coding Network with Multilayer Perceptron (CNMP) to overcome these problems by combining high-level features extracted by CNN and other manually selected common features. Xiao et al. [138] have used parallel attention module (PAM-DenseNet) for COVID-10 diagnosis, and their model can learn strong features automatically from channel-wise and spatial-wise which help in making the network to automatically detect the infected areas in CT scans of lungs without the need
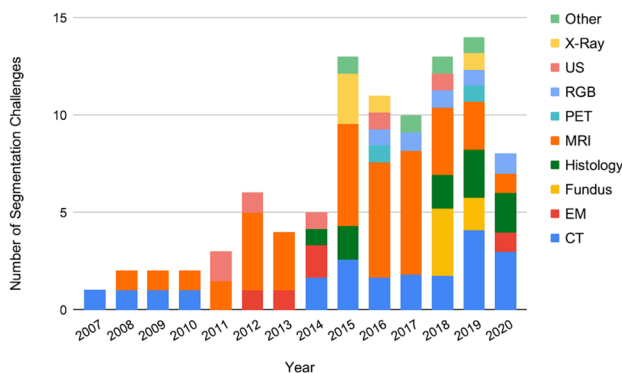

Fig. 8 The number of challenges related to segmentation in medical imaging from 2007 to 2020 listed on Grand Challenges regarding the imaging modalities

Table 1 DL models for medical images' segmentation

References	Algorithm	Organ (substructure)	Dataset	Performance	Citation
Dong et al. [87]	U-Net FCN	Brain (tumor)	BRATS-2015	DSC=0.88, 0.87, 0.81 For complete, core, and enhancing tumor, respectively	427
Soltaninejad et al. [88]	Random Forest + FCN for feature learning	Brain (tumor)	BRATS-2013	DSC=0.88, 080 and 0.73 for complete tumor, core, and enhancing tumor, respectively	14
Havaei et al. [84]	2D-CNN (DNN)	Brain	BRATS-2015	DSC=0.88, 0.79, 0.73 For complete, core, and enhancing tumor, respectively	1974
Wang et al. [89]	Cascaded CNN with residual connections	Brain (tumor)	BRATS-2017	DSC=0.87, 0.77, 0.78 For complete, core, and enhanced tumor, respectively	296
Chen et al. [90]	DeconvNet	Brain (ischemic lesions)	Locally collected from 761 patient	Dice coefficient=0.88	149
Li et al. [91]	GAN	Brain (tumor)	BRATS-2017	DSC=0.87, 0.72, 0.68 For complete, core, and enhanced tumor	32
Zhou et al. [92]	3D variant of FusionNet (One-pass Multi-task Network (OM-Net))	Brain (tumor)	BraTS-2018	DSC=0.916 (WT), 0.827 (TC), 0.807 (EC)	34
Isensee et al. [93]	3d-Unets	Brain (tumor)	BraTS-2017	DSC=0.850 (WT), 0.740 (TC), 0.640 (EC)	300
Korfiatis et al. [94]	AutoEncoder	Brain (white matter hyperintensities)	BRATS-2015 + locally collected glioma dataset	Dice coefficient=0.88	48
Liu et al. [95]	landmark-based deep multi-instance learning (LDMIL)	Brain (Alzheimer)	1526 MRIs (ADNI-1, ADNI-2, MIRIAD)	F-score=0.33 Accuracy=76.9%	160
Liskowski et al. [96]	DNN	Eye	DRIVE, STARE, and CHASE databases	ROC=0.99 Sensitivity>0.87	593
Fang et al. [97]	CNN with graph search (CNN-GS)	Eye	60 OCT volumes from 20 patients	–	315
Shankaranarayana et al. [98]	Conditional GAN, ResU-net	Eye	RIM-ONE dataset	F-score (disc=0.97, cup=0.94) IOU (disc=0.89, cup=0.76)	68
Fu et al. [99]	M-Net and Polar Transformation	Eye	ORIGA	AUC=0.8508 SCES dataset: AUC=0.8998	337
Son et al. [33]	U-Net, GAN	Eye	DRIVE, STARE Databases	DRIVE dataset: DSC=0.829 ROC=0.9803 STARE dataset, DSC=0.834 ROC=0.9838	117
Li et al. [86]	Custom CNN	Liver (tumor)	30 CT scans	Dice coefficient=0.80	150
Peijun Hu et al. [100]	Custom 3D-CNN	Liver (structure)	121 CT scans locally collected + 30 scans from SILVER07 dataset	Average score=80.3±4.5	129
Ben Cohen et al. [101]	FCN	Liver (lesion)	20 CT scans locally collected + 20 CT scans from SILVER07 Dataset	Average Dice index of 0.89	150

Table 1 (continued)

| References	Algorithm	Organ (substructure)	Dataset	Performance	Citation
Yang et al. [102]	deep image-to-image network with Adversarial Training (DI2IN-AN)	Liver (structure)	1000 3D-CT volumes	Mean Dice Score = 0.95	159
Cheng et al. [103]	Stacked Denoising Auto Encoder (SDAE)	Breast	1400 CT scans locally collected	Accuracy = (82.4 ± 4.5) %	528
Al-Antari et al. [104]	YOLO + full resolution convolutional network (FrCN)	Breast	410 X-ray mammography images from (INbreast database)	Accuracy = 98.96%, MCC = 97.62%, F1-score = 99.24%, AUC = 93.05%	128
Skourt et al. [105]	U-Net	Lung (parenchyma (nodule) segmentation)	(LIDC-IDRI) dataset	dice coefficient index = 0.9502	115
Kalinovsky et al. [106]	Encoder-Decoder CNN(ED-CNN)	Lung	354 X-Ray chest images	Min-Dice score = 0.926, Max-Dice score = 0.974, Accuracy = 0.96	46
Roy et al. [107]	Custom CNN (Reg-STN) (CNN + Reg-STN + SORD model) SORD: Soft ORDinal regression Regularized Spatial Transformer Networks (Reg-STN)	Lung (COVID-19)	(ICLUS-DB) 277 lung ultrasound (LUS) videos from 35 patients,	Accuracy = 96, Binary Dice score = 0.75, F1-score = 71.4%	131
Murphy et al. [108]	U-Net and CNN	Lung (COVID-19)	Private CXR images (24 678 chest radiographs)	AUC = 0.81	101
Kline et al. [109]	Custom CNN	Kidney	2400 MRI locally generated images	Dice score = 0.97 ± 0.01, Jaccard score = 0.94 ± 0.03	67
Jinlian Ma et al. [110]	Custom CNN	Thyroid (nodule)	US images from 250 patient	Dice coefficient = 0.92	63
Zhang et al. [111]	multiple supervised residual network (MSRN)	Bone (osteosarcoma)	1900 CT images from 15 osteosarcoma patients	Dice similarity coefficient (DSC) = 89.22%, a sensitivity = 88.74% and a F1-measure = 0.9305	17
Li Yu et al. [112]	Custom CNN	Heart (Fetal Left Ventricle)	51 patients US	Dice coefficient = 94.5%	60
Jafari et al. [113]	Custom CNN	Skin (melanoma)	126 clinical images	Accuracy = 98.7%	60
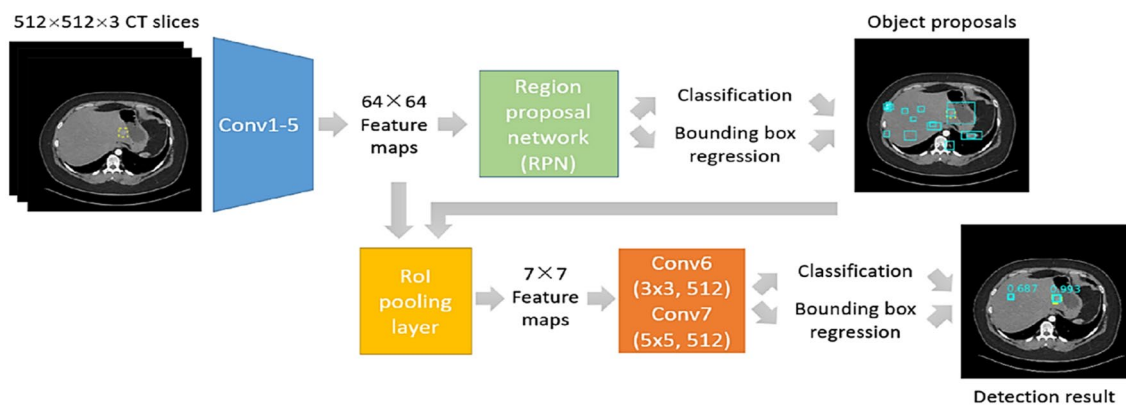
Fig. 9 Lesion detection algorithm flowchart [118]

of manual delineation. As any deep learning application, classification task is performed on different body organs for detecting diseases' patterns. Back in 1995, a CNN model was developed for detecting lung nodules from X-ray of chest [139]; classifying medical images is essential part for clinical aiding and further treatments, for example detecting and classifying pneumonia presence from chest X-ray scans [140]; CNN-based models have introduced various stratifies to better the classification performance especially when using small datasets, for example data augmentation [141, 142]. GANs' network was widely used for data augmentation and image synthesis [143]. Another robust strategy is transfer learning [61]. Rajpurkar et al. have used custom DenseNet for classifying 14 different diseases using chest X-ray from the chestXray14 dataset [129]. Li et al. have used 3D-CNN for interpolating the missing pixels data between MRI and PET modalities, where they have reconstructed PET images from MRI images from the (ADNI) dataset of Alzheimer disease which contain MRI and PET images [144]. Xiuli Bi et al. [31] have also worked on Alzheimer disease diagnosing using a CNN architecture for feature extraction and unsupervised predictor for the final diagnosis results on (ADNI-1 1.5 T) dataset and achieved accuracy of 97.01% for AD vs. MCI, and 92.6% for MCI vs. NC. Another 3D-CNN architecture employed in an autoencoder architecture is also used to classify Alzheimer disease using transfer learning on a pre-trained CAD Dementia dataset, they have reported accuracy of 99% on the publicly dataset ADNI, and the fine-tuning process is done in a supervised manner [145]. Diabetic Retinopathy (DR) could be diagnosed using fundus photographs of the eye, Abramoff et al.[146] have used custom CNN inspired from Alexnet and VGGNet to train a device (IDx-DR) version X2.1 on a dataset of 1.2 million DR images to record 0.98 AUC score. Figure 10 shows the classification of medical images. A few notable results found in literature are summarized in Table 3.

3.4 Image registration

Image registration means to allow images' spatial alignment to a common anatomical field. Previously, image registration was done manually by clinical experts, but after deep learning, image registration has changed [176–178]. Practically, this task is considered main scheme in medical images, and it relies on aligning and establishing accurate anatomical correspondences between a source image and target image using transformations. In the main theme of image registration, both handcrafted and selected features are employed in a supervised manner. Wu et al. [179, 180] have employed unsupervised deep learning approach for learning the basis filters which in turn represent image's patches and detect the correspondence detection for image registration. Yang et al. [177] have used an autoencoder architecture for predicting of deformation diffeomorphic metrics mapping (LDDMM) to get fast deformable image registration and the results shows improvements in computational time. Commonly, image registration is employed for spinal surgery or neurosurgery in form of localization of spinal bony or tumor landmarks to facilitate the spinal screw implant or tumor removal operation. Miao et al. [181] have trained a customized CNN on X-ray images to register 3D models of hand implant and knee implant onto 2D X-ray images for pose estimation. An overview of registration operation is shown in Table 4, which shows a summary of medical images registration as an application of deep learning.

3.5 Image characterization

Characterization of a disease within deep learning is a stage of computer-aided diagnosis (CADx) systems. For example, radiomics is an expansion of CAD systems for other tasks such as prognosis, staging, and cancer subtypes' determination. In fact, characterization of a disease will rely on the disease type in the first place and on the clinical questions

Table 2 Deep learning applications to medical imaging detection

References	Organ	CNN architecture	Dataset	Evaluation metrics	Citations
Zhang et al. [119]	Brain (landmarks)	FCN	3D T1-weighted (MR) images of 700 subjects(350/350)	Error (mean _ SD) = 2.94 _ 1.58 mm 1200 brain landmarks were detected	109
Zhang et al. [119]	Prostate	FCN	3D computed tomography (CT) images of 73 subjects	Error (mean _ SD) = 3:34 _ 2:55 7 prostate landmarks were detected	109
Nakao et al. [120]	Brain (cerebral aneurysms)	Custom CNN	Locally collected MRI dataset	Accuracy = 94.2%	74
Tsehay et al. [121]	Prostate (cancer)	DCNN	196 MRI locally collected dataset (train = 144, test = 52)	0.86 = AUC	30
Korsuk Sirinukunwattana et al. [122]	Abdomen (nucleus) colon cancer)	Spatially Constrained Convolutional Neural Network (SC-CNN)	2000+ histology images	Recall = 0.827 F-score = 0.802 Median distance = 2.236	740
Setio et al. [123]	Lung (pulmonary nodule)	Custom CNN	888 CT scans of LIDC-IDRI dataset	Competition Performance Metric score (CPM) = 0.827 AUC = 0.996	806
Li et al. [124]	Lung (COVID-19)	U-Net and ResNet-50	4356 chest CT exams from 3,322 patients	AUC = 0.96	695
Eduardo Luz et al. [125]	Lung (COVID-19)	EfficientNet	13,569 X-ray images	overall accuracy of 93.9%, COVID-19 sensitivity of 96.8%	46
Kassania et al. [126]	Lung (COVID-19)	DenseNet -121 and ResNet-50	50 Chest X-ray and 150 CT scans	DenseNet -121Accuracy = 99%, ResNet-50 accuracy = 98%	67
Wang et al. [127]	Breast(cancer)	Autoencoder	Histopathology Camelyon16 dataset consists of a total of 400 whole slide images (WSIs)	AUC = 0.995 for classification score of 0.7051 for the tumor localization	658
Dou et al. [128]	Lung (Pulmonary Nodule)	3D-CNN	LUNA16 dataset 888 CT scans	CPM score = 0.827	357
Rajpurkar et al. [129]	Chest (Pneumonia)	CheXnet (CNN)	ChestX-ray14 dataset over 100,000 X-ray images	F1-Score = 0.435 (95% CI 0.387, 0.481),	1184
Jinlian Ma et al. [130]	Thyroid(nodules)	Cascade CNN	21,523 ultrasound images	AUC = 98.51% confidence interval (CI) = 95%	78
Baka et al. [131]	Vertebral	U-Net	Ultrasound and X-ray (train = 25 test = 19)	Recall = 0.88 F-means = 0.90 Accuracy = 92%	40
Alex et al. [132]	Brain (Lesion)	GANs	MRI BraTS dataset	whole tumor dice score of 0.69, sensitivity of 91%	45
Bogunović et al. [133]	Eye (Fluid)		112 Optical coherence tomography (OCT)	AUC = 1.0 Accuracy = 90%	49
Varun Gulshan et al. [134]	Eye (diabetic retinopathy)	Inception-V3	1748 images from Messidor-2 dataset	AUC = 0.99, Sensitivity = 87%, Specificity = 98.50%	3437
Varun Gulshan et al. [134]	Eye (diabetic retinopathy)	Inception-V3	9963 from EyePACS-1 dataset	AUC = 0.991, Sensitivity = 90.3 Specificity = 98.10	3437
Hoo Chang Shin et al. [61]	Thoraco-abdominal lymph node (LN)	Custom CNN	983 2D CT from 176 patients	AUC = 0.94 for abdomen AUC = 0.95 for mediastinum	3188

Table 2 (continued)

References	Organ	CNN architecture	Dataset	Evaluation metrics	Citations
Nasrullah et al. [135]	Lung (Nodules)	(Custom CNN) 3D CMixNet with faster R-CNN	888 CT scan low-dosed lung (LUNA16 dataset)	x free response receiver-operating characteristic (FROC) = 94.21%	55
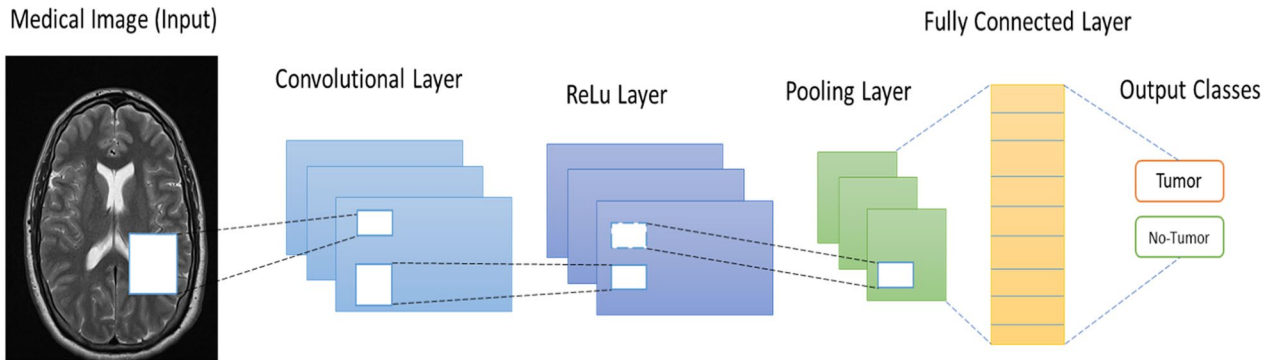
Fig. 10 Classification of brain tumor using general CNN architecture

related to it. There is two ways used for features extraction, either handcrafted features extraction or deep learned features, in the first, radiomic features is similar to radiologist's way of interpretation and analysis of medical images. These features might include tumor size, texture, and shape. In literature, the handcrafted features are used for many purposes, such as tumor aggressiveness, the probability of having cancer in the future, and the malignancy probability [190, 191]. There are two main categories for characterization, lesion characterization and tissue characterization. In deep learning applications of medical imaging, each computerized medical image requires some normalization plus customization to be handled and suited to the task and image modality. Conventional CAD is used for lesion characterization. For example, to track the growth of lung nodules, the characterization task is needed for the nodules and the change of lung nodules over time, and this will help of reducing the false-positive of lung cancer diagnosis. Another example of tumor characterization is found in imaging genomics, where the radiomic features are used as phenotypes for associative analysis with genomics and histopathology. A good report which was done with multi-institutes' collaboration about breast phenotype group through TGCA/TCIA [192–194]. Tissue characterization is to examine when particular tumor areas are not relevant. The main focus in this type of characterization is on the healthy tissues that are susceptible for future disease; also focusing on the diffuse disease such as interstitial lung disease and liver disease [195]. Deep learning has used conventional texture analysis for lung tissue. The characterization of lung pattern using patches can be informative of the disease which commonly is interpreted by

radiologists. Many researchers have employed DL models with different CNN architectures for interstitial lung disease classification characterized by lung tissue sores [149, 196]. CADx is not only a detection/localization task only, but it is classification and characterization task as well. Finding the likelihood of disease subtyping is the output of a DL model and characteristic features' presentation of a disease too. For the characterization task, especially with limited dataset, CNN models are not trained from scratch in general, data augmentation is an essential tool for this application, and performing CNN on dynamic contrast-enhanced MRI is important too. For example, while using VGG-19-Net, researchers have used DCE-MRI temporal images with pre-contrast, first post-contrast, and the second post-contrast MR images as an input to the RGB channels. Antropova et al. [197] have used the maximum intensity projections (MIP) as an input to their CNN model. Table 5 shows some highlighted literature of characterization which includes diagnosis and prognosis.

3.6 Prognosis and staging

Prognosis and staging refer to the future prediction of a disease status for example after cancer identification, further treatment process through biopsies which give a track on the stage, molecular type, and genomics which finally provides information about prognosis and the further treatment process and options. Since most of the cancers are spatially heterogeneous, specialists and radiologists are interested about the information on spatial variations that medical imaging can provide. Mostly, many imaging biomarkers include only

Table 3 Deep learning applications to medical imaging classification

References	Anatomic Organ	DL architecture	Dataset	Results evaluation metrics	Citations
Anthimopoulos et al. [147]	Lung	Custom CNN	120 ILD CT scans (ILD CT scans 109 HRCT scans)+(Bern University Hospital, "Inselspital" and consists of 26 HRCT scans of ILD cases)	Accuracy=0.8561 Favg=0.8547	836
Nibali et al. [148]	Lung (Pulmonary nodule) _	ResNet	LIDC/IDRI dataset 1010 CT scans	Accuracy=89.9 AUROC=0.9459	136
Christodoulidis et al. [149]	Lung (tissues)	Custom CNN	(ALOT), (DTD), (FMD), (KTB), KTH-TIPS-2b, (HUG) Databases	F1-score $_{avg}$=0.8817	198
Wu et al. [150]	Lung (COVID-19)	ResNet-50	Chest CT images of 495 patients	(AUC)=0.732, accuracy=0.700, sensitivity=0.730 and specificity=0.615 in validation set	76
Farid et al. [151]	Lung (COVID-19)	Naïve Bayes	CT scans from online access Kaggle benchmark dataset,2020)	Accuracy of 96.07%	37
Pereira et al. [152]	Lung (COVID-19)	CNN	CXR, a database, named RYDLS-20	F1-Score of 0.89	161
Huynh et al. [153]	Breast	Custom CNN	Mammography, Private data (219 lesions)	AUC=0.86	340
Sun et al. [154]	Breast (Cancer)	Custom CNN	1874 mammogram images	Accuracy=0.8243 AUC=0.8818	160
Swati et al. [155]	Brain (tumor)	Custom CNN+transfer learning	3064 Contrast Enhanced (CE-MRI)	Accuracy=94.82%	90
Sajjad et al. [156]	Brain (tumor)	Custom CNN	3064 MRI	Accuracy=95.5%	180
Deepak et al. [157]	Brain (tumor)	Custom CNN	3064 brain MRI images from 233 patients (figshare dataset)	Accuracy=98%	150
Afshar et al. [158]	Brain (tumor)	Capsule networks (CapsNet)	3064 MRI images of 233 patients	Accuracy=86.56%	202
Gao et al. [159]	Brain (AD)	3D-CNN	285 datasets of 3D are collected from Navy General Hospital (CT scans)	Accuracy=87.6%	189
Bharati et al. [160]	Lung	hybrid CNN VGG Data STN (VDSNet)	5606 X-Ray images	Accuracy=73%, Fβ(0.5)-score=0.68, Recall=0.63, Precision=0.69	34
Nasrullah et al. [135]	Lung (Nodules)	Customized mixed link network 3D CMixNet+GBM+Biomarkers	1018 CT scans (LIDC-IDRI dataset)	Accuracy=94.17	52
Zhou et al. [161]	Breast (Cancer)	Weakly supervised 3D-CNN	Privately collected from 1537 female patients	Accuracy=83.7% Confidence Interval (CI)=95%	35
Zhang et al. [162]	Breast (tumor)	PGBM-RBM-SVM	227 shear-wave elastography (SWE)	Accuracy=93.4%, Sensitivity=88.6, Specificity=97.1 AUC=0.947	142
Yang et al. [163]	Brain (AD)	3D VggNet, 3D Resnet	MRI scans from ADNI dataset (47 AD, 56 NC)	86.3% AUC using 3D VggNet and 85.4% AUC using 3D ResNet	58

Table 3 (continued)

References	Anatomic Organ	DL architecture	Dataset	Results evaluation metrics	Citations
Schwyzer et al. [164]	Lung (cancer)	ANN	3936 PET slices	Standard dose): AUC = 0.989, sensitivity = 95.9%,, specificity = 98.1% (Ultralow dose): AUC = 0.970, sensitivity = 91.5%, specificity = 94.2%	44
de Carvalho Filho et al. [165]	Lung	Custom CNN	50,580 (14,184 malignant and 36,396 benign) nodules (LIDC-IDRI)	ACC = 92.63%, sensitivity = 90.7%, Specificity = 93.47% ROC = 0.934	33
Shen et al. [166]	Lung (cancer)	Multi-crop CNN MC-CNN	LIDC-IDRI (CT scans) dataset, 1010 patients	Accuracy = 87.14% AUC = 0.93, Sensitivity = 0.77, Specificity = 0.93	322
Ozturk et al. [167]	Lung (COVID-19)	Custom CNN (DarkCovidNet)	481 CXR images	Accuracy = 98.08% for binary classification, Accuracy = 87.02% for multi-class cases	679
Ucar et al. [168]	Lung (COVID-19)	SqueezNet	5949 Chest X-ray scans (COVIDx)	Accuracy = 98.3% F1-score = 0.983, Matthew Correlation Coefficient (MCC) = 0.974	205
Wang et al. [169]	Lung (COVID-19)	Custom CNN (COVID-Net)	13,975 CXR images across 13,870 patients	Accuracy94.3%, sensitivity = 96.8%	925
Rezvantalab [170]	Skin (cancer)	DenseNet 201, ResNet152	10,135 dermoscopy skin images Actinic Keratoses (AK), Basal cell carcinoma (BCC), Squamous cell carcinoma (SCC)	AUC (ResNet-152) = (94.40%) F1-score (DenseNet) = (89.01–85.13%), ROC AUC (98.79–98.16%). And for Basal cell carcinoma ROC AUC (99.30%)	36
Dorj et al. [171]	Skin (cancer)	ECOC SVM + CNN + Transfer learning of AlexNet	3753 images	Accuracy 95.1%, Sensitivity = 98.9% Specificity = 94.17	111
Esteva et al. [172]	Skin (cancer)	Google's Inception v3 CNN	129,450 clinical images	Accuracy = 93.33%	6173
Awais et al. [173]	Eye (Diabetic Macular Edema (DME))	CNN- VGG16	OCT (Optical Coherence Tomography)	Accuracy = 87.5%, SENSITIVITY = 93.5% SPECificity = 81%	37
Ting et al. [174]	Eye (diabetic retinopathy)	A deep learning system (DLS)	40 752 images	AUC range was 0.889–0.983 AUC was 0.936 for (dataset: n = 71,896)	796
Frid-Adar et al. [175]	Liver (tumor)	GAN	2D CT scans 182 cases (53 cysts, 64 metastases and 65 hemangiomas)	Sensitivity = 85.7%, Specificity = 92.4%	348

Table 4 Deep learning applications to medical imaging for registration

Ref	Anatomical structure	DL architecture	Dataset	Transformation	Citations
Miao et al. [181]	Chest	Custom CNN regression	Total Knee Arthroplasty (TKA) Kinematics X-ray + Virtual Implant Planning System (VIPS) + X-ray Echo Fusion (XEF)	Rigid	295
De vos et al. [182]	Brain	Deep learning network for deformable image registration (DIRNet)	Sunnybrook Cardiac Data (SCD) contains 45 cardiac cine MRI + handwritten digits from the MNIST database	Deformable	267
Sun et al. [183]	Brain	3D-CNN	MRI (RESECT dataset)	Deformable	15
Chen et al. [184]		Biogeography-based Optimization (BBO) algorithm	Retrospective Image Registration Evaluation Project (RIRE) with 6 modalities (CT, PET, MR-PD, MR-T1, MR-T2, MRMP-RAGE)	Geometric rigid transformation	31
Niethammer et al. [185]	Brain	CNN	MRI (LPBA (40) dataset 2D 3D CUMC (12), MGH(10), and IBSR(18) datasets)	Deformable	31
Wu et al. [186]	Brain	CNN	MRI, LONI, ADNI databases	Deformable	212
Kang et al. [187]	Abdomen	CNN	CT low-dose Grand Challenge	Contourlet transform + wavelet transform	486
Zhang et al. [162]	Head, Abdomen, Chest	CNN	Private CT scans	Deformable	142
Stankevièius et al. [188]	Eye	DNN	3153 retinal images	Rigid	6
Haskins et al. [189]	Prostate cancer	CNN	679 MR images from the National Institute of Health (NIH)	Deformable	53

Table 5 Deep learning applications to medical imaging for characterization

References	Anatomical site/task	Network	Dataset	Citation
Aly Mohamed et al. [198]	Breast (Cancer risk assessment)	Customized AlexNet	Mammogram images (Train = 14,000, test = 1850)	152
Juhun Lee et al. [199]	Breast (Cancer risk assessment)	Fine-tuned a pre-trained VGG-16Net	Mammograms 604 images	42
Antropova et al. [200]	Breast (Diagnosis)	Pre-trained VGGNet 19	Mammograms (245), MRI(690), US(1125)	188
Samala et al. [201]	Breast (Diagnosis)	AlexNet	Mammograms (train = 1545, test = 909) images	115
Kooi et al. [202]	Breast (Solitary cyst diagnosis)	VGGNet customized	1600 Mammograms lesions	93
Anthimopoulos et al. [147]	Lung (Interstitial disease)	Custom VGG	14,696 CT patches	849
Masood et al. [203]	Lung (Pulmonary nodule staging)	DFCNet	LIDC-IDRI database, RIDER, LungCT-Diagnosis, and Lung Nodule Analysis (LUNA) 2016 Dataset	92
Christodoulidis et al. [149]	Lung (Interstitial disease)	Custom CNN	CT patches (train = 36,106, test = 1050) from ALOT and KTH-TIPS-2b databases	200
González et al. [204]	Lung and chest (prognosis of chronic obstructive pulmonary disease (COPD))	Custom CNN	CT scans (train = 7983, test = 1000 COPDGene + 1,672 ECLIPSE)	112
Lao et al. [205]	Brain (Survival)	CNN-S with transfer learning	Multiparametric MRI (112 patients	284

the size and another simple enhancement procedures; therefore, the current investigators are more interested in including radiomic features and extending the knowledge from medical images. Some deep learning analysis have been investigated in cancerous tumors for prognosis and staging [192, 206]. The goal of prognosis is to analyze the medical images (MRI or ultrasound) of cancer and get the better presentation of it by gaining the prognostic biomarkers from the phenotypes of the image (e.g., size, margin morphology, texture, shape, kinetics, and variance kinetics). For example, Li et al. [192] found that texture phenotype enhancement can characterize the tumor pattern from MRI, which lead to prediction of the molecular classification of the breast cancers; in other words, the computer-extracted phenotypes provide promises regarding the quality of the breast cancer subtypes' discrimination which leads to distinct quantitative prediction in terms of the precise medicine. Moreover, with the enhancement of the texture entropy, the vascular uptake pattern related to the tumor became heterogeneous which in turn reflects the heterogeneous temperament of the angiogenesis and the treatment process applicability and this is termed as the virtual digital biopsy location based. Gonzalez et al. [204] have applied DL on thoracic CT scans for prediction of staging of chronic obstructive pulmonary disease (COPD). Hidenori et al. [207] have used CNN model for grading diabetic retinopathy and determining the treatment and prognosis which involves a non-typically visualized on fundoscopy of retinal area; their novel AI system suggests treatment and determines prognoses.

Another term related to staging and prognosis is survival prediction and disease outcome, Skrede et al. [208] have performed DL using a large dataset over 12 million pathology images to predict the survival outcome for colorectal cancer in its early stages, a common evaluation metric is Hazard function which indicate the risk measures of a patient after treatment, and their results yield a hazard ration of 3.84 for poor against good prognosis in the validation set cohort of 1122 patients, and a hazard ratio of 3.04 after adjusting for prognostic markers which contain T and N stages. Sillard et al. [209] used deep learning for predicting survival outcomes after hepatocellular carcinoma resection.

3.7 Medical imaging in COVID-19

Basically, after COVID-19 has been identified in 31 December 2019 [210] and it is based on polymerase chain reaction (PCR) test. However, it was found that it can be analyzed and diagnosed through medical imaging, even though most radiologists' societies do not recommend it, because it has similar features of various pneumonia diseases. Simpson et al. [211], have prospected a potential use of CT scans for clinical managing, and eventually,

they have proposed four standard categories for reporting COVID-19 languages. Mahmood et al. [212] have studied 12,270 patients and recommend to be subjected for CT screening for early detection of COVID-19 to limit the speedy spread of the disease. Another approach for classification of COVID-19 is using portable (PCXR) which uses chest X-ray scans instead of the expensive CT scans; furthermore, this has the potential of minimizing the chances of spreading the virus. For the identification of COVID-19, Pereira et al. [152] have flowed using chest X-ray scans using the portable manner. For the comparison of different screening methods, it was suggested by Soldati et al. [213], which stated the Lung Ultrasound (LUS) is needed to be compared with chest X-ray and CT scans to help designing better diagnostic system to be suitable for the technological resources. COVID-19 has gained the attention of deep learning researchers who have employed different DL models for the main pathologies for diagnosing this disease using different medical imaging modalities from different datasets. Starting with segmentation, a new proposed system for screening coronavirus disease was done by Butt et al. [214], who have employed 3D-CNN architecture for segmenting multiple volumes of CT scans; a classification step is included to categorize patches into COVID-19 from other pneumonia diseases, such as influenza and a viral pneumonia. After that, Bayesian function is used to calculate the final analysis report. Wang et al. [215] have performed their CNN model on chest X-ray images, for extracting the feature map, classification, regression, and finally the needed mask for segmentation. Another DL model using chest X-ray scans was introduced by Murphy et al. [108], using U-Net architecture for detecting of tuberculosis and finally classifying images, with AUC of 0.81. For the detection of COVID-19, Li et al. [124] have developed a new tool of deep learning to detect COVID-19 from CT scans; the main work consists of few steps starting from extracting the lungs as ROI using U-Net, then generating features using ResNet-50, and finally using fully connected layer for generating the probability score of COVID-19 and the final results have reported AUC of 0.96. Another COVID-19 detection system from X-rays and CT scans was proposed by Kassani et al. [126], who have used multiple models for their strategy, DenseNet 121 have achieved accuracy of 99%, and REsNet achieved accuracy of 98% after being trained by LightGBM, and also, they have used other backbone models such as MobileNet, Xception, and Inception-ResNet-V2,NASNe, and VGG-Net. For classification of COVID-19, Wu et al. [150] have used the fusion of DL networks, starting from segmenting lung regions using threshold-based method using CT scans, next using ResNet-50 to extract the features map which further is fed to fully

connected layer to record AUC of 0.732 and 70% accuracy. Ardakani et al. [216] have compared ten DL models for classification of COVID-19, including AlexNet, VGG-16, VGG-19, GoogleNet, MobileNet, Xception, ResNet-101, ResNet-18, ResNet-50, and SqueezNet. Where ResNet-101 has recorded the best results regarding sensitivity. A few used deep learning themes that have been used for different applications of COVID-19 are listed in Tables 1, 2, 3.

4 Deep learning schemes

4.1 Data augmentation

It was clearly that deep learning approach performs better than the traditional machine learning and shallow learning methods and other handcrafted feature extraction from images, because deep learning models learn image descriptors automatically for analysis. It is commonly possible to combine deep learning approach with the knowledge learned from the handcrafted features for analyzing medical images [153, 200, 217]. The main key feature of deep learning is the large-scale datasets which contain images from thousands of patients. Although some vast data of clinical images, reports, and annotations are recorded and stored digitally in many hospitals for example, Picture Archiving and Communication systems (PACS) and Oncology Information System (OIS), in practice, these kinds of large-scale datasets with semantic labels are an efficiency measure for deep learning models used in medical imaging analysis. As it is known that medical images face the lack of dataset, data augmentation has been used to create new samples either depending on the existing samples or using generative models to generate new images. The new augmented samples are emerged with the original samples; thus, the size of the dataset is increased with the variation in the data points. Data augmentation is used by default with deep learning due to its added efficiency, since it reduces the chance of overfitting and it eliminates the imbalanced issue while using multi-class datasets, because it increases the number of the training samples and this also helps in generalizing the models and enhance the testing results. The basic data augmentation techniques are simple and it was widely adopted in medical imaging, such as cropping, rotating, flipping, shearing, scaling, and translation of images [80, 218, 219]. Pezeshk et al. [220] have proposed mixing tool which can seamlessly merge a lesion patch into a CT scan or mammography modality, so the merged lesion patches can be augmented using the basic transformations and inserted to the lesion shape and characteristics.

Zhang et al. [221] have used DCNN for extracting features and obtaining image representations and similarity matrix too, their proposed data augmentation method is called unified learning of features representation, their model was trained on seed-labeled dataset, and authors intended to classify colonoscopy and upper endoscopy medical images. The second method to tackle limited datasets is to synthesize medical data using an object model or physics principles of image formation and using generative models schemes to serve as applicable medical examples and therefore increase the performance of any deep learning task at hand. The most used model for synthesizing medical data is Generative Adversarial Networks (GANs); for example [143], GANs were used to generate lesion samples which increase CNN performance while the classification task of liver lesions. Yang et al. [222] used Radon Transform for objects with different modeled conditions by adding noise to the data for synthesizing CT dataset and the trained CNN model does the estimation of high-dose projection from low-dose. Synthesizing medical images is used for different purposes; for example, Chen et al. [223] have generating training data for noise reduction for reconstructed CT scans by applying deep learning algorithm by synthesizing noisy projections from patient images. While, *CUI* et al. [224] have used simulated dynamic PET data and used stacked sparse autoencoders for dynamic PET reconstruction framework.

4.2 Datasets

Deep learning models are famous to be dataset hungry, and the good quality of dataset has been always the key-parameter for deep learning for learning computational models and provide trusted results. The task of deep learning models is more potential when handling medical data because the accuracy is highly needed, recently many publicly available datasets have been released online for evaluating the new developed DL models. Commonly, there are different repositories which provide useful compilations of the public datasets (e.g., Github, Kaggle, and other webpages). Comparing to the datasets for general computer vision tasks (thousands to million annotated images), medical imaging datasets are considered to be too small. According to the Conference on Machine Intelligence in Medical Imaging (C-MIMI) that was held in 2016 [225], ML and DL are starving for large-scale annotated datasets, and the most common regularities and specifications (e.g., sample size, cataloging and discovery, pixel data, metadata, and post-processing) related to medical images datasets are mentioned in this white paper. Therefore, different trends in medical imaging community have started to adopt different approaches for generating and

Fig. 11 Flowchart of medical images data handling

make deep learning algorithms much efficient and adequate. There are some common algorithms used for medical image features' extractors, such as Gray-Level-Run-Length-Matrix (GLRM), Local Binary Patterns (LBP), Local Tetra Patterns (LTrP), Completed Local Binary Patterns (CLBP), and Gray-Level-Co-Occurrence Matrix (GLCM); these techniques are used in the first place before applying the main DL algorithm for different medical imaging tasks.

GLCM: is a common used feature extractor by which it searches for the textural patterns and their nature within gray-level gradients [234]. The main extracted features through this technique are autocorrelation, contrast, Dissimilarity, correlation, cluster prominence, energy, homogeneity, variance, entropy, difference variance, sum variance, cluster shade, sum entropy, information measure of correlation.

LBP: is another famous feature extractor which uses the locally regional statistical features [235]. The main theme of this technique is to select a central pixel and the rest pixels along a circle are taken to be binary encoded as 0 if their values are less than the central pixel, and 1 for the pixels which have values greater than the central pixel. In histogram statistics, these binary codes are encoded to decimal numbers.

Gray-Level Run Length Matrix (GLRLM): this method removes the higher order statistical texture data. In case of the maximum gray dimensions G, the image is repeatedly re-quantizing to aggregate the network. The mathematical formula of GLRLM is given as follows:

$$K(\theta) = (r(u, v)/\theta, 0 \le u \le N_r, 0 \le v \le K\max, \qquad (4)$$

where (u,v) refers to the sizes of the array values, N_r refers to the maximum gray-level values, and $K_{\max}$ is the more length.

Raj et al. [236] have used both GLCM and GLRLM as the main features' extraction techniques for extracting the optimal features from the pre-processed medical images, which further the optimal features have improved the final results of classification task. Figure 12 shows the features extraction and selection types used for dimensionality reduction.

increasing the number of samples in dataset, such as generative models, data augmentation, and weakly supervised learning, to avoid overfitting on the small dataset and finally provide an end-to-end fashion reliable deep learning model. Martin et al. [226] have described the fundamental steps for preparing the medical imaging datasets for the usage of AI applications. Fig. 11 shows the flowchart of the process; moreover, they have listed the current limitations and problem of data availability of such datasets. Examples of popular used databases for medical images analysis which exploit deep learning were listed in [227]. In this paper, we provide the typically mostly used datasets in the literature of medical imaging which are exploited by deep learning approaches in Table 6.

4.3 Feature's extraction and selection

Feature extraction is the tool of converting training data and trying to establish as maximum features as possible to

Table 6 Public available datasets used for medical imaging

Dataset/references	Anatomical organ	Image modality	No. of images/patients
BRATS [228]	Brain	MRI	–
ADNI [229]	Brain	MRI	819 patients
ChestX-ray14 [129]	Lung	X-ray	30,000 + patients
LIDC-IDRI [230]	Lung	CT	1018 patients
ILD [231]	Lung	CT	120 patients
TCIA	Multiple organs with cancer	CT, MRI, PET	30.9 million images/37,568 patients
DRIVE [232]	Eye	SLO	400 patients
STARE [233]	Eye	SLO	400 images
ISIC 2018	Skin	JPEG	2600 images
SILVER07	Liver	CT	–
COVIDx [169]	Lung	CXR	13,975 CXR images across 13,870

4.3.1 Feature selection techniques

Analysis of Variance—ANOVA: is a statistical model by which it evaluates and compares two or more experiments averages. The idea behind this model is that the difference between means are substantial to evaluate the performance of two estimates [237]. Surendiran et al. [238] have used the stepwise ANOVA Discriminant Analysis (DA) for mammogram masses' classification.

The basic steps of performing ANOVA on data distribution are

1. Defining Hypothesis:
2. Calculating the sum of squares
   It is used to determine the dispersion from datapoints and it can be written as

$$\text{Sum of Squares} = \sum_{i=0} (X_i - \overline{X})^2. \tag{5}$$

ANOVA performs F test to compare the variance difference between groups and within groups. And this can be done using total sum of squares which is defined as the distance between each point from the grand mean x-bar.
3. Determining the degree of freedom
4. Calculating F value
5. Acceptance or rejection of null hypothesis

Principal Component Analysis (PCA): is considered as the most used tool for extracting structural features from potentially high-dimensional datasets. It extracts the eigenvectors (q) which are connected to (q) largest eigenvalues from an input distribution. PCA results develop new features that are independent of another. The main goal of PCA is to apply linear transformation for obtaining a new set of samples, so that the components of y are uncorrelated, [239]. The linear transform is given as follows:

$$y = A^T x, \tag{6}$$

where x is the input element vector $\in R_I$, after that the PCA algorithm will choose the most significant components (y), and the main steps to do this are summarized as follows:

1. Standardize and normalization of the datapoints: after calculating the mean and standard deviation of the input distribution

$$X_{\text{new}} = \frac{X - \text{mean}(X)}{\text{std}(X)}. \tag{7}$$

2. Calculating the covariance matrix from the input datapoints:

$$C[i,j] = \text{cov}(x_i, x_j). \tag{8}$$

3. From the covariance matrix extract the eigenvalues:

$$C = V \sum V^{-1}. \tag{9}$$

4. Choosing k eigenvectors with the highest eigenvalues by sorting the eigenvalues and eigenvectors, k refers to the number of dimensions in the dataset

$$\sum_{\text{sort}} = \text{sort}(\sum) \bigvee_{\text{sort}} = \text{sort}(\bigvee \sum_{\text{sort}}). \tag{10}$$

Another major feature of PCA algorithm is used for feature dimensionality reduction.

In medical imaging, PCA was used mostly for dimensionality reduction, Wu et al. [240] have used PCA-based nearest neighbor for estimation of local structure distribution and extracted the entire connected tree, and in their results over retinal fundus data, they have achieved state-of-the-art results by producing more information regarding the tree structure.

PCA was also used as a data augmentation process before training the discriminative CNN for different medical imaging tasks; for capturing the important characteristics of natural images, different algorithms were compared to perform data augmentation [241] (Fig. 12).
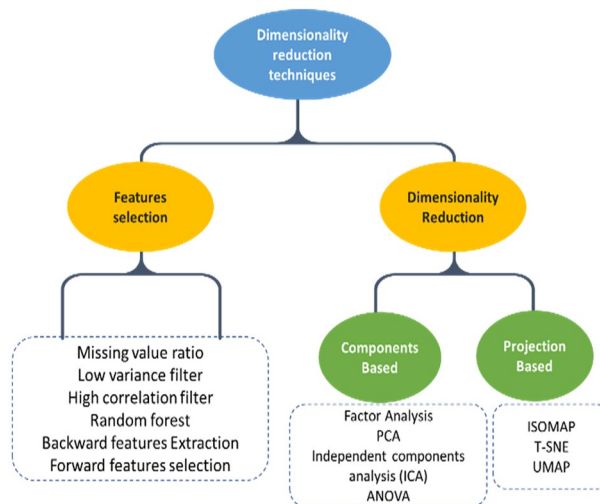

Fig. 12 Features' extraction and selection types used for dimensionality reduction

4.4 Evaluation metrics

For the purpose of evaluating and measuring the performance of deep learning models while validating medical images, different evaluation metrics are used according to some specific regularities and criteria. For example, some particular evaluation metrics are used with specific tasks like Dice score and $F1$-score are mostly used for segmentation, while accuracy and sensitivity are mostly used for classification task. Here, we will focus on the most used performance measurement metrics in the literature and will cover the metrics mentioned in our tables of comparison.

1. **The Dice coefficient** is the most used metric for segmentation task for validating the medical images. It is common also to use dice score to measure reproducibility [242]. The general formula to calculate the Dice coefficient is

$$\text{Dice} = \frac{2\left|S_g^1 \cap S_t^1\right|}{\left|S_g^1\right| + \left|S_t^1\right|} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}. \quad (11)$$

2. **Jaccard-index (similarity coefficient) [JAC]:**
   Jaccard-index is a statistic metric used for finding the similarities between sample-sets. It is defined as the ratio between the size of intersection and the size of union of the sample-set, and the mathematical formula is given by

$$\text{JAC} = \frac{\left|S_g^1 \cap S_t^1\right|}{\left|S_g^1\right| + \left|S_t^1\right|} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}. \quad (12)$$

   From the formula above, we note that the JAC-index is always greater than the dice score and the relation between the two metrics is defined by

$$\text{JAC} = \frac{\left|S_g^1 \cap S_t^1\right|}{\left|S_g^1\right| + \left|S_t^1\right|} = \frac{2\left|S_g^1 \cap S_t^1\right|}{2\left(\left|S_g^1\right| + \left|S_t^1\right| - \left|S_g^1 \cap S_t^1\right|\right)} = \frac{\text{DICE}}{2 - \text{DICE}}. \quad (13)$$

3. **True-Positive Rate (TPR):**
   Also is called as Sensitivity and Recall, is used to maximize the prediction of a particular class and it measures the portion of the positive voxels from the ground truth which also were identified as positive when performing segmentation process. It is given by the formula

$$\text{TPR} = \text{Sensitivity} = \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (14)$$

4. **True-Negative Rate (TNR):**
   Also called specificity, it measures the number of negative voxels (background) from the ground truth which are also identified to be negative after the segmentation process, and it is given by the formula

$$\text{TNR} = \text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}. \quad (15)$$

   However, both TNR and TPR metrics are not used commonly for medical images' segmentation due to their sensibility to the segments size.

5. **Accuracy [ACC]**
   Accuracy means exactly how good the DL model at guessing the right labels (ground truth). Accuracy is commonly used to validate the classification task and it is given using the formula

$$\text{ACC} = \frac{\text{no. of correct predictions}}{\text{total number of predictions}} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \quad (16)$$

6. **F1-Score:**
   It is used to get the best precision and recall together; thus, the $F1$-score is called the harmonic mean of precision and recall values; it is given by the formula

$$F_1 = \left(\frac{\text{recall}^{-1} + \text{Precision}^{-1}}{2}\right)^{-1} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}. \quad (17)$$

   The predictive accuracy of a classification model is related to the $F1$-score; when $F1$-score is higher means, we have better classification accuracy.

7. **F-beta score:**
   It is a combination of advantages of precision and recall metrics when both the False-Negative (FN) and False-Positive (FP) have equal importance. F-beta-score is given using the same formula for $F1$-score by altering its formula a bit by including an adjustable parameter (beta), and the formula became

$$F_\beta = \left(1 + \beta^2\right) \cdot \frac{\text{precision} \cdot \text{recall}}{\left(\beta^2 \cdot \text{precision}\right) + \text{recall}}. \quad (18)$$

   This evaluation metric measures the effectiveness of a DL model according to a user who attaches beta times.

8. **AUC-ROC:**
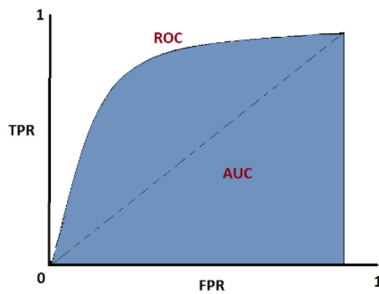   Receiver-Operating Characteristics Curve (ROC) is a graph between True-Positive Rate (TPR) (sensitiv-

Fig. 13 ROC and AUC graph

ity) and False-Positive Rate (FPR) (1- specificity) by which it shows the performance of classification model, and the plot is characterized at different classification thresholds. The biggest advantage of ROC curve is that its independency of the change in number of responders and response rate.

AUC is the area under curve of ROC, and it measures the 2D area under the ROC curve which in turn means the integral of the ROC curve from 0 to the AUC measures the aggregate performance of classification at all the possible thresholds. One way to understand the AUC is as the probability that a model classifies random positive samples more than random negative samples. The ROC curve is shown in Fig. 13.

5 Discussion and conclusion

5.1 Technical challenges

In this overview paper, we have presented a review of the previous literature of deep learning applications in medical imaging. It contributed three main sections: first, we have presented the core of deep learning concepts considering the main highlights of understanding of basic frameworks in medical images analysis. The second section contains the main applications of deep learning in medical imaging (e.g., segmentation, detection, classification, and registration) and we have presented a comprehensive review of the literature. The criteria that we have built our overview consists of the mostly cited papers, the mostly recent (from 2015 to 2021), and the papers with better results. The third major part of this paper is focused on the deep learning themes regarding some challenges and the future directions of addressing those challenges. Besides focusing on the quality of the mostly recent works, we have highlighted the suitable solutions for different challenges in this field and the future directions that have been concluded from different scientific perspective. Medical imaging can get the benefit from other fields of deep learning, that have been encouraged

from collaborative research works from computer vision communities, and furthermore, this collaboration is used to overcome the lack of medical dataset using transfer learning. Cho et al. [243] have answered the question of how much is the size of medical dataset needed to train a deep learning model. Creating synthetic medical images is another solution presented in deep learning using Variational Autoencoders (VAEs) and GANs for tackling the limited labeled medical data. For instance, Guibas et al. [244] have used 2 GANs for segmenting and then generating new retinal fundus images successfully. Another applications of GANs for segmentation and synthetic data generation were found [132, 245].

Data or class imbalance [246] is considered a critical problem in medical imaging, and it refers to that medical images that are used for training are skewed toward non-pathological images; rare diseases have less number of training examples which cause the problem of imbalanced data which lead to incorrect results. Data augmentation represents good solution for this, because it increases the number of samples of the small classes. Away from dataset challenges strategies, there are algorithmic modification strategies which are used to improve DL models' performance for data imbalance issue [247].

Another important non-technical challenge is the public reception of humans that the results are being analyzed using DL models (not human). In some papers in our report, DL models have outperformed specialists in medicine (e.g., dermatologists and radiologists) and mostly in image recognition tasks. Yet, a moral culpability may arise whenever a patient is mistakenly diagnosed or morbidity cases may arise too when using DL-based diagnostic, since the work of a DL algorithms is considered a black box. However, the continued development and evolving of DL models might take a major role in the medicine as it is involving in various facets of our life.

AI systems have started to emerge in hospitals from a clinical perspective. Bruijne [248] have presented five challenges facing the broader family of deep learning which is machine learning in medical imaging field including the data preprocessing of different modalities, interpretation of results to clinical practice, improving the access of medical data, and training the models with little training data. These challenges further have addressed the future directions of DL models improvement. Another solutions of small datasets were reported in [8, 249].

DL models' architectures were found not to be the only factor that provides quality results, where data augmentation and preprocessing techniques are also substantial tools for a robust and efficient system. The big question is that how to benefit from the results of DL models for the best of medical images analysis in the community.

Considering the historical developments of ML techniques in medical imaging gives us a future perspective

how DL models will continue to improve in the same field. Accordingly, medical images quality and data annotations are crucial for proper analysis. A significant concept is the relevance between statistical sense and clinical sense, even though the statistical analysis are quiet important in research, but in this field, researchers should not lose the sight from clinical perspective; in other words, even when a good CNN models provides good answers from the statistical perspective, it does not mean that it will replace a radiologist even after using all the helping techniques like data augmentation and adding more layers to get better accuracies.

5.2 Future promises

After reviewing literature and the most competitive challenges that face deep leaning in medical imaging, we concluded that three aspects will probably carry the revolution of DL according to most of researchers, which are the availability of large-scale datasets, advances in deep learning algorithms, and the computational power for data processing and evaluation of DL models. Thus, most of the DL techniques are directed into the above aspects for alleviating the DL performance more; moreover, the need for investigations to improve data harmonization, standards developments which is needed for reporting and evaluation, and accessibility of larger annotated data such as the public datasets which lead to better independent benchmarks' services. Some of the interesting applications in medical imaging was proposed by Nie et al. [250], by which they have used GANs to generate or CT scans from MRI images for brain; the benefit of such work will reduce the risk of patients being exposed to ionizing radiation from CT scanners, which also reserve patients' safety. Another significant perspective relies on increasing the resolution and quality of medical images and also reduces the blurriness from CT scans and MRI images which means getting higher resolution with lower costs and better results, because it has lower field strength [251].

The new trends' technology of deep learning approach is concerned about medical data collection. Wearable technologies are getting the interest of the new research which provide the benefits of flexibility, real-time monitoring of patients, and the immediate communication of the collected data. Whenever the data become available, Deep learning and AI will start to use the unsupervised data exploration, which in turn will provide better analysis power plus suggesting better treatments' methodologies in healthcare. In summary, the new trends of AI in healthcare pass through stages; the quality of performance (QoP) related to deep learning will lead to standardization in terms of wearable technology which represent the next stage of healthcare applications and personalized treatment. Diagnosing and treatment depend on specialists, but with deep learning enabled, some small changes and signs in human body can be seen and early detection becomes possible which in turn will launch the treatment process of pre-stage of diseases. DL model optimization mainly focuses on the network architecture, while the standard term of optimization means the distribution and standardization with respect to other parts of DL (e.g., optimizers, loss functions, preprocessing and post-processing, etc.). In many cases to achieve better diagnosis, medical images are not sufficient, where another data are required to be combined (e.g., historical medical reports, genetic information, lab values, and other non-image data), though by linking and normalizing these data with medical images will lead to better diagnosis of diseases, more accurately through analyzing these data in higher dimensions.

Declarations

References

1. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553), 436–444 (2015). https://doi.org/10.1038/nature14539
2. Akhtar, N., Mian, A.: Threat of adversarial attacks on deep learning in computer vision: a survey. IEEE Access **6**, 14410–14430 (2018). https://doi.org/10.1109/ACCESS.2018.2807385
3. Sutskever, I., Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks. Adv. Neural Inf. Process. Syst. **4**(January), 3104–3112 (2014)
4. Smith-Bindman, R., et al.: Use of diagnostic imaging studies and associated radiation exposure for patients enrolled in large integrated health care systems, 1996–2010. JAMA **307**(22), 2400–2409 (2012). https://doi.org/10.1001/jama.2012.5960
5. Rubin, D.L.: Measuring and improving quality in radiology: meeting the challenge with informatics. Radiographics **31**(6), 1511–1527 (2011). https://doi.org/10.1148/rg.316105207
6. Recht, M.P., et al.: Integrating artificial intelligence into the clinical practice of radiology: challenges and recommendations. Eur. Radiol. **30**(6), 3576–3584 (2020). https://doi.org/10.1007/s00330-020-06672-5
7. Bosma, M., van Beuzekom, M., Vähänen, S., Visser, J., Koffeman, E.: The influence of edge effects on the detection properties of Cadmium Telluride. In: 2011 IEEE Nuclear Science Symposium Conference Record IEEE, pp. 4812–4817 (2011)
8. Litjens, G., et al.: A survey on deep learning in medical image analysis. Med. Image Anal. **42**(December 2012), 60–88 (2017). https://doi.org/10.1016/j.media.2017.07.005
9. LeCun, Y., et al.: Backpropagation applied to handwritten zip code recognition. Neural Comput. **1**(4), 541–551 (1989). https://doi.org/10.1162/neco.1989.1.4.541
10. Lecun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. IEEE **86**(11), 2278–2324 (1998). https://doi.org/10.1109/5.726791
11. Sutskever, I., Martens, J. and Hinton, G.: Generating text with recurrent neural networks. In: Proc. 28th Int. Conf. Mach. Learn. ICML 2011, pp. 1017–1024 (2011)

12. Balkanski, E., Rubinstein, A. and Singer, Y.: The power of optimization from samples. In: Advances in Neural Information Processing Systems, 2016, vol. 29. Available: https://proceedings.neurips.cc/paper/2016/file/c8758b517083196f05ac29810b924aca-Paper.pdf

13. Karpathy, A and Fei-Fei, L.: Deep Visual-Semantic Alignments for Generating Image Descriptions - Karpathy_Deep_Visual-Semantic_Alignments_2015_CVPR_paper.pdf. Cvpr (2015)

14. Shin, H.C., Roberts, K., Lu, L., Demner-Fushman, D., Yao, J., Summers, R.M.: Learning to read chest x-rays: recurrent neural cascade model for automated image annotation. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. **2016**, 2497–2506 (2016). https://doi.org/10.1109/CVPR.2016.274

15. Cui, R., Liu, M.: RNN-based longitudinal analysis for diagnosis of Alzheimer's disease. Comput. Med. Imaging Graph. **73**, 1–10 (2019). https://doi.org/10.1016/j.compmedimag.2019.01.005

16. Zhang, J., Zuo, H.: A deep RNN for CT image reconstruction. Proc. SPIE (2020). https://doi.org/10.1117/12.2549809

17. Qin, C., et al.: Joint learning of motion estimation and segmentation for cardiac MR image sequences, vol. 11071. Springer International Publishing, LNCS (2018)

18. Ben-Cohen, A., Mechrez, R., Yedidia, N. and Greenspan, H.: Improving CNN training using disentanglement for liver lesion classification in CT. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2019, pp. 886–889, https://doi.org/10.1109/EMBC.2019.8857465

19. Liao, H., Lin, W.-A., Zhou, S.K., Luo, J.: ADN: artifact disentanglement network for unsupervised metal artifact reduction. IEEE Trans. Med. Imaging **39**(3), 634–643 (2020). https://doi.org/10.1109/TMI.2019.2933425

20. Qin, C., Shi, B., Liao, R., Mansi, T., Rueckert, D., Kamen, A.: Unsupervised deformable registration for multi-modal images via disentangled representations, vol. 11492. Springer International Publishing, LNCS (2019)

21. Creswell, A., Bharath, A. A.: Denoising adversarial autoencoders. IEEE transactions on neural networks and learning systems, **30**(4), 968–984 (2018)

22. Lopez Pinaya, W.H., Vieira, S., Garcia-Dias, R., Mechelli, A.: Autoencoders. Elsevier Inc. (2019)

23. Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P.A.: Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. J. Mach. Learn. Res. **11**, 3371–3408 (2010)

24. Ranzato, M.A., Poultney, C., Chopra, S., LeCun, Y.: Efficient learning of sparse representations with an energy-based model. Adv. Neural Inf. Process. Syst. (2007). https://doi.org/10.7551/mitpress/7503.003.0147

25. Kingma, D.P. and Welling, M.: Auto-encoding variational bayes. In: 2nd International Conference of Learning Representation. ICLR 2014 - Conf. Track Proc., no. Ml, pp. 1–14 (2014)

26. Rifai, S., Vincent, P., Muller, X., Glorot, X. and Bengio, Y.: Contractive auto-encoders: explicit invariance during feature extraction. In: Proceeding 28th International Conference of Machine Learning. ICML 2011, no. 1, pp. 833–840 (2011)

27. Li, C., Xu, K., Zhu, J. and Zhang, B.: Triple generative adversarial nets. In: Advances in Neural Information Processing Systems, vol. 2017-Decem, pp. 4089–4099 (2017)

28. Goodfellow, I. et al.: Generative adversarial nets. In: Advances in Neural Information Processing Systems, 2014, vol. 27. Available: https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf

29. Shrivastava, A., Pfister, T., Tuzel, O., Susskind, J., Wang, W. and Webb, R.: Learning from simulated and unsupervised images through adversarial training. In: Proceeding - 30th IEEE Conference of Computer Vision Pattern Recognition, CVPR 2017, vol. 2017-Janua, pp. 2242–2251, (2017) https://doi.org/10.1109/CVPR.2017.241

30. Hacihaliloglu, J.R.B.I., Singer, E.A., Foran, D.J.: For Classification of Prostate Histopathology Whole-Slide Images, vol. 1. Springer International Publishing, Berlin (2018)

31. Bi, X., Li, S., Xiao, B., Li, Y., Wang, G., Ma, X.: Computer aided Alzheimer's disease diagnosis by an unsupervised deep learning technology. Neurocomputing **392**, 296–304 (2020). https://doi.org/10.1016/j.neucom.2018.11.111

32. Baumgartner, C.F., Koch, L.M., Tezcan, K.C., Ang, J.X., Konukoglu, E.: Visual feature attribution using wasserstein GANs. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (2018). https://doi.org/10.1109/CVPR.2018.00867

33. Son, J., Park, S.J. and Jung, K.-H.: Retinal Vessel Segmentation in Fundoscopic Images with Generative Adversarial Networks (2017). Available: http://arxiv.org/abs/1706.09318

34. Dou, Q., et al.: PnP-AdaNet: plug-and-play adversarial domain adaptation network with a benchmark at cross-modality cardiac segmentation. CoRR, vol. abs/1812.0, (2018). Available: http://arxiv.org/abs/1812.07907

35. Welander, P., Karlsson, S. and Eklund, A.: Generative adversarial networks for image-to-image translation on multi-contrast {MR} images - {A} comparison of CycleGAN and {UNIT}. *CoRR* vol. abs/1806.0, (2018). Available: http://arxiv.org/abs/1806.07777

36. Kazeminia, S., et al.: GANs for medical image analysis. Artif. Intell. Med. **109**, 101938 (2020). https://doi.org/10.1016/j.artmed.2020.101938

37. Ackley, D.H., Hinton, G.E., Sejnowski, T.J.: A learning algorithm for boltzmann machines. Cogn. Sci. **9**(1), 147–169 (1985). https://doi.org/10.1016/S0364-0213(85)80012-4

38. Paul, S.: Information processing in dynamical systems: foundations of harmony theory. J. Jpn. Soc. Fuzzy Theory Syst. **4**(2), 220–228 (1986)

39. van Tulder, G., de Bruijne, M.: Combining generative and discriminative representation learning for lung CT analysis with convolutional restricted boltzmann machines. IEEE Trans. Med. Imaging **35**(5), 1262–1272 (2016). https://doi.org/10.1109/TMI.2016.2526687

40. Hinton, G.E., Osindero, S., Teh, Y.-W.: A fast learning algorithm for deep belief nets. Neural Comput. **18**(7), 1527–1554 (2006). https://doi.org/10.1162/neco.2006.18.7.1527

41. Khatami, A., Khosravi, A., Nguyen, T., Lim, C.P., Nahavandi, S.: Medical image analysis using wavelet transform and deep belief networks. Expert Syst. Appl. **86**, 190–198 (2017). https://doi.org/10.1016/j.eswa.2017.05.073

42. Reddy, A.V.N., et al.: Analyzing MRI scans to detect glioblastoma tumor using hybrid deep belief networks. J. Big Data. (2020). https://doi.org/10.1186/s40537-020-00311-y

43. Kaur, M., Singh, D.: Fusion of medical images using deep belief networks. Cluster Comput. **23**(2), 1439–1453 (2020). https://doi.org/10.1007/s10586-019-02999-x

44. Zhou, Z., et al.: Models genesis: generic autodidactic models for 3D medical image analysis. In: Medical Image Computing and Computer Assisted Intervention -- MICCAI 2019, pp. 384–393 (2019)

45. Zhu, J., Li, Y., Hu, Y., Ma, K., Zhou, S.K., Zheng, Y.: Rubik's Cube+: a self-supervised feature learning framework for 3D medical image analysis. Med. Image Anal. **64**, 101746 (2020). https://doi.org/10.1016/j.media.2020.101746

46. Azizi, S., et al.: Big self-supervised models advance medical image classification. no. 1, (2021). Available: http://arxiv.org/abs/2101.05224

47. Nie, D., Gao, Y., Wang, L. and Shen, D.: ASDNet: attention based semi-supervised deep networks for medical image segmentation. In: Medical Image Computing and Computer Assisted Intervention -- MICCAI 2018, pp. 370–378 (2018)

48. Bai, W., et al.: Semi-supervised learning for network-based cardiac MR image segmentation. In: Medical Image Computing and Computer-Assisted Intervention − MICCAI 2017, pp. 253–260 (2017)

49. Liu, Q., Yu, L., Luo, L., Dou, Q., Heng, P.A.: Semi-supervised medical image classification with relation-driven self-ensembling model. IEEE Trans. Med. Imaging 39(11), 3429–3440 (2020). https://doi.org/10.1109/TMI.2020.2995518

50. Kervadec, H., Dolz, J., Tang, M., Granger, E., Boykov, Y., Ben Ayed, I.: Constrained-CNN losses for weakly supervised segmentation. Med. Image Anal. 54, 88–99 (2019). https://doi.org/10.1016/j.media.2019.02.009

51. Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., Summers, R.M.: ChestX-ray: hospital-scale chest x-ray database and benchmarks on weakly supervised classification and localization of common thorax diseases. Adv. Comput. Vis. Pattern Recognit. (2019). https://doi.org/10.1007/978-3-030-13969-8_18

52. Shi, G., Xiao, L., Chen, Y., Zhou, S.K.: Marginal loss and exclusion loss for partially supervised multi-organ segmentation. Med. Image Anal. 70, 101979 (2021). https://doi.org/10.1016/j.media.2021.101979

53. Roth, H.R., Yang, D., Xu, Z., Wang, X., Xu, D.: Going to extremes: weakly supervised medical image segmentation. Mach. Learn. Knowl. Extr. 3(2), 507–524 (2021). https://doi.org/10.3390/make3020026

54. Schlegl, T., Seeböck, P., Waldstein, S.M., Langs, G., Schmidt-Erfurth, U.: f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks. Med. Image Anal. 54, 30–44 (2019). https://doi.org/10.1016/j.media.2019.01.010

55. Hu, Y., et al.: Weakly-supervised convolutional neural networks for multimodal image registration. Med. Image Anal. 49, 1–13 (2018). https://doi.org/10.1016/j.media.2018.07.002

56. Quellec, G., Laniard, M., Cazuguel, G., Abràmoff, M.D., Cochener, B. and Roux, C.: Weakly supervised classification of medical images. In: 2012 9th IEEE International Symposium on Biomedical Imaging (ISBI), pp. 110–113 (2012) doi: https://doi.org/10.1109/ISBI.2012.6235496

57. Abdullah Al, W., Yun, I.D.: Partial policy-based reinforcement learning for anatomical landmark localization in 3D medical images. IEEE Trans. Med. Imaging 39(4), 1245–1255 (2020). https://doi.org/10.1109/TMI.2019.2946345

58. Smith, R.L., Ackerley, I.M., Wells, K., Bartley, L., Paisey, S. and Marshall, C.: Reinforcement learning for object detection in PET imaging. In: 2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), pp. 1–4 (2019). doi: https://doi.org/10.1109/NSS/MIC42101.2019.9060031

59. Park, J., Jo, S., Lee, J., and Sun, W.: Color image classification on neuromorphic system using reinforcement learning. In: 2020 International Conference on Electronics, Information, and Communication (ICEIC), pp. 1–2 (2020). doi: https://doi.org/10.1109/ICEIC49074.2020.9051310

60. Pan, S.J., Yang, Q.: A survey on transfer learning. IEEE Trans. Knowl. Data Eng. 22(10), 1345–1359 (2010). https://doi.org/10.1109/TKDE.2009.191

61. Shin, H.C., et al.: Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans. Med. Imaging 35(5), 1285–1298 (2016). https://doi.org/10.1109/TMI.2016.2528162

62. Xue, D., et al.: An application of transfer learning and ensemble learning techniques for cervical histopathology image classification. IEEE Access 8, 104603–104618 (2020). https://doi.org/10.1109/ACCESS.2020.2999816

63. Tajbakhsh, N., et al.: Convolutional neural networks for medical image analysis: full training or fine tuning? IEEE Trans. Med.

Imaging 35(5), 1299–1312 (2016). https://doi.org/10.1109/TMI.2016.2535302

64. Krizhevsky, B.A., Sutskever, I., Hinton, G.E.: Cnn实际训练的. Commun. ACM 60(6), 84–90 (2012)

65. Lu, S., Lu, Z., Zhang, Y.D.: Pathological brain detection based on AlexNet and transfer learning. J. Comput. Sci. 30, 41–47 (2019). https://doi.org/10.1016/j.jocs.2018.11.008

66. Simonyan, K. and Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: 3rd International Conference of Learning Representation. ICLR 2015 - Conf. Track Proc., pp. 1–14 (2015)

67. Sahiner, B., et al.: Deep learning in medical imaging and radiation therapy. Med. Phys. 46(1), e1–e36 (2019). https://doi.org/10.1002/mp.13264

68. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z.: Rethinking the inception architecture for computer vision. In: Proceeding of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2016-Decem, pp. 2818–2826 (2016). https://doi.org/10.1109/CVPR.2016.308

69. Szegedy, C., Ioffe, S., Vanhoucke, V. and Alemi, A.A.: Inception-v4, inception-ResNet and the impact of residual connections on learning. In: 31st AAAI Conference of Artificial Intelligence. AAAI 2017, pp. 4278–4284 (2017)

70. Gao, F., Wu, T., Chu, X., Yoon, H., Xu, Y., Patel, B.: Deep residual inception encoder–decoder network for medical imaging synthesis. IEEE J. Biomed. Heal. Informatics 24(1), 39–49 (2020). https://doi.org/10.1109/JBHI.2019.2912659

71. A. {Szegedy, Christian and Liu, Wei and Jia, Yangqing and Sermanet, Pierre and Reed, Scott and Anguelov, Dragomir and Erhan, Dumitru and Vanhoucke, Vincent and Rabinovich, "{Going Deeper With Convolutions}e," 2015, [Online]. Available: Szegedy, Christian, et al. %22Going deeper with convolutions.%22 Proceedings of the IEEE conference on computer vision and pattern recognition. 2015

72. He, K., Zhang, X.,S. Ren, S. and Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition., vol. 2016-Decem, pp. 770–778 (2016). https://doi.org/10.1109/CVPR.2016.90

73. Wang, Q., Shen, F., Shen, L., Huang, J., Sheng, W.: Lung nodule detection in CT images using a raw patch-based convolutional neural network. J. Digit. Imaging 32(6), 971–979 (2019). https://doi.org/10.1007/s10278-019-00221-3

74. Mantas, J., Hasman, A., Househ, M.S., Gallos, P., Zoulias, E.: Preface. Stud. Health Technol. Inform. 272, v (2020). https://doi.org/10.3233/SHTI272

75. Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N. and Liang, J.: UNet++: A nested U-net architecture for medical image segmentation. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. pp. 3–11 (2018)

76. Mohammed Senan, E., Waselallah Alsaade, F., Ibrahim Ahmed Al-mashhadani, M., Aldhyani, T.H.H., Hmoudal-Adhaileh, M.: Classification of histopathological images for early detection of breast cancer using deep learning. J. Appl. Sci. Eng. 24(3), 323–329 (2021). https://doi.org/10.6180/jase.202106_24(3).0007

77. Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K.Q.: Densely connected convolutional networks. In: Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017, vol. 2017-Janua. pp. 2261–2269 (2017). https://doi.org/10.1109/CVPR.2017.243

78. Mahmood, F., Yang, Z., Ashley, T. and Durr, N.J.: Multimodal Densenet (2018). Available: http://arxiv.org/abs/1811.07407

79. Xu, X., Lin, J., Tao, Y. and Wang, X.: An improved DenseNet method based on transfer learning for fundus medical images. In:

2018 7th International Conference on Digital Home (ICDH). pp. 137–140 (2018). https://doi.org/10.1109/ICDH.2018.00033

80. Ronneberger, O., Fischer, P. and Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2015. pp. 234–241 (2015)

81. Forouzanfar, M., Forghani, N., Teshnehlab, M.: Parameter optimization of improved fuzzy c-means clustering algorithm for brain MR image segmentation. Eng. Appl. Artif. Intell. **23**(2), 160–168 (2010). https://doi.org/10.1016/j.engappai.2009.10.002

82. Akkus, Z., Galimzianova, A., Hoogi, A., Rubin, D.L., Erickson, B.J.: Deep learning for brain MRI segmentation: state of the art and future directions. J. Digit. Imaging **30**(4), 449–459 (2017). https://doi.org/10.1007/s10278-017-9983-4

83. Milletari, F., Navab, N. and Ahmadi, S.-A.: V-Net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV). pp. 565–571 (2016). https://doi.org/10.1109/3DV.2016.79

84. Havaei, M., et al.: Brain tumor segmentation with Deep Neural Networks. Med. Image Anal. **35**, 18–31 (2017). https://doi.org/10.1016/j.media.2016.05.004

85. Asgari Taghanaki, S., Abhishek, K., Cohen, J.P., Cohen-Adad, J., Hamarneh, G.: Deep semantic segmentation of natural and medical images: a review. Artif. Intell. Rev. **54**(1), 137–178 (2021). (**Springer Netherlands**)

86. Li, W., Jia, F., Hu, Q.: Automatic segmentation of liver tumor in CT images with deep convolutional neural networks. J. Comput. Commun. **03**(11), 146–151 (2015). https://doi.org/10.4236/jcc.2015.311023

87. Dong, H., Yang, G., Liu, F., Mo, Y. and Guo Y.: Automatic brain tumor detection and segmentation using U-net based fully convolutional networks. In: Medical Image Understanding and Analysis, pp. 506–517 (2017)

88. Soltaninejad, M., Zhang, L., Lambrou, T., Allinson, N. and Ye, X.: Multimodal MRI brain tumor segmentation using random forests with features learned from fully convolutional neural network. (2017). Available: http://arxiv.org/abs/1704.08134

89. Ker, J., Wang, L., Rao, J., Lim, T.: Deep learning applications in medical image analysis. IEEE Access **6**, 9375–9379 (2017). https://doi.org/10.1109/ACCESS.2017.2788044

90. Chen, L., Bentley, P., Rueckert, D.: Fully automatic acute ischemic lesion segmentation in DWI using convolutional neural networks. NeuroImage Clin. **15**(January), 633–643 (2017). https://doi.org/10.1016/j.nicl.2017.06.016

91. Li, Z., Wang, Y. and Yu, J.: Brain tumor segmentation using an adversarial network. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. pp. 123–132 (2018)

92. Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., Jambawalikar, S. R.: Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. (2018). arXiv:1811.02629

93. Isensee, F., Kickingereder, P., Wick, W., Bendszus, M. and Maier-Hein, K. H.: Brain tumor segmentation and radiomics survival prediction: contribution to the BRATS 2017 challenge. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, pp. 287–297 (2018)

94. Korfiatis, P., Kline, T.L., Erickson, B.J.: Automated segmentation of hyperintense regions in FLAIR MRI using deep learning. Tomography **2**(4), 334–340 (2016)

95. Liu, M., Zhang, J., Adeli, E., Shen, D.: Landmark-based deep multi-instance learning for brain disease diagnosis. Med. Image Anal. **43**, 157–168 (2018). https://doi.org/10.1016/j.media.2017.10.005

96. Liskowski, P., Krawiec, K.: Segmenting retinal blood vessels with deep neural networks. IEEE Trans. Med. Imaging **35**(11), 2369–2380 (2016). https://doi.org/10.1109/TMI.2016.2546227

97. Fang, L., Cunefare, D., Wang, C., Guymer, R.H., Li, S., Farsiu, S.: Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and graph search. Biomed. Opt. Express **8**(5), 2732–2744 (2017). https://doi.org/10.1364/BOE.8.002732

98. Shankaranarayana, S.M., Ram, K., Mitra, K. and Sivaprakasam, M.: Joint optic disc and cup segmentation using fully convolutional and adversarial networks. In: Fetal, Infant and Ophthalmic Medical Image Analysis. pp. 168–176 (2017)

99. Fu, H., Cheng, J., Xu, Y., Wong, D.W.K., Liu, J., Cao, X.: Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. IEEE Trans. Med. Imaging **37**(7), 1597–1605 (2018). https://doi.org/10.1109/TMI.2018.2791488

100. Hu, P., Wu, F., Peng, J., Liang, P., Kong, D.: Automatic 3D liver segmentation based on deep learning and globally optimized surface evolution. Phys. Med. Biol. **61**(24), 8676–8698 (2016). https://doi.org/10.1088/1361-6560/61/24/8676

101. Ben-Cohen, A., Diamant, I., Klang, E., Amitai, M. and Greenspan, H.: Fully convolutional network for liver segmentation and lesions detection. In: Deep Learning and Data Labeling for Medical Applications, pp. 77–85 (2016)

102. Yang, D., et al.: Automatic liver segmentation using an adversarial image-to-image network. In: Medical Image Computing and Computer Assisted Intervention − MICCAI 2017. pp. 507–515 (2017)

103. Cheng, J.Z., et al.: Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and pulmonary nodules in CT scans. Sci. Rep. **6**(March), 1–13 (2016). https://doi.org/10.1038/srep24454

104. Al-antari, M.A., Al-masni, M.A., Choi, M.T., Han, S.M., Kim, T.S.: A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification. Int. J. Med. Inform. **117**(April), 44–54 (2018). https://doi.org/10.1016/j.ijmedinf.2018.06.003

105. Ait Skourt, B., El Hassani, A., Majda, A.: Lung CT image segmentation using deep neural networks. Procedia Comput. Sci. **127**, 109–113 (2018). https://doi.org/10.1016/j.procs.2018.01.104

106. Kalinovsky, A. and Kovalev, V.: Lung image segmentation using deep learning methods and convolutional neural networks. Int. Conf. Pattern Recognit. Inf. Process., no. July 2017, 21–24 (2016)

107. Roy, S., et al.: Deep learning for classification and localization of COVID-19 markers in point-of-care lung ultrasound. IEEE Trans. Med. Imaging **39**(8), 2676–2687 (2020). https://doi.org/10.1109/TMI.2020.2994459

108. Murphy, K., et al.: COVID-19 on chest radiographs: a multi-reader evaluation of an artificial intelligence system. Radiology **296**(3), E166–E172 (2020). https://doi.org/10.1148/radiol.2020201874

109. Kline, T.L., et al.: Performance of an artificial multi-observer deep neural network for fully automated segmentation of polycystic kidneys. J. Digit. Imaging **30**(4), 442–448 (2017). https://doi.org/10.1007/s10278-017-9978-1

110. Ma, J., Wu, F., Jiang, T., Zhao, Q., Kong, D.: Ultrasound image-based thyroid nodule automatic segmentation using convolutional neural networks. Int. J. Comput. Assist. Radiol. Surg. **12**(11), 1895–1910 (2017). https://doi.org/10.1007/s11548-017-1649-7

111. Zhang, R., Huang, L., Xia, W., Zhang, B., Qiu, B., Gao, X.: Multiple supervised residual network for osteosarcoma segmentation in CT images. Comput. Med. Imaging Graph. **63**(January), 1–8 (2018). https://doi.org/10.1016/j.compmedimag.2018.01.006

112. Yu, L., Guo, Y., Wang, Y., Yu, J., Chen, P.: Segmentation of fetal left ventricle in echocardiographic sequences based on dynamic convolutional neural networks. IEEE Trans. Biomed. Eng. **64**(8), 1886–1895 (2017). https://doi.org/10.1109/TBME.2016.2628401

113. Jafari, M.H., Nasr-Esfahani, E., Karimi, N., Soroushmehr, S.M.R., Samavi, S., Najarian, K.: Extraction of skin lesions from non-dermoscopic images for surgical excision of melanoma. Int. J. Comput. Assist. Radiol. Surg. **12**(6), 1021–1030 (2017). https://doi.org/10.1007/s11548-017-1567-8

114. Yang, D., Zhang, S., Yan, Z., Tan, C., Li, K. and Metaxas, D.: Automated anatomical landmark detection ondistal femur surface using convolutional neural network. In: 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI). pp. 17–21 (2015). doi: https://doi.org/10.1109/ISBI.2015.7163806

115. Orlando, J.I., Prokofyeva, E., del Fresno, M., Blaschko, M.B.: An ensemble deep learning based approach for red lesion detection in fundus images. Comput. Methods Programs Biomed. **153**, 115–127 (2018). https://doi.org/10.1016/j.cmpb.2017.10.017

116. Yang, X., et al.: Co-trained convolutional neural networks for automated detection of prostate cancer in multi-parametric MRI. Med. Image Anal. **42**, 212–227 (2017). https://doi.org/10.1016/j.media.2017.08.006

117. Dou, Q., et al.: Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks. IEEE Trans. Med. Imaging **35**(5), 1182–1195 (2016). https://doi.org/10.1109/TMI.2016.2528129

118. Yan, K., Wang, X., Lu, L., Summers, R.M.: DeepLesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning. J. Med. Imaging **5**(3), 1–11 (2018). https://doi.org/10.1117/1.JMI.5.3.036501

119. Zhang, J., Liu, M., Shen, D.: Detecting anatomical landmarks from limited medical imaging data using two-stage task-oriented deep neural networks. IEEE Trans. Image Process. **26**(10), 4753–4764 (2017). https://doi.org/10.1109/TIP.2017.2721106

120. Nakao, T., et al.: Deep neural network-based computer-assisted detection of cerebral aneurysms in MR angiography. J. Magn. Reson. Imaging **47**(4), 948–953 (2018). https://doi.org/10.1002/jmri.25842

121. Tsehay, Y., et al.: Biopsy-guided learning with deep convolutional neural networks for prostate cancer detection on multiparametric MRI Imaging Biomarkers and Computer-Aided Diagnosis Laboratory, Department of Radiology and Imaging Science, National Institute of Health, C. In: 2017 IEEE 14th Int. Symp. Biomed. Imaging (ISBI 2017), pp. 642–645 (2017)

122. Sirinukunwattana, K., Raza, S.E.A., Tsang, Y.W., Snead, D.R.J., Cree, I.A., Rajpoot, N.M.: Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. IEEE Trans. Med. Imaging **35**(5), 1196–1206 (2016). https://doi.org/10.1109/TMI.2016.2525803

123. Setio, A.A.A., et al.: Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. IEEE Trans. Med. Imaging **35**(5), 1160–1169 (2016). https://doi.org/10.1109/TMI.2016.2536809

124. Li, L., Qin, L., Xu, Z., Yin, Y., Wang, X., Kong, B., Bai, J., Lu, Y., Fang, Z., Song, Q., Cao, K.: Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT. Radiology **296**(2), E65–E71 (2020)

125. Luz, E., et al.: Towards an effective and efficient deep learning model for COVID-19 patterns detection in X-ray images. Res. Biomed. Eng. (2021). https://doi.org/10.1007/s42600-021-00151-6

126. Kassania, S.H., Kassanib, P.H., Wesolowskic, M.J., Schneidera, K.A., Detersa, R.: Automatic detection of coronavirus disease (COVID-19) in X-ray and CT images: a machine learning based approach. Biocybern. Biomed. Eng. **41**(3), 867–879 (2021). https://doi.org/10.1016/j.bbe.2021.05.013

127. Wang, D., Khosla, A., Gargeya, R., Irshad, H. and Beck, A.H.: Deep learning for identifying metastatic breast cancer. pp. 1–6 (2016). Available: http://arxiv.org/abs/1606.05718

128. Dou, Q., Chen, H., Yu, L., Qin, J., Heng, P.A.: Multilevel contextual 3-D CNNs for false positive reduction in pulmonary nodule detection. IEEE Trans. Biomed. Eng. **64**(7), 1558–1567 (2017). https://doi.org/10.1109/TBME.2016.2613502

129. Rajpurkar, P., et al.: CheXNet: radiologist-level pneumonia detection on chest x-rays with deep learning. pp. 3–9, (2017). Available: http://arxiv.org/abs/1711.05225

130. Ma, J., Wu, F., Jiang, T., Zhu, J., Kong, D.: Cascade convolutional neural networks for automatic detection of thyroid nodules in ultrasound images. Med. Phys. **44**(5), 1678–1691 (2017). https://doi.org/10.1002/mp.12134

131. Baka, N., Leenstra, S., Van Walsum, T.: Ultrasound aided vertebral level localization for lumbar surgery. IEEE Trans. Med. Imaging **36**(10), 2138–2147 (2017). https://doi.org/10.1109/TMI.2017.2738612

132. Alex, V., Safwan, P.K.M., Chennamsetty, S.S., Krishnamurthi, G.: Generative adversarial networks for brain lesion detection. Med. Imaging 2017 Image Process. (2017). https://doi.org/10.1117/12.2254487

133. Bogunovic, H., et al.: RETOUCH: the retinal OCT fluid detection and segmentation benchmark and challenge. IEEE Trans. Med. Imaging **38**(8), 1858–1874 (2019). https://doi.org/10.1109/TMI.2019.2901398

134. Gulshan, V., et al.: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. J. Am. Med. Assoc. **316**(22), 2402–2410 (2016). https://doi.org/10.1001/jama.2016.17216

135. Nasrullah, N., Sang, J., Alam, M.S., Mateen, M., Cai, B., Hu, H.: Automated lung nodule detection and classification using deep learning combined with multiple strategies. Sensors (2019). https://doi.org/10.3390/s19173722

136. Domingues, I., Cardoso, J. S.: Mass detection on mammogram images: a first assessment of deep learning techniques. In: 19th Portuguese Conference on Pattern Recognition (RECPAD) (2013)

137. Lai, Z., Deng, H.: Medical image classification based on deep features extracted by deep model and statistic feature fusion with multilayer perceptron. Comput. Intell. Neurosci. (2018). https://doi.org/10.1155/2018/2061516

138. Xiao, B., et al.: PAM-DenseNet: a deep convolutional neural network for computer-aided COVID-19 diagnosis. IEEE Trans. Cybern. (2021). https://doi.org/10.1109/TCYB.2020.3042837

139. Lo, S.-C.B., Lou, S.-L.A., Lin, J.-S., Freedman, M.T., Chien, M.V., Mun, S.K.: Artificial convolution neural network techniques and applications for lung nodule detection. IEEE Trans. Med. Imaging **14**(4), 711–718 (1995). https://doi.org/10.1109/42.476112

140. World Health Organization: Standardization of interpretation of chest radiographs for the diagnosis of pneumonia in children / World Health Organization Pneumonia Vaccine Trial Investigators' Group (2001). Available: http://www.who.int/iris/handle/10665/66956

141. Ding, J., Chen, B., Liu, H., Huang, M.: Convolutional neural network with data augmentation for SAR target recognition. IEEE Geosci. Remote Sens. Lett. **13**(3), 364–368 (2016). https://doi.org/10.1109/LGRS.2015.2513754

142. Perez, L. and Wang, J.: The effectiveness of data augmentation in image classification using deep learning. CoRR, vol. abs/1712.0 (2017). Available: http://arxiv.org/abs/1712.04621

143. 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 **321**, 321–331 (2018). https://doi.org/10.1016/j.neucom.2018.09.013

144. Li, R., et al.: Deep learning based imaging data completion for improved brain disease diagnosis. In: Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2014. pp. 305–312 (2014)

145. Hosseini-Asl, E., Gimel'farb, G.L. and El-Baz, A.: Alzheimer's disease diagnostics by a deeply supervised adaptable 3D convolutional network. CoRR vol. abs/1607.0 (2016). Available: http://arxiv.org/abs/1607.00556

146. Abràmoff, M.D., et al.: Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. Invest. Ophthalmol. Vis. Sci. **57**(13), 5200–5206 (2016). https://doi.org/10.1167/iovs.16-19964

147. Anthimopoulos, M., Christodoulidis, S., Ebner, L., Christe, A., Mougiakakou, S.: Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. IEEE Trans. Med. Imaging **35**(5), 1207–1216 (2016). https://doi.org/10.1109/TMI.2016.2535865

148. Nibali, A., He, Z., Wollersheim, D.: Pulmonary nodule classification with deep residual networks. Int. J. Comput. Assist. Radiol. Surg. **12**(10), 1799–1808 (2017). https://doi.org/10.1007/s11548-017-1605-6

149. Christodoulidis, S., Anthimopoulos, M., Ebner, L., Christe, A., Mougiakakou, S.: Multisource transfer learning with convolutional neural networks for lung pattern analysis. IEEE J. Biomed. Heal. Informatics **21**(1), 76–84 (2017). https://doi.org/10.1109/JBHI.2016.2636929

150. Wu, X., et al.: Deep learning-based multi-view fusion model for screening 2019 novel coronavirus pneumonia: a multicentre study. Eur. J. Radiol. **128**(March), 1–9 (2020). https://doi.org/10.1016/j.ejrad.2020.109041

151. Farid, A.A., Selim, G.I., Khater, H.A.A.: A novel approach of CT images feature analysis and prediction to screen for corona virus disease (COVID-19). Int. J. Sci. Eng. Res. **11**(03), 1141–1149 (2020). https://doi.org/10.14299/ijser.2020.03.02

152. Pereira, R.M., Bertolini, D., Teixeira, L.O., Silla, C.N., Costa, Y.M.G.: COVID-19 identification in chest X-ray images on flat and hierarchical classification scenarios. Comput. Methods Programs Biomed. **194**, 105532 (2020). https://doi.org/10.1016/j.cmpb.2020.105532

153. Huynh, B.Q., Li, H., Giger, M.L.: Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. J. Med. Imaging **3**(3), 034501 (2016). https://doi.org/10.1117/1.jmi.3.3.034501

154. Sun, W., Tseng, T.L.B., Zhang, J., Qian, W.: Enhancing deep convolutional neural network scheme for breast cancer diagnosis with unlabeled data. Comput. Med. Imaging Graph. **57**, 4–9 (2017). https://doi.org/10.1016/j.compmedimag.2016.07.004

155. Swati, Z.N.K., et al.: Brain tumor classification for MR images using transfer learning and fine-tuning. Comput. Med. Imaging Graph. **75**, 34–46 (2019). https://doi.org/10.1016/j.compmedimag.2019.05.001

156. Sajjad, M., Khan, S., Muhammad, K., Wu, W., Ullah, A., Baik, S.W.: Multi-grade brain tumor classification using deep CNN with extensive data augmentation. J. Comput. Sci. **30**, 174–182 (2019). https://doi.org/10.1016/j.jocs.2018.12.003

157. Deepak, S., Ameer, P.M.: Brain tumor classification using deep CNN features via transfer learning. Comput. Biol. Med. **111**(June), 103345 (2019). https://doi.org/10.1016/j.compbiomed.2019.103345

158. Afshar, P., Mohammadi, A. and Plataniotis, K. N.: Brain tumor type classification via capsule networks. In: 2018 25th IEEE International Conference on Image Processing (ICIP), pp. 3129–3133 (2018). https://doi.org/10.1109/ICIP.2018.8451379

159. Gao, X.W., Hui, R., Tian, Z.: Classification of CT brain images based on deep learning networks. Comput. Methods Programs Biomed. **138**, 49–56 (2017). https://doi.org/10.1016/j.cmpb.2016.10.007

160. Bharati, S., Podder, P., Mondal, M.R.H.: Hybrid deep learning for detecting lung diseases from X-ray images. Inform. Med. Unlocked **20**, 100391 (2020). https://doi.org/10.1016/j.imu.2020.100391

161. Zhou, J., et al.: Weakly supervised 3D deep learning for breast cancer classification and localization of the lesions in MR images. J. Magn. Reson. Imaging **50**(4), 1144–1151 (2019). https://doi.org/10.1002/jmri.26721

162. Zhang, Q., et al.: Deep learning based classification of breast tumors with shear-wave elastography. Ultrasonics **72**, 150–157 (2016). https://doi.org/10.1016/j.ultras.2016.08.004

163. Yang, C., Rangarajan, A., Ranka, S.: Visual explanations from deep 3D Convolutional Neural Networks for Alzheimer's disease classification. AMIA … Annu. Symp. proceedings. AMIA Symp. **2018**, 1571–1580 (2018)

164. Schwyzer, M., et al.: Automated detection of lung cancer at ultralow dose PET/CT by deep neural networks—initial results. Lung Cancer **126**(November), 170–173 (2018). https://doi.org/10.1016/j.lungcan.2018.11.001

165. de Carvalho Filho, A.O., Silva, A.C., de Paiva, A.C., Nunes, R.A., Gattass, M.: Classification of patterns of benignity and malignancy based on CT using topology-based phylogenetic diversity index and convolutional neural network. Pattern Recognit. **81**, 200–212 (2018). https://doi.org/10.1016/j.patcog.2018.03.032

166. Shen, W., et al.: Multi-crop Convolutional Neural Networks for lung nodule malignancy suspiciousness classification. Pattern Recognit. **61**, 663–673 (2017). https://doi.org/10.1016/j.patcog.2016.05.029

167. Ozturk, T., Talo, M., Yildirim, E.A., Baloglu, U.B., Yildirim, O., Rajendra Acharya, U.: Automated detection of COVID-19 cases using deep neural networks with X-ray images. Comput. Biol. Med. **121**(April), 103792 (2020). https://doi.org/10.1016/j.compbiomed.2020.103792

168. Ucar, F., Korkmaz, D.: COVIDiagnosis-Net: Deep Bayes-SqueezeNet based diagnosis of the coronavirus disease 2019 (COVID-19) from X-ray images. Med. Hypotheses **140**(April), 109761 (2020). https://doi.org/10.1016/j.mehy.2020.109761

169. Wang, L., Lin, Z.Q., Wong, A.: COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. Sci. Rep. **10**(1), 1–12 (2020). https://doi.org/10.1038/s41598-020-76550-z

170. Rezvantalab, A., Safigholi, H. and Karimijeshni, S.: Dermatologist level dermoscopy skin cancer classification using different deep learning convolutional neural networks algorithms. (2018). Available: http://arxiv.org/abs/1810.10348

171. Dorj, U.O., Lee, K.K., Choi, J.Y., Lee, M.: The skin cancer classification using deep convolutional neural network. Multimed. Tools Appl. **77**(8), 9909–9924 (2018). https://doi.org/10.1007/s11042-018-5714-1

172. Esteva, A., et al.: Dermatologist-level classification of skin cancer with deep neural networks. Nature **542**(7639), 115–118 (2017). https://doi.org/10.1038/nature21056

173. Awais, M., Muller, H., Tang, T. B. and Meriaudeau, F.: Classification of SD-OCT images using a deep learning approach. In: Proc. 2017 IEEE Int. Conf. Signal Image Process. Appl. ICSIPA 2017, vol. c, pp. 489–492 (2017). https://doi.org/10.1109/ICSIPA.2017.8120661

174. Ting, D.S.W., et al.: Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with

diabetes. J. Am. Med. Assoc. **318**(22), 2211–2223 (2017). https://doi.org/10.1001/jama.2017.18152

175. Frid-Adar, M., Klang, E., Amitai, M., Goldberger, J. and Greenspan, H.: Synthetic data augmentation using GAN for improved liver lesion classification. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), pp. 289–293 (2018). https://doi.org/10.1109/ISBI.2018.8363576

176. Krebs, J., et al.: Robust non-rigid registration through agent-based action learning. In: Medical Image Computing and Computer Assisted Intervention − MICCAI 2017. pp. 344–352 (2017)

177. Yang, X., Kwitt, R., Styner, M., Niethammer, M.: Quicksilver: fast predictive image registration—a deep learning approach. Neuroimage **158**, 378–396 (2017). https://doi.org/10.1016/j.neuroimage.2017.07.008

178. Sokooti, H., de Vos, B., Berendsen, F., Lelieveldt, B. P. F., Išgum, I. and Staring, M.: Nonrigid image registration using multi-scale 3D Convolutional Neural Networks. In: Medical Image Computing and Computer Assisted Intervention − MICCAI 2017, pp. 232–239 (2017)

179. Wu, G., Kim, M., Wang, Q., Gao, Y., Liao, S. and Shen, D.: Unsupervised deep feature learning for deformable registration of MR brain images. In: Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2013, pp. 649–656 (2013)

180. Wu, G., Kim, M., Wang, Q., Munsell, B.C., Shen, D.: Scalable high-performance image registration framework by unsupervised deep feature representations learning. IEEE Trans. Biomed. Eng. **63**(7), 1505–1516 (2016). https://doi.org/10.1109/TBME.2015.2496253

181. Miao, S., Wang, Z.J., Liao, R.: A CNN regression approach for real-time 2D/3D registration. IEEE Trans. Med. Imaging **35**(5), 1352–1363 (2016). https://doi.org/10.1109/TMI.2016.2521800

182. de Vos, B.D., Berendsen, F. F., Viergever, M. A., Staring, M., Išgum, I.: End-to-end unsupervised deformable image registration with a convolutional neural network. In: Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 10553 LNCS, pp. 204–212 (2017). doi: https://doi.org/10.1007/978-3-319-67558-9_24

183. Sun, L., Zhang, S.: Deformable MRI-ultrasound registration using 3D convolutional neural network, vol. 11042. Springer International Publishing, LNCS (2018)

184. Chen, Y., He, F., Li, H., Zhang, D., Wu, Y.: A full migration BBO algorithm with enhanced population quality bounds for multimodal biomedical image registration. Appl. Soft Comput. J. **93**, 106335 (2020). https://doi.org/10.1016/j.asoc.2020.106335

185. Niethammer, M., Kwitt, R., Vialard, F.X.: Metric learning for image registration. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. **2019**, 8455–8464 (2019). https://doi.org/10.1109/CVPR.2019.00866

186. Wang, S., Kim, M., Wu, G., Shen, D.: Scalable high performance image registration framework by unsupervised deep feature representations learning. Deep Learn. Med. Image Anal. **63**(7), 245–269 (2017). https://doi.org/10.1016/B978-0-12-810408-8.00015-8

187. Kang, E., Min, J., Ye, J.C.: A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction. Med. Phys. **44**(10), e360–e375 (2017). https://doi.org/10.1002/mp.12344

188. Abanoviè, E., Stankevièius, G. and Matuzevièius, D.: Deep Neural Network-based feature descriptor for retinal image registration. In: 2018 IEEE 6th Workshop on Advances in Information, Electronic and Electrical Engineering (AIEEE), pp. 1–4 (2018)https://doi.org/10.1109/AIEEE.2018.8592033

189. Haskins, G., et al.: Learning deep similarity metric for 3D MR–TRUS image registration. Int. J. Comput. Assist. Radiol. Surg. **14**(3), 417–425 (2019). https://doi.org/10.1007/s11548-018-1875-7

190. Giger, M.L., Chan, H.-P., Boone, J.: Anniversary paper: history and status of CAD and quantitative image analysis: the role of medical physics and AAPM. Med. Phys. **35**(12), 5799–5820 (2008). https://doi.org/10.1118/1.3013555

191. Giger, M.L., Karssemeijer, N., Schnabel, J.A.: Breast image analysis for risk assessment, detection, diagnosis, and treatment of cancer. Annu. Rev. Biomed. Eng. **15**(1), 327–357 (2013). https://doi.org/10.1146/annurev-bioeng-071812-152416

192. Li, H., et al.: Quantitative MRI radiomics in the prediction of molecular classifications of breast cancer subtypes in the TCGA/TCIA data set. NPJ Breast Cancer **2**(1), 16012 (2016). https://doi.org/10.1038/npjbcancer.2016.12

193. Guo, W., et al.: Prediction of clinical phenotypes in invasive breast carcinomas from the integration of radiomics and genomics data. J. Med. Imaging **2**(4), 1–12 (2015). https://doi.org/10.1117/1.JMI.2.4.041007

194. Li, H., et al.: MR imaging radiomics signatures for predicting the risk of breast cancer recurrence as given by research versions of MammaPrint, Oncotype DX, and PAM50 Gene assays. Radiology **281**(2), 382–391 (2016). https://doi.org/10.1148/radiol.2016152110

195. Katsuragawa, S., Doi, K., MacMahon, H., Monnier-Cholley, L., Ishida, T., Kobayashi, T.: Classification of normal and abnormal lungs with interstitial diseases by rule-based method and artificial neural networks. J. Digit. Imaging **10**(3), 108–114 (1997). https://doi.org/10.1007/BF03168597

196. Kim, G.B., et al.: Comparison of shallow and deep learning methods on classifying the regional pattern of diffuse lung disease. J. Digit. Imaging **31**(4), 415–424 (2018). https://doi.org/10.1007/s10278-017-0028-9

197. Antropova, N.O., Abe, H., Giger, M.L.: Use of clinical MRI maximum intensity projections for improved breast lesion classification with deep convolutional neural networks. J. Med. Imaging **5**(1), 1–6 (2018). https://doi.org/10.1117/1.JMI.5.1.014503

198. Mohamed, A.A., Berg, W.A., Peng, H., Luo, Y., Jankowitz, R.C., Wu, S.: A deep learning method for classifying mammographic breast density categories. Med. Phys. **45**(1), 314–321 (2018). https://doi.org/10.1002/mp.12683

199. Lee, J., Nishikawa, R.M.: Automated mammographic breast density estimation using a fully convolutional network. Med. Phys. **45**(3), 1178–1190 (2018). https://doi.org/10.1002/mp.12763

200. Antropova, N., Huynh, B.Q., Giger, M.L.: A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets. Med. Phys. **44**(10), 5162–5171 (2017). https://doi.org/10.1002/mp.12453

201. Samala, R.K., Chan, H.-P., Hadjiiski, L.M., Helvie, M.A., Cha, K.H., Richter, C.D.: Multi-task transfer learning deep convolutional neural network: application to computer-aided diagnosis of breast cancer on mammograms. Phys. Med. Biol. **62**(23), 8894–8908 (2017). https://doi.org/10.1088/1361-6560/aa93d4

202. Kooi, T., van Ginneken, B., Karssemeijer, N., den Heeten, A.: Discriminating solitary cysts from soft tissue lesions in mammography using a pretrained deep convolutional neural network. Med. Phys. **44**(3), 1017–1027 (2017). https://doi.org/10.1002/mp.12110

203. Masood, A., et al.: Computer-Assisted Decision Support System in Pulmonary Cancer detection and stage classification on CT images. J. Biomed. Inform. **79**(January), 117–128 (2018). https://doi.org/10.1016/j.jbi.2018.01.005

204. Gonzalez, G., et al.: Disease staging and prognosis in smokers using deep learning in chest computed tomography. Am. J.

Respir. Crit. Care Med. **197**(2), 193–203 (2018). https://doi.org/10.1164/rccm.201705-0860OC

205. Lao, J., et al.: A deep learning-based radiomics model for prediction of survival in Glioblastoma Multiforme. Sci. Rep. **7**(1), 1–8 (2017). https://doi.org/10.1038/s41598-017-10649-8

206. Garapati, S.S., et al.: Urinary bladder cancer staging in CT urography using machine learning. Med. Phys. **44**(11), 5814–5823 (2017). https://doi.org/10.1002/mp.12510

207. Takahashi, H., Tampo, H., Arai, Y., Inoue, Y., Kawashima, H.: Applying artificial intelligence to disease staging: deep learning for improved staging of diabetic retinopathy. PLoS One **12**(6), 1–11 (2017). https://doi.org/10.1371/journal.pone.0179790

208. Skrede, O.-J., et al.: Deep learning for prediction of colorectal cancer outcome: a discovery and validation study. Lancet **395**(10221), 350–360 (2020). https://doi.org/10.1016/S0140-6736(19)32998-8

209. Saillard, C., et al.: Predicting survival after hepatocellular carcinoma resection using deep learning on histological slides. Hepatology **72**(6), 2000–2013 (2020). https://doi.org/10.1002/hep.31207

210. "No Title," https://www.who.int/news-room/fact-sheets/detail. 2019, [Online]. Available: https://www.who.int/news-room/fact-sheets/detail

211. Simpson, S., et al.: Radiological society of North America expert consensus document on reporting chest CT findings related to COVID-19: endorsed by the society of thoracic radiology, the American College of Radiology, and RSNA. Radiol. Cardiothorac. Imaging **2**(2), e200152 (2020). https://doi.org/10.1148/ryct.2020200152

212. Mahmood, A., Gajula, C., Gajula, P.: COVID 19 diagnostic tests: a study of 12,270 patients to determine which test offers the most beneficial results. Surg. Sci. **11**(04), 82–88 (2020). https://doi.org/10.4236/ss.2020.114011

213. Soldati, G., et al.: Is there a role for lung ultrasound during the COVID-19 pandemic? J. Ultrasound Med. **39**(7), 1459–1462 (2020). https://doi.org/10.1002/jum.15284

214. Butt, C., Gill, J., Chun, D. et al.: RETRACTED ARTICLE: Deep learning system to screen coronavirus disease 2019 pneumonia. Appl Intell (2020). https://doi.org/10.1007/s10489-020-01714-3

215. Wang, B., Wu, Z., Khan, Z. U., Liu, C. and Zhu, M.: Deep convolutional neural network with segmentation techniques for chest x-ray analysis. In: 2019 14th IEEE Conference on Industrial Electronics and Applications (ICIEA). pp. 1212–1216 (2019). https://doi.org/10.1109/ICIEA.2019.8834117

216. Ardakani, A.A., Kanafi, A.R., Acharya, U.R., Khadem, N., Mohammadi, A.: Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: results of 10 convolutional neural networks. Comput. Biol. Med. **121**, 103795 (2020). https://doi.org/10.1016/j.compbiomed.2020.103795

217. Li, H., Giger, M.L., Huynh, B.Q., Antropova, N.O.: Deep learning in breast cancer risk assessment: evaluation of convolutional neural networks on a clinical dataset of full-field digital mammograms. J. Med. Imaging **4**(4), 1–6 (2017). https://doi.org/10.1117/1.JMI.4.4.041304

218. Roth, H.R., et al.: DeepOrgan: multi-level deep convolutional networks for automated pancreas segmentation. In: Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2015. pp. 556–564 (2015)

219. Asperti, A. and Mastronardo, C.: The effectiveness of data augmentation for detection of gastrointestinal diseases from endoscopical images. CoRR, vol. abs/1712.0 (2017). Available: http://arxiv.org/abs/1712.03689

220. Pezeshk, A., Petrick, N., Chen, W., Sahiner, B.: Seamless lesion insertion for data augmentation in CAD training. IEEE Trans.

Med. Imaging **36**(4), 1005–1015 (2017). https://doi.org/10.1109/TMI.2016.2640180

221. Zhang, C., Tavanapong, W., Wong, J., de Groen, P.C. and Oh, J.: Real data augmentation for medical image classification. In: Intravascular Imaging and Computer Assisted Stenting, and Large-Scale Annotation of Biomedical Data and Expert Label Synthesis. pp. 67–76 (2017)

222. Yang, X., et al.: Low-dose x-ray tomography through a deep convolutional neural network. Sci. Rep. **8**(1), 2575 (2018). https://doi.org/10.1038/s41598-018-19426-7

223. Chen, H., et al.: Low-dose CT via convolutional neural network. Biomed. Opt. Express **8**(2), 679–694 (2017). https://doi.org/10.1364/BOE.8.000679

224. Cui, J., Liu, X., Wang, Y., Liu, H.: Deep reconstruction model for dynamic PET images. PLoS One **12**(9), 1–21 (2017). https://doi.org/10.1371/journal.pone.0184667

225. Kohli, M.D., Summers, R.M., Geis, J.R.: Medical image data and datasets in the era of machine learning—whitepaper from the 2016 C-MIMI meeting dataset session. J. Digit. Imaging **30**(4), 392–399 (2017). https://doi.org/10.1007/s10278-017-9976-3

226. Willemink, M.J., et al.: Preparing medical imaging data for machine learning. Radiology **295**(1), 4–15 (2020). https://doi.org/10.1148/radiol.2020192224

227. Altaf, F., Islam, S.M.S., Akhtar, N., Janjua, N.K.: Going deep in medical image analysis: concepts, methods, challenges, and future directions. IEEE Access **7**(3), 99540–99572 (2019). https://doi.org/10.1109/ACCESS.2019.2929365

228. Menze, B.H., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans. Med. Imaging **34**(10), 1993–2024 (2015). https://doi.org/10.1109/TMI.2014.2377694

229. Petersen, R.C., et al.: Alzheimer's disease neuroimaging initiative (ADNI). Neurology **74**(3), 201–209 (2010). https://doi.org/10.1212/WNL.0b013e3181cb3e25

230. Armato, S.G., et al.: Lung image database consortium: developing a resource for the medical imaging research community. Radiology **232**(3), 739–748 (2004). https://doi.org/10.1148/radiol.2323032035

231. Depeursinge, A., Vargas, A., Platon, A., Geissbuhler, A., Poletti, P.-A., Müller, H.: Building a reference multimedia database for interstitial lung diseases. Comput. Med. Imaging Graph. **36**(3), 227–238 (2012). https://doi.org/10.1016/j.compmedimag.2011.07.003

232. Staal, J., Abràmoff, M.D., Niemeijer, M., Viergever, M.A., Van Ginneken, B.: Ridge-based vessel segmentation in color images of the retina. IEEE Trans. Med. Imaging **23**(4), 501–509 (2004). https://doi.org/10.1109/TMI.2004.825627

233. Hoover, A.D., Kouznetsova, V., Goldbaum, M.: Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. IEEE Trans. Med. Imaging **19**(3), 203–210 (2000). https://doi.org/10.1109/42.845178

234. Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural features for image classification. IEEE Trans. Syst. Man. Cybern. **SMC-3**(6), 610–621 (1973). https://doi.org/10.1109/TSMC.1973.4309314

235. Çamlica, Z., Tizhoosh, H. R. and Khalvati, F.: Medical image classification via SVM using LBP features from saliency-based folded data. In: 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA). pp. 128–132 (2015). https://doi.org/10.1109/ICMLA.2015.131

236. Raj, R.J.S., Shobana, S.J., Pustokhina, I.V., Pustokhin, D.A., Gupta, D., Shankar, K.: Optimal feature selection-based medical image classification using deep learning model in internet of medical things. IEEE Access **8**, 58006–58017 (2020). https://doi.org/10.1109/ACCESS.2020.2981337

237. Miller Jr, R. G.: Beyond ANOVA: basics of applied statistics. CRC press (1997)

238. Surendiran, A., Vadivel, B.: Feature selection using stepwise ANOVA discriminant analysis for mammogram mass classification. Int. J. Signal Image Process. **2**(1), 17 (2011). https://www.researchgate.net/profile/Surendiran_Balasubramanian/publication/258052973_Feature_selection_using_stepwise_ANOVA_discriminant_analysis_for_mammogram_mass_classification/links/0c96052942c3e97cda000000.pdf

239. Theodoridis, D.: Sergios and Pikrakis, Aggelos and Koutroumbas, Konstantinos and Cavouras, *Introduction to pattern recognition: a matlab approach*. 2010

240. Wu, A., Xu, Z., Gao, M., Mollura, D.J.: Deep vessel tracking: a generalized probabilistic approach via deep learning Aaron Wu, Ziyue Xu?, Mingchen Gao, Mario Buty, Daniel J. Mollura Department of Radiology and Imaging Sciences, National Institutes of Health, Bethesda, MD 20892. Isbi **2016**, 1363–1367 (2016)

241. Hussain, Z., Gimenez, F., Yi, D., Rubin, D.: Differential data augmentation techniques for medical imaging classification tasks. AMIA … Annu. Symp. Proc. AMIA Symp. **2017**, 979–984 (2017)

242. Zou, K.H., et al.: Statistical validation of image segmentation quality based on a spatial overlap index1: scientific reports. Acad. Radiol. **11**(2), 178–189 (2004). https://doi.org/10.1016/S1076-6332(03)00671-8

243. Cho, J., Lee, K., Shin, E., Choy, G. and Do, S.: Medical image deep learning with hospital {PACS} dataset. CoRR vol. abs/1511.0, (2015). Available: http://arxiv.org/abs/1511.06348

244. Guibas, J. T., Virdi, T. S. and Li, P. S.: Synthetic medical images from dual generative adversarial networks. CoRR vol. abs/1709.0, (2017). Available: http://arxiv.org/abs/1709.01872

245. Moeskops, P., Veta, M., Lafarge, M. W., Eppenhof, K. A. J. and Pluim, J. P. W.: Adversarial training and dilated convolutions for brain MRI segmentation. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. pp. 56–64 (2017)

246. Mazurowski, M.A., Habas, P.A., Zurada, J.M., Lo, J.Y., Baker, J.A., Tourassi, G.D.: Training neural network classifiers for medical decision making: the effects of imbalanced datasets on classification performance. Neural Netw. **21**(2), 427–436 (2008). https://doi.org/10.1016/j.neunet.2007.12.031

247. Galar, M., Fernandez, A., Barrenechea, E., Bustince, H. and Herrera, F.: A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. **42**(4), 463–484 (2012). https://doi.org/10.1109/TSMCC.2011.2161285

248. de Bruijne, M.: Machine learning approaches in medical image analysis: from detection to diagnosis. Med. Image Anal. **33**, 94–97 (2016). https://doi.org/10.1016/j.media.2016.06.032

249. Lundervold, A.S., Lundervold, A.: An overview of deep learning in medical imaging focusing on MRI. Z. Med. Phys. **29**(2), 102–127 (2019). https://doi.org/10.1016/j.zemedi.2018.11.002

250. Nie, D., Cao, X., Gao, Y., Wang, L. and Shen, D.: Estimating CT image from MRI data using 3D fully convolutional networks. In: Deep Learning and Data Labeling for Medical Applications, pp. 170–178 (2016)

251. Ledig, C., et al.: "\href{https://ieeexplore.ieee.org/abstract/document/8099502}{Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network}," *Cvpr*, vol. 2, no. 3, p. 4, 2017. Available: http://openaccess.thecvf.com/content_cvpr_2017/papers/Ledig_Photo-Realistic_Single_Image_CVPR_2017_paper.pdf

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.