INVITED PAPER

A Review of Deep Learning in Medical Imaging: Imaging Traits, Technology Trends, Case Studies With Progress Highlights, and Future Promises

In this article, the authors highlight both clinical needs and technical challenges in medical imaging and describe how emerging trends in deep learning are addressing these issues.

By S. Kevin Zhou, *Fellow IEEE*, Hayit Greenspan, Christos Davatzikos, *Fellow IEEE*, James S. Duncan, *Life Fellow IEEE*, Bram van Ginneken, Anant Madabhushi, *Fellow IEEE*, Jerry L. Prince, *Fellow IEEE*, Daniel Rueckert, *Fellow IEEE*, and Ronald M. Summers

Manuscript received August 1, 2020; revised December 12, 2020; accepted January 13, 2021. Date of publication February 26, 2021; date of current version April 30, 2021. The work of Anant Madabhushi was supported in part by the National Institutes of Health under Award 1U24CA199374-01, Award R01CA202752-01A1, Award R01CA208236-01A1, Award R01CA216579-01A1, Award R01CA220581-01A1, Award 1U01CA239055-01, Award 1U54CA254566-01, Award 1U01CA248226-01, and Award 1R43EB028736-01 and in part by the VA Merit Review Award IBX004121A from the Biomedical Laboratory Research and Development Service of the United States Department of Veterans Affairs. The work of Hayit Greenspan was supported in part by the Israeli Science Foundation (ISF) and in part by the Ministry of Science & Technology. The work of Ronald M. Summers was supported by the National Institutes of Health Clinical Center.
(S. Kevin Zhou and Hayit Greenspan contributed equally to this work.)
(Corresponding author: S. Kevin Zhou.)

S. Kevin Zhou is with the School of Biomedical Engineering, University of Science and Technology of China, Hefei 230052, China, and also with the Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China (e-mail: zhoushaohua@ict.ac.cn).
Hayit Greenspan is with the Department of Biomedical Engineering, Faculty of Engineering, Tel-Aviv University, Tel-Aviv 69978, Israel.
Christos Davatzikos is with the Radiology Department, University of Pennsylvania, Philadelphia, PA 19104 USA, and also with the Electrical and Systems Engineering Department, University of Pennsylvania, Philadelphia, PA 19104 USA.
James S. Duncan is with the Department of Biomedical Engineering, Yale University, New Haven, CT 06520 USA, and also with the Department of Radiology and Biomedical Imaging, Yale University, New Haven, CT 06520 USA.
Bram van Ginneken is with the Radboud University Medical Center, 6525 GA Nijmegen, The Netherlands.
Anant Madabhushi is with the Department of Biomedical Engineering, Case Western Reserve University, Cleveland, OH 44106 USA, and also with the Louis Stokes Cleveland Veterans Administration Medical Center, Cleveland, OH 44106 USA.
Jerry L. Prince is with the Electrical and Computer Engineering Department, Johns Hopkins University, Baltimore, MD 21218 USA.
Daniel Rueckert is with the Klinikum rechts der Isar, Technical University of Munich (TU Munich), 81675 Munich, Germany, and also with the Department of Computing, Imperial College London, London SW7 2AZ, U.K.
Ronald M. Summers is with the National Institutes of Health Clinical Center, Bethesda, MD 20892 USA.

Digital Object Identifier 10.1109/JPROC.2021.3054390

ABSTRACT | Since its renaissance, deep learning (DL) has been widely used in various medical imaging tasks and has achieved remarkable success in many medical imaging applications, thereby propelling us into the so-called artificial intelligence (AI) era. It is known that the success of AI is mostly attributed to the availability of big data with annotations for a single task and the advances in high-performance computing. However, medical imaging presents unique challenges that confront DL approaches. In this survey article, we first present traits of medical imaging, highlight both clinical needs and technical challenges in medical imaging, and describe how emerging trends in DL are addressing these issues. We cover the topics of network architecture, sparse and noisy labels, federating learning, interpretability, uncertainty quantification, and so on. Then, we present several case studies that are commonly found in clinical practice, including digital pathology and chest, brain, cardiovascular, and abdominal imaging. Rather than presenting an exhaustive literature survey, we instead describe some prominent research highlights related to these case study applications. We conclude with a discussion and presentation of promising future directions.

KEYWORDS | Deep learning (DL); medical imaging; survey.

I. Overview

Medical imaging [1] exploits physical phenomena, such as light, electromagnetic radiation, radioactivity, nuclear magnetic resonance (MR), and sound to generate visual representations or images of external or internal tissues of

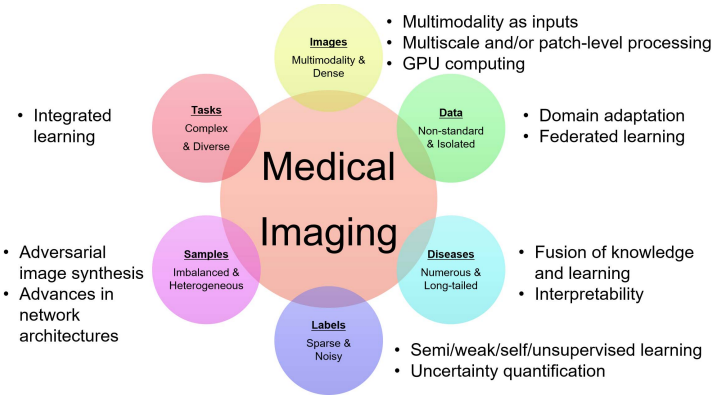

Fig. 1. ***Main traits of medical imaging and the associated technological trends for addressing these traits.***

the human body or a part of the human body in a noninvasive manner or via an invasive procedure. The most commonly used imaging modalities in clinical medicine include X-ray radiography, computed tomography (CT), MR imaging (MRI), ultrasound, and digital pathology. Imaging data account for about 90% of all healthcare data[1] and, hence, is one of the most important sources of evidence for clinical analysis and medical intervention.

A. Traits of Medical Imaging

As described in the following and illustrated in Fig. 1, medical imaging has several traits that influence the suitability and nature of deep learning (DL) solutions. Note that these traits are not necessarily unique to medical imaging. For example, satellite imaging shares the first trait described in the following with medical imaging.

Medical images have multiple modalities and are dense in pixel resolution. There are many existing imaging modalities, and new modalities, such as spectral CT, are being routinely invented. Even for commonly used imaging modalities, the pixel or voxel resolution has become higher, and the information density has increased. For example, the spatial resolution of clinical CT and MRI has reached the submillimeter level, and the spatial resolution of ultrasound is even better, while its temporal resolution exceeds real time.

Medical image data are isolated and acquired in nonstandard settings. Although medical imaging data exist in large numbers in the clinic, due to a lack of standardized acquisition protocols, there is a large variation in terms of equipment and scanning settings, leading to the so-called "distribution drift" phenomenon. Due to patient privacy and clinical data management requirements, images are scattered among different hospitals and imaging centers, and truly centralized open-source medical big data are rare.

The disease patterns in medical images are numerous, and their incidence exhibits a long-tailed distribution. Radiology gamuts ontology [2] defines 12 878 "symptoms" (conditions that lead to results) and 4662 "diseases" (imaging findings). The incidence of the disease has a typical long-tailed distribution: while a small number of common diseases have sufficient observed cases for large-scale analysis, most diseases are infrequent in the clinic. In addition, novel contagious diseases that are not represented in the current ontology, such as the outbreak of COVID-19, occur with some frequency.

The labels associated with medical images are sparse and noisy. Labeling or annotating a medical image is time-consuming and expensive. Also, different tasks require different forms of annotation, which creates the phenomenon of label sparsity. Because of variable experience and different conditions, both interuser and intrauser labeling inconsistencies are high [3], and labels must, therefore, be considered to be noisy. In fact, the establishment of gold standards for image labeling remains an open issue.

Samples are heterogeneous and imbalanced: In the already labeled images, the appearance varies from sample to sample, with its probability distribution being multimodal. The ratio between positive and negative samples is extremely uneven. For example, the number of pixels belonging to a tumor is usually one to many orders of magnitude less than that of normal tissue.

Medical image processing and analysis tasks are complex and diverse: Medical imaging has a rich body of tasks. At the technical level, there is an array of technologies, including reconstruction, enhancement, restoration, classification, detection, segmentation, and registration. When these technologies are combined with multiple image modalities and numerous disease types, a very large number of highly complex tasks associated with numerous applications are formed and should be addressed.

B. Clinical Needs and Applications

Medical imaging is often a key part of the medical diagnosis and treatment process. Typically, a radiologist reviews the acquired medical images and writes a summarizing report of their findings. The referring physician defines a diagnosis and treatment plan based on the images and radiologist's report. Often, medical imaging is ordered as part of a patient's follow-up to verify successful

[1]"The Digital Universe Driving Data Growth in Healthcare," published by EMC with research and analysis from IDC (12/13).

treatment. In addition, images are becoming an important component of invasive procedures, being used both for surgical planning and real-time imaging during the procedure itself.

As a specific example, we can look at what we term the "radiology challenge" [4], [5]. In the past decade, with the development of technologies related to the image acquisition process, imaging devices have improved in speed and resolution. For example, in 1990, a CT scanner might acquire 50–100 slices, whereas today's CT scanners might acquire 1000–2500 slices per case. A single whole slide digital pathology image corresponding to a single prostate biopsy core can easily occupy 10 GB of space at 40× magnification. Overall, there are billions of medical imaging studies conducted per year, worldwide, and this number is growing.

Most interpretations of medical images are performed by physicians and, in particular, by radiologists. Image interpretation by humans, however, is limited due to human subjectivity, the large variations across interpreters, and fatigue. Radiologists who review cases have limited time to review an ever-increasing number of images, which leads to missed findings, long turn-around times, and a paucity of numerical results or quantification. This, in turn, drastically limits the medical community's ability to advance toward more evidence-based personalized healthcare.

Artificial intelligence (AI) tools, such as DL technology, can provide support to physicians by automating image analysis, leading to what we can term "computational radiology" [6], [7]. Among the automated tools that can be developed are *detection* of pathological findings, *quantification* of disease extent, *characterization* of pathologies (e.g., into benign versus malignant), and assorted software tools that can be broadly characterized as *decision support*. This technology can also extend physicians' capabilities to include the characterization of 3-D and time-varying events, which are often not included in today's radiological reports because of both limited time and limited visualization and quantification tools.

C. Key Technologies and Deep Learning

Several key technologies arise from the various medical imaging applications, including the following [8]–[11].

1) *Medical image reconstruction [12],* which aims to form a visual representation (aka an image) from signals acquired by a medical imaging device, such as a CT or MRI scanner. Reconstruction of high-quality images from low doses and/or fast acquisitions has important clinical implications.
2) *Medical image enhancement,* which aims to adjust the intensities of an image so that the resultant image is more suitable for display or further analysis. Enhancement methods include denoising, super-resolution, MR bias field correction [13], and image harmonization [14]. Recently, much research has focused on modality translation and synthesis, which can be considered as image enhancement steps.
3) *Medical image segmentation [15],* which aims to assign labels to pixels so that the pixels with the same label form a segmented object. Segmentation has numerous applications in clinical quantification, therapy, and surgical planning.
4) *Medical image registration [16],* which aims to align the spatial coordinates of one or more images into a common coordinate system. Registration finds wide use in population analysis, longitudinal analysis, and multimodal fusion and is also commonly used for image segmentation via label transfer.
5) *Computer-aided detection (CADe) and diagnosis (CADx) [17]:* CADe aims to localize or find a bounding box that contains an object (typically a lesion) of interest. CADx aims to further classify the localized lesion as benign/malignant or one of the multiple lesion types.
6) *Others technologies* include landmark detection [18], image or view recognition [19], automatic report generation [20], and so on.

In mathematics, the above technologies can be regarded as function approximation methods, which approximates the true mapping F that takes an image (or multiple images if multimodality is accessible) as input and outputs a specific y, $y = F(x)$. The definition of y varies depending on the technology, which itself depends on the application or task. In CADe, y denotes a bounding box. In image registration, y is a deformation field. In image segmentation, y is a label mask. In image enhancement, y is a quality-enhanced image typically of the same size[2] of the input image x.

There are many ways to approximate F; however, DL [21], a branch of machine learning (ML), is one of the most powerful methods for function approximation. Since its renaissance, DL has been widely used in various medical imaging tasks and has achieved substantial success in many medical imaging applications. Because of its focus on learning rather than modeling, the use of DL in medical imaging represents a substantial departure from previous approaches in medical imaging. Take supervised DL as an example. Assume that a training data set $\{(x_n, y_n); n = 1, \ldots, N\}$ is available and that a deep neural network is parameterized by θ, which includes the number of layers, the number of nodes of each layer, the connecting weights, the choices of activation functions, and so on. The neural network that is found to approximate F can be written as $\phi_{\hat{\theta}}(x)$, where $\hat{\theta}$ are the parameters that minimize the so-called loss function

$$L(\theta) = \frac{1}{N}\sum_{n=1}^{N} l(\phi_\theta(x_n), y_n) + R_1(\phi_\theta(x_n)) + R_2(\theta) \quad (1)$$

[2]Image super-resolution generates an output image that has a different size from the input image.

where $l(\phi_\theta(x), y)$ is the itemwise loss function that penalizes the prediction error, $R_1(\phi_\theta(x_n))$ reflects the prior belief about the output, and $R_2(\theta)$ is a regularization term about the network parameters. Although the neural network $\phi_{\hat{\theta}}(x)$ does represent a type of model, it is generally thought of as a "black box" since it does not represent a designed model based on well-understood physical or mathematical principles.

There are many survey papers on DL-based key technologies for medical image analysis [22]–[31]. To differentiate the present review paper from these works, we specifically omit any presentation of the technical details of DL itself, which is no longer considered new and is well-covered in numerous other works, and focus instead on the connections between the emerging DL approaches and the specific needs in medical imaging and on several case examples that illustrate the state of the art.

D. Historical Perspective

Here, we briefly outline the development timeline of DL in medical imaging. DL was termed one of the ten breakthrough technologies of 2013 [32]. This followed the 2012 large-scale image categorization challenge that introduced the CNN superiority on the ImageNet data set [33]. At that time, DL emerged as the leading machine-learning tool in the general imaging and computer vision domains, and the medical imaging community began a debate about whether DL would be applicable in the medical imaging space. The concerns were due to the challenges that we have outlined above, with the main challenge being the lack of sufficient labeled data, known as the *data challenge*.

Several steps can be pointed to as *enablers* of the DL technology within the medical imaging space: in 2015–2016, techniques were developed using "transfer learning" (TL) (or what was also called "learning from nonmedical features" [34]) to apply the knowledge gained via solving a source problem to a different but related target problem. A key question was whether a network pretrained on natural imagery would be applicable to medical images. Several groups showed this to be the case (e.g., [34]–[36]); using the deep network trained based on ImageNet and fine-tuning to a medical imaging task was helpful in order to speed up training convergence and improve accuracy.

In 2017–2018, *synthetic data augmentation* emerged as a second solution to process limited data sets. Classical augmentation is a key component of any network training. Still, key questions to address were whether it was possible to synthesize medical data using schemes, such as generative modeling, and whether the synthesized data would serve as viable medical examples and would, in practice, increase the performance of the medical task at hand. Several works across varying domains demonstrated that this was, in fact, the case. In [37], for example, synthetic image augmentation based on the generative adversarial network (GAN) was shown to generate lesion image samples that were not recognized as synthetic by the expert radiologists and also increased CNN performance in classifying liver lesions. GANs, variational encoders, and variations on these are still being explored and advanced in recent works, as will be described in Section I-E.

For image segmentation, one of the key contributions that emerged from the medical imaging community was the U-Net architecture [38]. Originally designed for microscopic cell segmentation, the U-Net has proven to efficiently and robustly learn effective features for many medical image segmentation tasks.

E. Emerging Deep Learning Approaches

1) Network Architectures: Deep neural networks have a larger model capacity and stronger generalization capability than shallow neural networks. Deep models trained on large-scale annotated databases for a single task achieve outstanding performances, far beyond traditional algorithms, or even human capability.

Making it deeper: Starting from AlexNet [33], there was a research trend to make networks deeper, as represented by VGGNet [39], Inception Net [40], and ResNet [41]. The use of skip connections makes a deep network more trainable as in DenseNet [42] and U-Net [38]. U-net was first proposed to tackle segmentation, while the other networks were developed for image classification. Deep supervision [43] further improves discriminative power.

Adversarial and attention mechanisms: In the GAN, Goodfellow *et al.* [44] propose to accompany a generative model with a discriminator that tells whether a sample is from the model distribution or the data distribution. Both the generator and discriminator are represented by deep networks, and their training is done via a minimax optimization. Adversarial learning is widely used in medical imaging [27], including medical image reconstruction [12], image quality enhancement [14], and segmentation [45].

Attention mechanism [46] allows automatic discovery of "where" and "what" to focus on when describing image contents or making a holistic decision. Squeeze and excitation [47] can be regarded as a channel attention mechanism. Attention is combined with GAN in [48] and with U-Net in [49].

Neural architecture search (NAS) and lightweight design: NAS [50] aims to automatically design the architecture of a deep network for high performance geared toward a given task. Zhu *et al.* [51] successfully apply NAS to volumetric medical image segmentation. Lightweight design [52], [53], on the other hand, aims to design the architecture for computational efficiency on resource-constrained devices, such as mobile phones while maintaining accuracy.

2) Annotation Efficient Approaches: To address sparse and noisy labels, we need DL approaches that are efficient with respect to annotations. Thus, a key idea is to leverage the power and robustness of feature representation capability derived from existing models and data even

though the models or data are not necessarily from the same domain or for the same task and to adapt such representation to the task at hand. To do this, there are a handful of methods proposed in the literature [28], including TL, domain adaptation, self-supervised learning, semisupervised learning, and weakly/partially supervised learning.

TL aims to apply the knowledge gained via solving a source problem to a different but related target problem. One commonly used TL method is to use the deep network trained based on ImageNet and fine-tune it to a medical imaging task in order to speed up training convergence and improve accuracy. With the availability of a large number of annotated data sets, such TL methods [36] achieve remarkable success. However, ImageNet consists of natural images, and its pretrained models are for 2-D images only and are not necessarily the best for medical images, especially for small-sample settings [54]. Liu *et al.* [55] propose a 3-D anisotropic hybrid network that effectively transfers convolutional features learned from 2-D images to 3-D anisotropic volumes. Chen *et al.* [56] combine multiple data sets from several medical challenges with diverse modalities, target organs, and pathologies and learn one 3-D network that provides an effective pretrained model for 3-D medical image analysis tasks.

Domain adaptation is a form of TL in which the source and target domains have the same feature space but different distributions. In [57], domain-invariant features are learned via an adversarial mechanism that attempts to classify the domain of the input data. Zhang *et al.* [58] propose to synthesize and segment multimodal medical volumes using GANs with the cycle and shape consistencies. In [59], a domain adaptation module that maps the target input to features that are aligned with source domain feature space is proposed for cross-modality biomedical image segmentation, using a domain critic module for discriminating the feature space of both domains. Huang *et al.* [60] propose a universal U-Net comprising domain-general and domain-specific parameters to deal with multiple organ segmentation tasks on multiple domains. This *integrated learning* mechanism offers a new possibility of dealing with multiple domains and even multiple heterogeneous tasks.

Self-supervised learning, a form of unsupervised learning, learns a representation through a proxy task, in which the data provide supervisory signals. Once the representation is learned, it is fine-tuned by using annotated data. The models' genesis method [61] uses a proxy task of recovering the original image using a distorted image as input. The possible distortions include nonlinear gray-value transformation, local pixel shuffling, and image out-painting and in-painting. Zhu *et al.* [62] propose solving Rubik's cube proxy task that involves three operations, namely cube ordering, cube rotating, and cube masking. This allows the network to learn features that are invariant to translation and rotation and robust to noise as well.

Semisupervised learning often trains a model using a small set of annotated images, then generates pseudolabels for a large set of images with annotations, and learns a final model by mixing up both sets of images. Bai *et al.* [63] implement such a method for cardiac MR segmentation. Nie *et al.* [64] propose an attention-based semisupervised deep network for segmentation. It adversarially trains a segmentation network, from which a confidence map is computed as a region-attention-based semisupervised learning strategy to include the unlabeled data for training.

Weakly or partially supervised learning: Wang *et al.* [65] solve a weakly supervised multilabel disease classification from a chest X-ray. To relax the stringent pixelwise annotation for image segmentation, weakly supervised methods that use image-level annotations [66] or weak annotations, such as dots and scribbles [67], are proposed. For multiorgan segmentation, Shi *et al.* [68] learn a single multiclass network from a union of multiple data sets, each with low sample size and partial organ label, using newly proposed marginal loss and exclusion loss. Schlegl *et al.* [69] build a deep model from only normal images to detect abnormal regions in a test image.

Unsupervised learning and disentanglement: Unsupervised learning does not rely on the existence of annotated images. A disentangled network structure is designed with an adversarial learning strategy that promotes the statistical matching of deep features, which has been widely used. In medical imaging, unsupervised learning and disentanglement have been used in image registration [70], motion tracking [71], artifact reduction [72], improving classification [73], domain adaptation [74], and general modeling [75].

3) Embedding Knowledge Into Learning: Knowledge arises from various sources, such as imaging physics, statistical constraints, and task specifics and ways of embedding into a DL approach, vary too. For chest X-ray disease classification, Li *et al.* [76], [77] encode anatomy knowledge embedded in unpaired CT into a deep network that decomposes a chest X-ray into lung, bone, and the remaining structures (see Fig. 2). With augmented bone-suppressed images, classification performance is improved in predicting 11 out of 14 common lung diseases. In [78], lung radiographs are enhanced by learning to extract lung structures from CT-based simulated X-ray (DRRs) and fusing them with the original X-ray image. The enhancement is shown to augment results of pathology characterization in real X-ray images. In [79], a dual-domain network is proposed to reduce metal artifacts on both the image and sinogram domains, which are seemingly integrated into one differential framework through a Radon inverse layer, rather than two separate modules.

4) Federated Learning: To combat issues related to data privacy, data security, and data access rights, it has become increasingly important to have the capability of learning a common, robust algorithmic model through distributed computing and model aggregation strategies so that no data are transferred outside a hospital or an imaging lab. This research direction is called federated learning (FL) [80], which is in contrast to conventional centralized

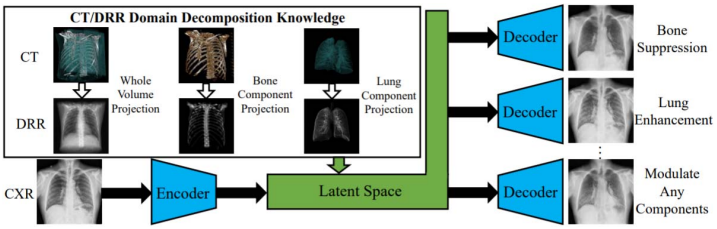

Fig. 2. ***Leveraging the anatomy knowledge embedded in CT to decompose a chest X-ray [76].***

learning with all the local data sets uploaded to one server. There are many ongoing research challenges related to federate learning, such as reduced communication burden [81], data heterogeneity in various local sites [82], and vulnerability to attacks [83].

Despite its importance, work on FL in medical imaging has only been reported recently. Sheller *et al.* [84] present the first use of FL for multi-institutional DL model without sharing patient data and report similar brain lesion segmentation performances between the models trained in a federated or centralized way. Li *et al.* [85] study several practical FL methods while protecting data privacy for brain tumor segmentation on the BraTS data set and demonstrate a tradeoff between model performance and privacy protection costs. Recently, FL is applied, together with domain adaptation, to train a model with boosted analysis performance and reliable discovery of disease-related biomarkers [86].

5) Interpretability: Clinical decision-making relies heavily on evidence gathering and interpretation. Lacking evidence and interpretation makes it difficult for physicians to trust the ML model's prediction, especially in disease diagnosis. In addition, interpretability is also the source of new knowledge. Murdoch *et al.* [87] define interpretable ML as leveraging machine-learning models to extract relevant knowledge about domain relationships contained in data, aiming to provide insights for a user into a chosen domain problem. Most interpretation methods are categorized as model-based and posthoc interpretability. The former is about constraining the model so that it readily provides useful information (such as sparsity and modularity) about the uncovered relationships. The latter is about extracting information about what relationships the model has learned.

Model-based interpretability: For cardiac MRI classification [88], diagnostically meaningful concepts in the latent space are encoded. In [89], when training the model for healthy and hypertrophic cardiomyopathy classification, it leverages interpretable task-specific anatomic patterns learned from 3-D segmentations.

Posthoc interpretability: In [90], the feature importance scores are calculated for graph neural network by comparing its interpretation ability with the random forest. Li *et al.* [91] propose a brain biomarker interpretation method through a frequency-normalized sampling strategy to corrupt an image. In [92], various interpretability methods are evaluated in the context of semantic segmentation of polyps from colonoscopy images. In [93], a hybrid RBM-random forest system on brain lesion segmentation is learned with the goal of enhancing interpretability of automatically extracted features.

6) Uncertainty Quantification: It characterizes the model prediction with confidence measure [94], which can be regarded as a method of posthoc interpretability, even though, often, the uncertainty measure is calculated along with the model prediction. Recently, there are emerging works that quantify uncertainty in DL methods for medical image segmentation [95]–[97], lesion detection [98], chest X-ray disease classification [99], and diabetic retinopathy grading [100], [101]. One additional extension to uncertainty is its combination with the knowledge that the given labels are noisy. Works are now starting to emerge that take into account label uncertainty in the modeling of the network architecture and its training [102].

II. CASE STUDIES WITH PROGRESS HIGHLIGHTS

Given that DL has been used in a vast number of medical imaging applications, it is nearly infeasible to cover all possible related literature in a single paper. Therefore, we cover several selected cases that are commonly found in clinical practices, which includes chest, neuro, cardiovascular, abdominal, and microscopy imaging. Furthermore, rather than presenting an exhaustive literature survey for each studied case, we provide some prominent progress highlights in each case study.

A. Deep Learning in Thoracic Imaging

Lung diseases have high mortality and morbidity. In the top ten causes of death worldwide, we find lung cancer, chronic obstructive pulmonary disease (COPD), pneumonia, and tuberculosis (TB). At the moment of writing this overview, COVID-19 has a death rate comparable to TB. Imaging is highly relevant to diagnose, plan treatment, and learn more about the causes and mechanisms underlying these and other lung diseases. Next to that, pulmonary complications are common in hospitalized patients. As a result, chest radiography is by far the most common radiological examination, often comprising over a third of all studies in a radiology department.

Plain radiography and CT are the two most common modalities to image the chest. The high contrast in density between air-filled lung parenchyma and tissue makes CT

ideal for *in vivo* analysis of the lungs, obtaining high-quality and high-resolution images even at very low radiation dose. Nuclear imaging (PET or PET/CT) is used for diagnosing and staging oncology patients. MRI is somewhat limited in the lungs but can yield unique functional information. Ultrasound imaging is also difficult because sound waves reflect strongly at boundaries of air and tissue, but point-of-care ultrasound is used at the emergency department and is widely used to monitor COVID-19 patients for which the first decision support applications based on DL have already appeared [103].

1) Segmentation of Anatomical Structures: For analysis and quantification from chest CT scans, automated segmentation of major anatomical structures is an important prerequisite. Recent publications demonstrate convincingly that DL is now the state-of-the-art method to achieve this. This is evident from inspecting the results of LOLA11,[3] a competition started in 2011 for lung and lobe segmentation in chest CT. The test data set for this challenge includes many challenging cases with lungs affected by severe abnormalities. For years, the best results were obtained by interactive methods. In 2019 and 2020, seven fully automatic methods based on U-Nets or variants thereof made the top ten for lung segmentation, and for lobe segmentation, two recent methods obtained results outperforming the best interactive methods. Both these methods [104], [105] were trained on thousands of CT scans from the COPDGene study [106], illustrating the importance of large high-quality data sets to obtain good results with DL. These data are publicly available on request. Both methods use a multiresolution U-Net like architecture with several customizations. Gerard *et al.* [107] integrate a previously developed method for finding the fissures. Xie *et al.* [105] add a nonlocal module with self-attention and fine-tune their method on data of COVID-19 suspects to accurately segment the lobes in scans affected by ground glass and consolidations.

Segmentation of the vasculature, separated into arteries and veins, and the airway tree, including labeling of the branches and segmentation of the bronchial wall, is another important area of research. Although methods that use convolutional networks in some of their steps have been proposed, developing an architecture entirely based on DL that can accurately track and segment intertwined tree structures and take advantage of the known geometry of these complex structures is still an open challenge.

2) Detection and Diagnosis in Chest Radiography: Recently, the number of publications on detecting abnormalities in the ubiquitous chest X-ray has increased enormously. This trend has been driven by the availability of large public data sets, such as ChestXRay14 [65], CheXpert [108], MIMIC [109], and PadChest [110], totaling 868k images. Labels for the presence or absence of over 150 different abnormal signs were gathered by text-mining the accompanying radiology reports. This makes the labels noisy. Most publications use a standard approach of inputting the entire image in a popular convolutional network architecture. Methodological contributions include novel ways of preprocessing the images, handling the label uncertainty and a large number of classes, suppressing the bones [77], and exploiting self-supervised learning as a way of pretraining. So far, only a few publications analyze multiple exams of the same patient to detect interval change or analyze the lateral views.

3) Decision Support in Lung Cancer Screening: Following the positive outcome of the NLST trial, the United States has started a screening program for heavy smokers for early detection of lung cancer with annual low-dose CT scans. Many other countries worldwide are expected to follow suit. In the United States, screening centers have to use a reporting system called Lung-RADS [111]. Reading lung cancer screening CT scans is time-consuming, and therefore, automating the various steps in Lung-RADS has received a lot of attention.

The most widely studied topic is nodule detection [112]. Nodules may represent lung cancer. Many DL systems were compared in the LUNA16 challenge.[4] Lung-RADS classifies scans in categories based on the most suspicious nodule, and this is determined by the nodule type and size. DL systems to determine nodule type have been proposed [113], and measuring the size can be done by traditional methods based on thresholding and mathematical morphology but also with DL networks. Finally, Lung-RADS contains the option to directly refer scans with a nodule that is highly suspicious for cancer. Many DL systems to estimate nodule malignancy have been proposed.

The advantage of automating the LUNG-RADS guidelines step-by-step is that this leads to an explainable AI solution that can directly support radiologists in their reading workflow. Alternatively, one could ask a computer to directly predict if a CT scan contains an actionable lung cancer. This was the topic of a Kaggle challenge organized in 2017,[5] in which almost 2000 teams competed for a one million dollar prize. The top ten solutions all used DL and are open source. Two years later, a team from Google published an implementation [114] following the approach of the winning team in the Kaggle challenge, employing modern architectures, such as a 3-D inflated inception architecture (I3D) [115]. The I3D architecture builds upon the Inception v1 model for 2-D image classification but inflates the filters and pooling kernels into 3-D. This enables the use of an image classification model pretrained with 2-D data for a 3-D image classification task. This article showed that the model outperformed six radiologists that followed Lung-RADS. The model was also extended to handle follow-up scans where it obtained performance slightly below human experts.

[3]https://lola11.grand-challenge.org/

[4]https://luna16.grand-challenge.org/

[5]https://www.kaggle.com/c/data- science-bowl-2017/

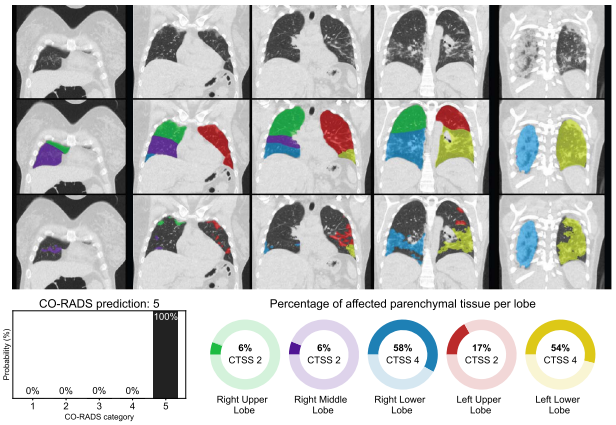

Fig. 3. ***Example output of the CORADS-AI system for a COVID-19 case. Top row: coronal slices. Second row: lobe segmentation. Bottom row: detected abnormal areas of patchy ground glass and consolidation typical for COVID-19 infection. The CO-RADS prediction and CT severity score per lobe are displayed below the images.***

4) COVID-19 Case Study: As an illustration of how DL with pretrained elements can be used to rapidly build applications, we briefly discuss the development of two tools for COVID-19 detection, for chest radiographs and chest CT. In March 2020, many European hospitals were overwhelmed by patients presenting at the emergency care with respiratory complaints. There was a shortage of molecular testing capacity for COVID-19, and the turnaround time for test results was often days. Hospitals, therefore, used chest X-ray or CT to obtain a working diagnosis and decide whether to hospitalize patients and how to treat them. In just six weeks, researchers from various Dutch and German hospitals, research institutes, and a company managed to create a solution for COVID-19 detection from an X-ray and from a CT scan. Fig. 3 shows the result of this CORADS-AI system for a COVID-19 positive case.

The X-ray solution started from a convolutional network using local and global labels, pretrained to detect TB [116], fine-tuned using public and private data of patients with and without pneumonia to detect pneumonia in general, and subsequently fine-tuned on X-ray data from patients of a Dutch hospital in a COVID-19 hotspot. The system was subsequently evaluated on 454 chest radiographs from another Dutch hospital and shown to perform comparably to six chest radiologists [117]. The system is currently being field-tested in Africa.

The CT solution, called CO-RADS [118], aimed to automate a clinical reporting system for CT of COVID-19 suspects. This system assesses the likelihood of COVID-19 infection on a scale from CO-RADS 1 (highly unlikely) to CO-RADS 5 (highly likely) and quantifies the severity of disease using a score per lung lobe from 0 to 5 depending on percentage affected lung parenchyma for a maximum CT severity score of 25 points. The previously mentioned lobe segmentation [105] was employed. Abnormal areas in the lung were segmented using a 3-D U-net built with the nnU-Net framework [119] in a cross-validated fashion with 108 scans and corresponding reference delineations to segment ground-glass opacities and consolidation in the lungs. The CT severity score was derived from the segmentation results by computing the percentage of affected parenchymal tissue per lobe. nnU-Net was compared with several other approaches and performed best. For assessing the CO-RADS score, the previously mentioned I3D architecture [115] performed best.

B. Deep Learning in Neuroimaging

In recent years, DL has seen a dramatic rise in popularity within the neuroimaging community. Many neuroimaging tasks, including segmentation, registration, and prediction, now have DL-based implementations. In addition, through the use of deep generative models and adversarial training, DL has enabled new avenues of research in complex image synthesis tasks. With the increasing availability of large and diverse pooled neuroimaging studies, DL offers interesting prospects for improving accuracy and generalizability while reducing inference time and the need for complex preprocessing. CNNs, in particular, have allowed for efficient network parameterization and spatial invariance, both of which are critical when dealing with high-dimensional neuroimaging data. The learnable feature reduction and selection capabilities of CNNs have proven effective in high-level prediction and analysis tasks and have reduced the need for highly specific domain knowledge. Specialized networks, such as U-Nets [38], V-Nets [120], and GANs [44], are also popular in neuroimaging and have been leveraged for a variety of segmentation and synthesis tasks.

1) Neuroimage Segmentation and Tissue Classification: Accurate segmentation is an important preprocessing step that informs much of the downstream analytic and predictive tasks done in neuroimaging. Commonly used tools, such as FreeSurfer [122], rely on atlas-based methods, whereby an atlas is deformably registered to the scan, which requires time-consuming optimization problems to be solved. Proposed DL-based approaches, however, are relatively computationally inexpensive during inference. Recent research has focused on important tasks, such as DL-based brain extraction [123], cortical and subcortical segmentation [124]–[127], and tumor and lesion segmentation [128], [129]. Some interesting research has looked at improving the generalization performance of DL-based segmentation methods across neuroimaging data sets imaged at different scanners. In particular, Kamnitsas *et al.* [57] have proposed a training schema, which leverages adversarial training to learn scanner invariant feature representations. They use an adversarial network to classify the origin of the input data based on the downstream feature representation learned by the segmentation network. By penalizing the segmentation network for improved performance of the adversarial network, they show improved segmentation generalization across data

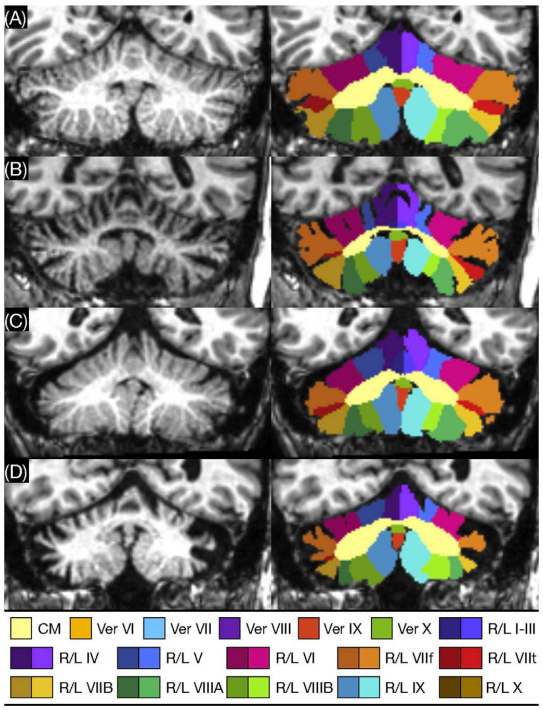

Fig. 4. ***Cerebellum parcellation by the ACAPULCO cascaded deep networks method. Lobule labels are shown for (a) healthy subject and subjects with (b) spinocerebellar ataxia (SCA) type 2, (c) SCA type 3, and (d) SCA type 6 [121].***

sets. Brain tumor segmentation has been another active area of research in the neuroimaging community where DL has shown promise. In the past, brain tumor data sets have remained relatively small, particularly ones with subjects imaged at a single institution. The Brain Tumor Segmentation Challenge (BraTS) [130] has provided the community with an accessible data set and a way to benchmark various approaches against one another. While it has been seen that DL has difficulty in training on data sets with relatively few scans, new architectures and training methods are becoming increasingly effective at this. Havaei *et al.* [128] demonstrate the performance of their glioblastoma segmentation network on the BraTS data set, achieving high accuracy while being substantially faster than previous methods. Another task where deep networks are finding increasing success is semantic segmentation in which anatomical labels are not necessarily well-defined by image intensity changes but can be identified by relative anatomical locations. A good example is that of cerebellum parcellation where deep networks performed best in a recent comparison of methods [131]. The even newer ACAPULCO method [121] uses two cascaded deep networks to produce cerebellar lobule labels, as shown in Fig. 4.

2) Deformable Image Registration: Image registration allows for imaging analysis in a single subject across imaging modalities and time points. DL-based deformable registration with neuroimaging data has proven to be a difficult problem, especially considering the lack of ground truth. Still, some unique and varied approaches have achieved state-of-the-art results with relatively fast run times [132]–[135]. Li and Fan [133] propose a fully convolutional "self-supervised" approach to learn the appropriate spatial transformations at multiple resolutions. Balakrishnan *et al.* [132] propose a method for unsupervised image registration, which attempts to directly compute the deformation field.

3) Neuroimaging Prediction: With many architectures being borrowed from the computer vision community, DL-based prediction in neuroimaging has quickly gained popularity. Traditionally, ML-based prediction on neuroimaging data has relied on careful feature selection/engineering, often taking the form of regional summary measures, which may not account for all the informative variation for a particular task, whereas, in DL, it is common to work with raw imaging data, where the appropriate feature representations can be learned through optimization. This can be particularly useful for high-level prediction tasks in which we do not know what imaging features will be informative. Furthermore, by working on the raw image, reliance on complex and time-consuming preprocessing can be reduced. In recent years, a large amount of work has been published on DL-based prediction tasks, such as brain age prediction [136]–[138], Alzheimer's disease classification and trajectory modeling [139]–[142], and schizophrenia classification [143], [144]. Some work has considered the use of deep Siamese networks for longitudinal image analysis. Siamese networks have gained popularity for their success in facial recognition. They work by jointly optimizing a set of weights on two images with respect to some distance metric between them. This setup makes them effective at identifying longitudinal changes on some chosen dimensions. Bhagwat *et al.* [145] consider the use of longitudinal Siamese network for the prediction of future Alzheimer's disease onset, using two early time points. They show substantially improved performance in identifying future Alzheimer's cases, with the use of two time points versus only a baseline scan.

4) Use of GANs in Neuroimaging: GANs have enabled complex image synthesis tasks in neuroimaging, many of which have no comparable analogs in traditional ML. GANs and their variants have been used in neuroimaging for cross-modality synthesis [146], motion artifact reduction [147], resolution upscaling [148]–[150], estimating full-dosage PET images from low-dosage PET [151]–[153], image harmonization [14], [154], heterogeneity analysis [155], and more. To help facilitate such work, the popular MedGAN [156] proposes a series of modifications and new loss functions to traditional GANs, aimed at preserving anatomically relevant information and fine details. They use auxiliary classifiers on the translated image to ensure that the resulting image feature representation is similar to the expected image representation for a given task. In addition, they use style-transfer loss in combination with an adversarial loss to ensure that fine structures and textural

details are matched in the translation. Some promising new work attempts to reduce the amount of radioactive tracer needed in PET imaging, potentially reducing associated costs and health risks. This problem can be framed as an image synthesis task, whereby the final image can be synthesized from the low-dose image. In [157], the pixel location information is integrated into a deep network for image synthesis. Kaplan and Zhu [153] propose a deep generative-based denoising method that uses paired scans of subjects imaged with both low- and full-dose PETs. They show that, despite a tenfold reduction in tracer material, they are able to preserve important edge, structural, and textural details. Consistent quantification in neuroimaging has been hampered for decades by the high variability in MR image intensities and resolutions between scans. Dewey *et al.* [14] use a U-Net style architecture and paired subjects who have been scanned with two different protocols to learn a mapping between the two sites. Resolution differences are addressed by applying a super-resolution method to the images acquired at lower resolutions [158]. They are able to use the network to reduce site-based variation, which improves consistency in segmentation between the two sites.

While DL in neuroimaging has certainly opened up many interesting avenues of investigation, certain areas still lack a rigorous understanding. Important lines of research, such as learning from limited data, optimal hyperparameter selection, domain adaptation, semisupervised designs, and improving robustness, require further investigation.

C. Deep Learning in Cardiovascular Imaging

The quantification and understanding of cardiac anatomy and function have been transformed by the recent progress in the field of data-driven DL. There has been significant recent work in a variety of subareas of cardiovascular imaging, including image reconstruction [159], end-to-end learning of cardiac pathology from images [160], and incorporation of nonimaging information (e.g., genetics [161] and clinical information) for analysis. Here, we briefly focus on three key aspects of DL in this field: cardiac chamber segmentation, cardiac motion/deformation analysis, and analysis of cardiac vessels. Motion tracking and segmentation both play crucial roles in the detection and quantification of myocardial chamber dysfunction and can help in the diagnosis of cardiovascular disease (CVD). Traditionally, these tasks are treated uniquely and solved as separate steps. Often times, motion tracking algorithms use segmentation results as an anatomical guide to sample points and regions of interest used to generate displacement fields [162]. In part due to this, there have also been efforts to combine motion tracking and segmentation.

Cardiac image segmentation is an important first step for many clinical applications. The aim is typically to segment the main chambers, e.g., the left ventricle (LV), right ventricle (RV), left atrium (LA), and right atrium (RA). This enables the quantification of parameters that describe cardiac morphology, e.g., LV volume, mass, or cardiac function, e.g., wall thickening and ejection fraction. There has been significant DL work on cardiac chamber segmentation, mostly characterized by the type of images (modalities) employed and whether the work is in 2-D or 3-D. One of the first efforts to apply a fully convolutional network (FCN) [163] to segment the LV, myocardium, and RV from 2-D short-axis cardiac MR images was done by Tran [164], significantly outperforming traditional methods in accuracy and speed. Since, this time, a variety of other FCN-based strategies have been developed [165], especially focusing on the popular U-Net approach, often including both 2-D and 3-D constraints (e.g., [166]), the incorporation of spatial and temporal contexts has also been an important research direction, including efforts to simultaneously segment the heart in both the end-diastolic and end-systolic states [167]. Shape-based constraints had previously been found useful for LV chamber segmentation using other types of ML (e.g., [168]) and were nicely included in an anatomically constrained DL strategy by [169]. This stacked convolutional autoencoder approach was successfully applied to LV segmentation from 3-D echocardiography data as well. Other important work has been aimed at atrial segmentation from MRI [170], whole heart segmentation from CT [171], and LV segmentation from 3-D ultrasound image sequences [172], the latter using a combination of atlas registration and adversarial learning. Progress in DL for cardiac segmentation is enabled by a number of ongoing challenges in the field [173], [174].

Cardiac motion tracking is key for deformation/strain analysis and is important for analyzing the mechanical performance of heart chambers. A variety of image registration, feature-based tracking, and regularization methods using both biomechanical models and data-driven learning have been developed. One special type of the data set useful for tracking is MRI tagging acquisitions, and DL has recently played a role in tracking these tags and quantifying the displacement information for motion tracking and analysis [176] using a combination of recurrent neural networks (RNNs) and convolutional neural networks (CNNs) to estimate myocardial strain from short-axis MRI tag image sequences. Estimating motion displacements and strain is also possible to do from both standard MR image sequences and 4-D echocardiography, most often by integrating ideas of image segmentation and mapping between frames using some type of image registration. Recent efforts for cardiac motion tracking from MR imaging have adopted approaches from the computer vision field, suggesting that the tasks of motion tracking and segmentation are closely related, and the information used to complete one task may complement and improve the overall performance of the other. In particular, an interesting DL approach proposed for joint learning of video object segmentation and optical

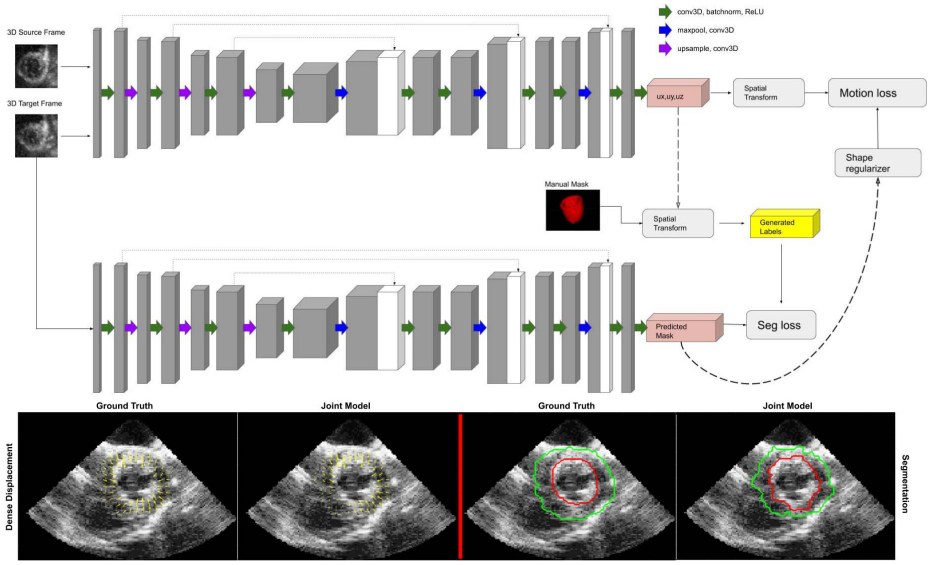

Fig. 5. *Top: 4-D two-channel CNN architecture for joint LV motion tracking/segmentation network. Bottom: 2-D echocardiography slices through 3-D canine results (left = displacements with ground truth interpolated from sonomicrometer crystals; right = LV endocardial (red) and epicardial (green) segmented boundaries with ground truth from human expert tracing). Reproduced from [175].*

flow (motion displacements) is termed SegFlow [177], an end-to-end unified network that simultaneously trains both tasks and exploits the commonality of these two tasks through bidirectional feature sharing. Among the first to integrate this idea into cardiac analysis was Qin *et al.* [71] who successfully implemented the idea of combining motion and segmentation on 2-D cardiac MR sequences by developing a dual Siamese style recurrent spatial transformer network and a fully convolutional segmentation network to simultaneously estimate motion and generate segmentation masks. This work was mainly aimed at 2-D MR images that have higher SNR than echocardiographic images and, therefore, more clearly delineated LV walls. It remains challenging to directly apply this approach to echocardiography. Very recent efforts by Ta *et al.* [175] (see Fig. 5) propose a 4-D (3-D+t) semisupervised joint network to simultaneously track LV motion while segmenting the LV wall. The network is trained in an iterative manner where the results from one branch influence and regularize the other. Displacement fields are further regularized by a biomechanically inspired incompressibility constraint that enforces realistic cardiac motion behavior. The proposed model is different from other models in which it expands the network to 4-D in order to capture out of plane motion. Finally, clinical interpretability of DL-derived motion information will be an important topic in the years ahead (e.g., [178]).

Cardiac vessel segmentation is another important task for cardiac image analysis and includes the segmentation of the vessels, including the great vessels (e.g., aorta, pulmonary arteries, and veins) and the coronary arteries. The segmentation of large vessels, such as the aorta, is important for accurate mechanical and hemodynamic characterization, e.g., for assessment of aortic compliance. Several DL approaches have been proposed for this segmentation task, including the use of RNNs in order to track the aorta in cardiac MR image sequences in the presence of noise and artifacts [179]. A similarly important task is the segmentation of the coronary arteries as a precursor to quantitative analysis for the assessment of stenosis or the simulation of blood flow simulation for the calculation of fractional flow reserve from CT angiography (CTA). The approaches for coronary artery segmentation can be divided into those approaches that extract the vessel centerline and those that segment the vessel lumen.

One end-to-end trainable approach for the extraction of the coronary centerline has been proposed in [180]. In this approach, the centerline is extracted using a multitask FCN, which simultaneously computes centerline distance maps and detects branch endpoints. The method generates single-pixel-wide centerlines with no spurious branches. An interesting aspect of this technique is that it can handle an arbitrary vessel tree with no prior assumption regarding the depth of the vessel tree or its bifurcation pattern. In contrast to this, Wolterink *et al.* [181] propose a CNN that is trained to predict the most likely direction and radius of the coronary artery within a local 3-D image patch. Starting from a seed point, the coronary artery is tracked by following the vessel centerline using the predictions of the CNN.

Alternative approaches to centerline extraction are based on techniques that instead aim to segment the vessel lumen, e.g., using CNN segmentation methods that perform segmentation by predicting vessel probability maps. One elegant approach has been proposed by Moeskops *et al.* [182]: Here, a single CNN is trained to perform three different segmentation tasks, including coronary artery

segmentation in cardiac CTA. Instead of performing voxelwise segmentation, Lee *et al.* [183] introduce a tubular shape prior for the vessel segments. This is implemented via a template transformer network, through which a shape template can be deformed via network-based registration to produce an accurate segmentation of the input image, as well as to guarantee topological constraints.

More recently, geometric DL approaches have also been applied for coronary artery segmentation. For example, Wolterink *et al.* [184] used graph convolutional networks for coronary artery segmentation. Here, vertices on the surface of the coronary artery are used as graph nodes, and their locations are optimized in an end-to-end fashion.

D. Deep Learning in Abdominal Imaging

Recently, there has been accelerating progress in automated detection, classification, and segmentation of abdominal anatomy and disease using medical imaging. Large public data sets, such as the MICCAI data decathlon and Deep Lesion data sets, have facilitated progress [185], [186].

Organs and lesions: Multiorgan approaches have been popular methods for anatomy localization and segmentation [187]. For individual organs, the liver, prostate, and spine are arguably the most accurately segmented structures and the most actively investigated with DL. Other organs of interest to DL researchers include the pancreas, lymph nodes, and bowel.

A number of studies have used U-Net to segment the liver and liver lesions and assess for hepatic steatosis [188]–[190]. Dice coefficients for liver segmentation typically exceed 95%. In the prostate, gland segmentation and lesion detection have been the subject of an SPIE/AAPM challenge (competition) and numerous publications [191], [192]. Several groups have used data sets, such as TCIA CT pancreas, to improve pancreas segmentation with Dice coefficients reaching the mid 80 percentile [193]–[196]. Automated detection of pancreatic cancer using DL has also been reported [196]. DL has been used for determining pancreatic tumor growth rates in patients with neuroendocrine tumors of the pancreas [197]. The spleen has been segmented with a Dice score of 0.962 [198]. Recently, marginal loss and exclusion loss [68] have been proposed to train a single multiorgan segmentation network from a union of partially labeled data sets.

Enlarged lymph nodes can indicate the presence of inflammation, infection, or metastatic cancer. Studies have assessed abdominal lymph nodes on CT in general and for specific diseases, such as prostate cancer [35], [199], [200]. The TCIA CT lymph node data set has enabled progress in this area [201].

In the bowel, CT colonography computer-aided polyp detection was a hot topic in abdominal CT image analysis over a decade ago. Recent progress with DL has been limited, but studies have reported improved electronic bowel cleansing, higher sensitivities, and lower false-positive rates for precancerous colonic polyp detection [199], [202]. DL using persistent homology has recently shown success for small bowel segmentation on CT [203]. Colonic inflammation can be detected on CT with DL [204]. Appendicitis can be detected on CT scans by pretraining with natural world videos [205]. The Inception V3 CNN could detect small bowel obstruction on abdominal radiographs [206].

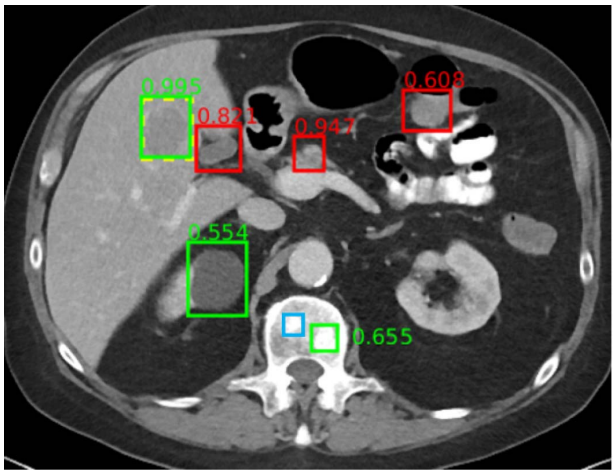

Fig. 6. *Example universal lesion detector for abdominal CT. In this axial image, through the upper abdomen, a liver lesion was correctly detected with high confidence (0.995). A renal cyst (0.554) and a bone metastasis (0.655) were also detected correctly. False positives include normal pancreas (0.947), gallbladder (0.821), and bowel (0.608). A subtle bone metastasis (blue box) was missed. Reproduced from [186].*

Kidney function can be predicted using DL of ultrasound images [207]. Potentially diffuse disorders, such as ovarian cancer and abnormal blood collections, were detectable using DL [208], [209]. Organs at risk for radiation therapy of the male pelvis, such as bladder and rectum, have been segmented on CT using U-Net [210].

Universal lesion detectors [186], [211] have been developed for body CT, including abdominal CT (see Fig. 6). The universal lesion detector identifies, classifies, and measures lymph nodes and a variety of tumors throughout the abdomen. This detector was trained using the publicly available Deep Lesion data set.

Opportunistic screening to quantify and detect underreported chronic diseases has been an area of recent interest. Example DL methods for opportunistic screening in the abdomen include automated bone mineral densitometry, visceral fat assessment, muscle volume and quality assessment, and aortic atherosclerotic plaque quantification [212]. Studies indicate that these measurements can be done accurately and generalize well to new patient populations. These opportunistic screening assessments also enable prediction of survival and cardiovascular morbidities, such as heart attack and stroke [212].

DL for abdominal imaging is likely to continue to advance rapidly. For translation to the clinic, some of the most important advances sought will be in demonstrating generalizability across different patient populations and variations in image acquisition.

E. Deep Learning in Microscopy Imaging

With the advent of whole slide scanning and the development of large digital data sets of tissue slide images, there has been a significant increase in applications of DL approaches to digital pathology data [213]. While the initial application of these approaches in the area of digital pathology primarily focused on its utility for detection and segmentation of individual primitives, such as lymphocytes and cancer nuclei, they have now progressed to addressing higher level diagnostic and prognostic tasks and also the application of DL approaches to predict the underlying molecular underpinning and mutational status of the disease. Briefly, in the following, we describe the evolving applications of DL approaches to digital pathology.

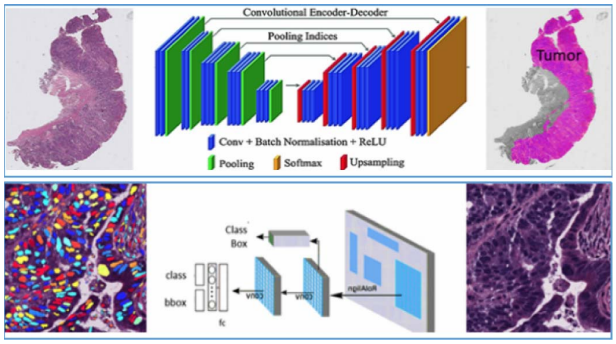

Fig. 7. *Application of DL for identifying cancerous regions from whole slide images and for identifying and segmenting different types of nuclei within the whole slide pathology images.*

1) Nuclei Detection and Segmentation: One of the early applications of DL to whole slide pathology images was in the detection and segmentation of individual nuclei. Xu *et al.* [214] present an approach using stacked spare autoencoder approach to identify the location of individual cancer nuclei on breast cancer pathology images. Subsequently, the work from Janowczyk and Madabhushi [215] demonstrates the utility of DL approaches for identifying and segmenting a number of different histologic primitives, including lymphocytes, tubules, mitotic figures, cancer extent, and also for classifying different disease categories pertaining to leukemia. The comprehensive tutorial also went into great detail with regard to best practices for annotation, network training, and testing protocols. Subsequently, Cruz-Roa *et al.* [216] demonstrate that CNNs could be applied for accurately identifying cancer presence and extent on whole slide breast cancer pathology images. The approach is shown to have a 100% accuracy at identifying the presence or absence of cancer on a slide or patient level. Subsequently, Cruz-Roa *et al.* [217] also demonstrate the use of a high-throughput adaptive sampling approach for improving the efficiency of the CNN presented previously. Bejnordi *et al.* [218] discuss the diagnostic assessment of DL algorithms for detection of lymph node metastases in women with breast cancer, as part of the CAMELYON16 challenge. The work finds that at least five DL algorithms perform comparably to a pathologist interpreting the slides in the absence of time constraints and that some DL algorithms achieve better diagnostic performance than a panel of eleven pathologists participating in a simulation exercise designed to mimic a routine pathology workflow. In a related study of lung cancer pathology images, Coudray *et al.* [219] train a deep CNN (inception V3) on WSIs from The Cancer Genome Atlas (TCGA) to accurately and automatically classify them into lung adenocarcinoma, squamous cell carcinoma, or normal lung tissue, yielding an area under the curve of 0.97.

One of the challenges with the CNN-based approaches described in [216]–[219] is the need for detailed annotations of the target of interest. This is a labor-intensive task given that annotations of disease extent typically need to be provided by pathologists who have a minimal amount of time to begin with. In a comprehensive paper by Campanella *et al.* [220], the team employs a weakly supervised approach for training a DL algorithm for identifying the presence or absence of cancer on a slide level. They are able to demonstrate on a large-scale study of over 44k WSIs from over 15k patients that, for prostate cancer, basal cell carcinoma, and breast cancer metastases to axillary lymph nodes, the corresponding areas under the curve are all above 0.98. The authors suggest that the approach could be used by pathologists to exclude 65%–75% of slides while retaining 100% cancer detection sensitivity.

2) Disease Grading: Pathologists can reliably identify disease type and extent on H&E slides and have made observations for decades that there are features of the disease that correlate with its behavior. However, they are unable to reproducibly identify or quantify these histologic hallmarks of disease behavior with enough rigor to use these features routinely to dictate disease outcome and treatment response. One of the areas to which DL has been applied is mimicking pathologist's identification of disease hallmarks, especially in the context of cancer. For instance, in prostate cancer, pathologists typically aim to place cancer into one of five different categories, referred to as the Gleason grade groups [221]. However, this grading system, as with many other cancers and diseases, is subject to interreader variability and disagreement. Consequently, a number of recent DL approaches for prostate cancer grading have been presented. Bulten *et al.* [222] and Ström *et al.* [223] both recently publish large cohort studies involving 1243 and 976 patients, respectively, and show that DL approaches could be used for achieving a performance of Gleason grading that is comparable to pathologists [224].

3) Mutation Identification and Pathway Association: Disease morphology reflects the sum of all temporal genetic and epigenetic changes and alterations within the disease. Recognizing this, some groups have begun to explore the role of DL approaches for identifying disease-specific mutations and associations with biological pathways. Oncotype DX is a 21-gene expression assay that is prognostic

and predictive of the benefit of adjuvant chemotherapy for early stage estrogen receptor-positive breast cancers. In two related studies, Romo-Bucheli *et al.* [225], [226] show that DL could be used to identify tubule density and mitotic index from pathology images and demonstrate a strong association between these measurements with the Oncotype DX risk categories (low, intermediate, and high) for breast cancers. Interestingly, tubule density and mitotic index are an important component of breast cancer grading. Microsatellite instability (MSI) is a condition that results from impaired DNA mismatch repair. To assess whether a tumor is MSI, genetic or immunohistochemical tests are required. A study by Kather *et al.* [227] shows that DL could predict MSI from histology images in gastrointestinal cancers with an AUC $= 0.84$. Coudray *et al.* [219] show that a DL network can be trained to recognize many of the commonly mutated genes in nonsmall cell lung adenocarcinoma. They show that six of these mutated genes—TK11, EGFR, FAT1, SETBP1, KRAS, and TP53—can be predicted from pathology images, with AUCs from 0.733 to 0.856.

4) Survival and Disease Outcome Prediction: More recently, there has been an interest in applying DL algorithms to pathology images to directly predict survival and disease outcome. In a recent paper, Skrede *et al.* [228] perform a large study of DL involving over 12M pathology image tiles from over 2000 patients to predict cancer-specific survival in early stage colorectal cancer patients. DL yields a hazard ratio for poor versus good prognosis of 3.84 (95% CI 2.72–5.43; $p < 0.0001$) in a validation cohort of 1122 patients and a hazard ratio of 3.04 (2.07–4.47; $p < 0.0001$) after adjusting for established prognostic markers, including T and N stages. Courtiol *et al.* [229] present an approach employing DL for predicting patient outcome in the case of mesothelioma. Saillard *et al.* [230] use DL to predict survival after hepatocellular carcinoma resection.

While the studies described above clearly reflect the growing influence and impact of DL on a variety of image analysis and classification problems in digital pathology, there are still concerns with regards to its interpretability, the need for large training sets, the need for annotated data, and generalizability. Attempts have been made to use approaches, such as visual attention mapping [231], to provide some degree of transparency with respect to where, in the image, the DL network appears to be focusing its attention. Another approach to imbue interpretability is via hybrid approaches, wherein DL is used to identify specific primitives of interest (e.g., lymphocytes) in the pathology images (in other words, using it as a detection and segmentation tool) and then deriving handcrafted features from these primitives (e.g., spatial patterns of arrangement of lymphocytes) to perform prognosis and classification tasks [232]–[234]. However, as Bera *et al.* [213] note in a recent review article, while DL approaches might be feasible for diagnostic indications, clinical tasks relating to outcome prediction and treatment response might still involve approaches that provide greater interpretability. While it seems highly likely that research in DL and its application to digital pathology are likely to continue to grow, it remains to be seen how these approaches fare in a prospective and clinical trial setting, which, in turn, might ultimately determine their translation to the clinic.

III. DISCUSSION

A. Technical Challenges Ahead

In this overview paper, many technological challenges across several medical domains and tasks have been reviewed. In general, most challenges are met by continuous improvement of solutions to the well-known *data challenge*. The community as a whole is continuously developing and improving TL-based solutions and data augmentation schemes. As systems are starting to be implemented across data sets, hospitals, and countries, a new spectrum of challenges is arising including *system robustness and generalization* across acquisition protocols, machines, and hospitals. Here, data preprocessing, continuous model learning, and fine-tuning across systems are a few of the new developments ahead. Detailed reviews of the topics presented herein, as well as additional topics, such as robustness to adversarial attacks [235], can be found in several recent DL review articles, such as [236].

B. How Do We Get New Tools Into the Clinic?

The question of whether DL tools are used in the clinic is often raised [5]. This question is particularly relevant because results in many tasks and challenges show radiologist-level performance. In several recent works conducted to estimate the utility of AI-based technology as an aid to the radiologist, it is consistently shown that human experts with AI perform better than those without AI [237]. The excitement in the field has led to the emergence of many AI medical imaging startup companies. Although there are some earlier studies [238]–[240] that present evidence that early CAD tools were not necessarily helpful in the tested scenarios and, to date, the translation of technologies from research to actual clinical use does not happen at a massive scale, there are a number of technologies that obtained the FDA approval.[6] In fact, a dedicated reimbursement code was recently awarded for an AI technology.[7] There are a variety of reasons for this delayed clinical translation, including users being cautious regarding the technology, specifically the prospect of being replaced by AI; the need to prove that the technology can address real user needs and bring quantifiable benefits; regulatory pathways that are long and costly; patient safety considerations; and economic factors, such as who will pay for AI tools.

The forecast going forward is that this is an emerging field, with enormous promise going forward. How would

[6]https://www.acrdsi.org/DSI-Services/FDA-Cleared-AI-Algorithms

[7]Federal Register/Vol. 85, No. 182/Friday, September 18, 2020/Rules and Regulations

we get there? An interesting possibility is that the worldwide experience with the COVID-19 pandemic will actually serve to bridge the gap between the *need* and *AI*—with users eager to receive support, and even the regulatory steps more adaptive to facilitate a transition of general computational tools, with COVID-AI related tools in particular. Within the last several months, we witnessed several interesting advances: the ability of AI to rapidly adapt from existing pretrained models to new disease manifestations of COVID-19, using the many tools described in this article and in [117] and [241]. Strong and robust DL-based solutions for COVID-19 detection, localization, quantification, and characterization start to support the initial diagnosis and more so the follow-up of hospitalized patients. AI-based tools are developed to support the assessment of disease severity, and recently, they start to provide tools for assessing treatment and predicting treatment success [242], [243]. Finally, numerous studies [244]–[247] in fields such as clinical neuroscience have shown that AI-based image evaluation can identify complex imaging patterns that are not perceptible with visual radiologic evaluation. For example, anatomical, functional, and connectomic signatures of many neuropsychiatric and neurologic diseases not only promise to redefine many of these diseases on a neurobiological basis and in a more precise manner but also to offer early detection and personalized risk estimates. Such capabilities can potentially significantly expand current clinical protocols and contribute to precision medicine.

C. Future Promise

As we envision future possibilities, one immediate step forward is to combine the image with additional clinical context, from patient record to additional clinical descriptors (such as blood tests, genomics, medications, vital signs, and nonimaging data, such as ECG). This step will provide a transition from *image space* to *patient-level* information. Collecting cohorts will enable the population-level statistical analysis to learn about disease manifestations, treatment responses, adverse reactions from and interactions between medications, and more. This step requires building complex infrastructure, along with the generation of new privacy and security regulations—between hospitals and academic research institutes, across hospitals, and in multinational consortia. As more and more data become available, DL and AI will enable unsupervised explorations within the data, thus providing for new discoveries of drugs and treatments toward the advancement and augmentation of healthcare as we know it. ■

Acknowledgment

The authors extend special acknowledgment to Vishnu Bashyam for his help. The mentioning of commercial products does not imply endorsement.

REFERENCES

[1] J. Beutel, H. L. Kundel, and R. L. Van Metter, *Handbook of Medical Imaging*, vol. 1. Bellingham, WA, USA: SPEI Press, 2000.
[2] J. J. Budovec, C. A. Lam, and C. E. Kahn, Jr., "Informatics in radiology: Radiology gamuts ontology: Differential diagnosis for the semantic Web," *RadioGraphics*, vol. 34, no. 1, pp. 254–264, Jan. 2014.
[3] K. H. Zou *et al.*, "Statistical validation of image segmentation quality based on a spatial overlap index1: Scientific reports," *Acad. Radiol.*, vol. 11, no. 2, pp. 178–189, 2004.
[4] D. L. Rubin, "Informatics in radiology: Measuring and improving quality in radiology: Meeting the challenge with informatics," *RadioGraphics*, vol. 31, no. 6, pp. 1511–1527, Oct. 2011.
[5] M. P. Recht *et al.*, "Integrating artificial intelligence into the clinical practice of radiology: Challenges and recommendations," *Eur. Radiol.*, vol. 30, pp. 3576–3584, Feb. 2020.
[6] C. Börgers and F. Natterer, *Computational Radiology and Imaging: Therapy and Diagnostics*, vol. 110. New York, NY, USA: Springer-Verlag, 2012.
[7] W. T. Tran *et al.*, "Computational radiology in breast cancer screening and diagnosis using artificial intelligence," *Can. Assoc. Radiol. J.*, vol. 72, no. 1, 2020, Art. no. 0846537120949974.
[8] J. Beutel, M. Sonka, H. L. Kundel, R. L. Van Metter, and J. M. Fitzpatrick, *Handbook of Medical Imaging: Medical Image Processing and Analysis*, vol. 2. Bellingham, WA, USA: SPIE Press, 2000.
[9] J. L. Prince and J. M. Links, *Medical Imaging Signals and Systems*. Upper Saddle River, NJ, USA: Prentice-Hall, 2006.
[10] I. Bankman, *Handbook of Medical Image Processing and Analysis*. Amsterdam, The Netherlands: Elsevier, 2008.
[11] S. K. Zhou, D. Rueckert, and G. Fichtinger, *Handbook of Medical Image Computing and Computer Assisted Intervention*. New York, NY, USA: Academic, 2019.
[12] G. Wang, Y. Zhang, X. Ye, and X. Mou, *Machine Learning for Tomographic Imaging*. Bristol, U.K.: IOP Publishing, 2019.
[13] M. Gaillochet, K. C. Tezcan, and E. Konukoglu, "Joint reconstruction and bias field correction for undersampled MR imaging," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2020, pp. 44–52.
[14] B. E. Dewey *et al.*, "DeepHarmony: A deep learning approach to contrast harmonization across scanner changes," *Magn. Reson. Imag.*, vol. 64, pp. 160–170, Dec. 2019.
[15] N. Tajbakhsh, L. Jeyaseelan, Q. Li, J. N. Chiang, Z. Wu, and X. Ding, "Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation," *Med. Image Anal.*, vol. 63, Jul. 2020, Art. no. 101693.
[16] Y. Fu, Y. Lei, T. Wang, W. J. Curran, T. Liu, and X. Yang, "Deep learning in medical image registration: A review," *Phys. Med. Biol.*, vol. 65, no. 20, Oct. 2020, Art. no. 20TR01.
[17] H.-P. Chan, L. M. Hadjiiski, and R. K. Samala, "Computer-aided diagnosis in the era of deep learning," *Med. Phys.*, vol. 47, no. 5, pp. e218–e227, 2020.
[18] D. Liu, K. S. Zhou, D. Bernhardt, and D. Comaniciu, "Search strategies for multiple landmark detection by submodular maximization," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2831–2838.
[19] Z. Xu *et al.*, "Less is more: Simultaneous view classification and landmark detection for abdominal ultrasound images," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent.* Cham, Switzerland: Springer, 2018, pp. 711–719.
[20] B. Jing, P. Xie, and E. Xing, "On the automatic generation of medical imaging reports," in *Proc. 56th Annu. Meeting Assoc. Comput. Linguistics, Long Papers*, vol. 1, 2018, pp. 2577–2586.
[21] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.
[22] H. Greenspan, B. van Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1153–1159, May 2016.
[23] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.
[24] S. K. Zhou, H. Greenspan, and D. Shen, *Deep Learning for Medical Image Analysis*. New York, NY, USA: Academic, 2017.
[25] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annu. Rev. Biomed. Eng.*, vol. 19, pp. 221–248, Jun. 2017.
[26] J. Ker, L. Wang, J. Rao, and T. Lim, "Deep learning applications in medical image analysis," *IEEE Access*, vol. 6, pp. 9375–9389, 2018.
[27] X. Yi, E. Walia, and P. Babyn, "Generative adversarial network in medical imaging: A review," *Med. Image Anal.*, vol. 58, Dec. 2019, Art. no. 101552.
[28] V. Cheplygina, M. de Bruijne, and J. P. W. Pluim, "Not-so-supervised: A survey of semi-supervised, multi-instance, and transfer learning in medical image analysis," *Med. Image Anal.*, vol. 54, pp. 280–296, May 2019.
[29] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, "Deep learning techniques for medical image segmentation: Achievements and challenges," *J. Digit. Imag.*, vol. 32, no. 4, pp. 582–596, Aug. 2019.
[30] J. S. Duncan, M. F. Insana, and N. Ayache, "Biomedical imaging and analysis in the age of big data and deep learning," *Proc. IEEE*, vol. 108, no. 1, pp. 3–10, Jan. 2020.
[31] G. Haskins, U. Kruger, and P. Yan, "Deep learning in medical image registration: A survey," *Mach. Vis. Appl.*, vol. 31, nos. 1–2, p. 8, Feb. 2020.
[32] (2013). *MIT Technology Review, 2013*. [Online]. Available: https://www.technologyreview.com/s/513696/deep-learning

[33] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[34] Y. Bar, I. Diamant, L. Wolf, S. Lieberman, E. Konen, and H. Greenspan, "Chest pathology detection using deep learning with non-medical training," in *Proc. IEEE 12th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2015, pp. 294–297.

[35] H.-C. Shin *et al.*, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, May 2016.

[36] V. Gulshan *et al.*, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *J. Amer. Med. Assoc.*, vol. 316, no. 22, pp. 2402–2410, 2016.

[37] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification," *Neurocomputing*, vol. 321, pp. 321–331, Dec. 2018.

[38] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*. Cham, Switzerland: Springer, 2015, pp. 234–241.

[39] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556

[40] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1–9.

[41] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[42] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 4700–4708.

[43] C.-Y. Lee, S. Xie, P. Gallagher, Z. Zhang, and Z. Tu, "Deeply-supervised nets," in *Proc. Artif. Intell. Statist.*, 2015, pp. 562–570.

[44] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[45] D. Yang *et al.*, "Automatic liver segmentation using an adversarial image-to-image network," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*. Cham, Switzerland: Springer, 2017, pp. 507–515.

[46] K. Xu *et al.*, "Show, attend and tell: Neural image caption generation with visual attention," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 2048–2057.

[47] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[48] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 7354–7363.

[49] O. Oktay *et al.*, "Attention U-net: Learning where to look for the pancreas," 2018, *arXiv:1804.03999*. [Online]. Available: http://arxiv.org/abs/1804.03999

[50] T. Elsken, J. H. Metzen, and F. Hutter, "Neural architecture search: A survey," 2018, *arXiv:1808.05377*. [Online]. Available: http://arxiv.org/abs/1808.05377

[51] Z. Zhu, C. Liu, D. Yang, A. Yuille, and D. Xu, "V-NAS: Neural architecture search for volumetric medical image segmentation," in *Proc. Int. Conf. 3D Vis. (3DV)*, Sep. 2019, pp. 240–248.

[52] A. G. Howard *et al.*, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*. [Online]. Available: http://arxiv.org/abs/1704.04861

[53] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6848–6856.

[54] M. Raghu, C. Zhang, J. Kleinberg, and S. Bengio, "Transfusion: Understanding transfer learning for medical imaging," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 3347–3357.

[55] S. Liu *et al.*, "3D anisotropic hybrid network: Transferring convolutional features from 2D images to 3D anisotropic volumes," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*. Cham, Switzerland: Springer, 2018, pp. 851–858.

[56] S. Chen, K. Ma, and Y. Zheng, "Med3D: Transfer learning for 3D medical image analysis," 2019, *arXiv:1904.00625*. [Online]. Available: http://arxiv.org/abs/1904.00625

[57] K. Kamnitsas *et al.*, "Unsupervised domain adaptation in brain lesion segmentation with adversarial networks," in *Proc. Int. Conf. Inf. Process. Med. Imag.* Cham, Switzerland: Springer, 2017, pp. 597–609.

[58] Z. Zhang, L. Yang, and Y. Zheng, "Translating and segmenting multimodal medical volumes with cycle- and shape-consistency generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9242–9251.

[59] Q. Dou, C. Ouyang, C. Chen, H. Chen, and P.-A. Heng, "Unsupervised cross-modality domain adaptation of ConvNets for biomedical image segmentations with adversarial loss," 2018, *arXiv:1804.10916*. [Online]. Available: http://arxiv.org/abs/1804.10916

[60] C. Huang, H. Han, Q. Yao, S. Zhu, and S. K. Zhou, "3D U^2-net: A 3D universal U-net for multi-domain medical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCIAI)*. Cham, Switzerland: Springer, 2019, pp. 291–299.

[61] Z. Zhou *et al.*, "Models genesis: Generic autodidactic models for 3d medical image analysis," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*. Cham, Switzerland: Springer, 2019, pp. 384–393.

[62] J. Zhu, Y. Li, Y. Hu, K. Ma, S. K. Zhou, and Y. Zheng, "Rubik's cube+: A self-supervised feature learning framework for 3D medical image analysis," *Med. Image Anal.*, vol. 64, Aug. 2020, Art. no. 101746.

[63] W. Bai *et al.*, "Semi-supervised learning for network-based cardiac MR image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*. Cham, Switzerland: Springer, 2017, pp. 253–260.

[64] D. Nie, Y. Gao, L. Wang, and D. Shen, "ASDNet: Attention based semi-supervised deep networks for medical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*. Cham, Switzerland: Springer, 2018, pp. 370–378.

[65] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2097–2106.

[66] Y. Xu, J.-Y. Zhu, E. I.-C. Chang, M. Lai, and Z. Tu, "Weakly supervised histopathology cancer image segmentation and classification," *Med. Image Anal.*, vol. 18, no. 3, pp. 591–604, Apr. 2014.

[67] H. Kervadec, J. Dolz, M. Tang, E. Granger, Y. Boykov, and I. B. Ayed, "Constrained-CNN losses for weakly supervised segmentation," *Med. Image Anal.*, vol. 54, pp. 88–99, May 2019.

[68] G. Shi, L. Xiao, Y. Chen, and S. K. Zhou, "Marginal loss and exclusion loss for partially supervised multi-organ segmentation," 2020, *arXiv:2007.03868*. [Online]. Available: http://arxiv.org/abs/2007.03868

[69] T. Schlegl, P. Seeböck, S. M. Waldstein, G. Langs, and U. Schmidt-Erfurth, "F-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks," *Med. Image Anal.*, vol. 54, pp. 30–44, May 2019.

[70] C. Qin, B. Shi, R. Liao, T. Mansi, D. Rueckert, and A. Kamen, "Unsupervised deformable registration for multi-modal images via disentangled representations," in *Proc. Inf. Process. Med. Imag. (IPMI)*. Cham, Switzerland: Springer, 2019, pp. 249–261.

[71] C. Qin *et al.*, "Joint learning of motion estimation and segmentation for cardiac MR image sequences," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*. Cham, Switzerland: Springer, 2018, pp. 472–480.

[72] H. Liao, W.-A. Lin, S. K. Zhou, and J. Luo, "ADN: Artifact disentanglement network for unsupervised metal artifact reduction," *IEEE Trans. Med. Imag.*, vol. 39, no. 3, pp. 634–643, Mar. 2020.

[73] A. Ben-Cohen, R. Mechrez, N. Yedidia, and H. Greenspan, "Improving CNN training using disentanglement for liver lesion classification in CT," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (IEEE-EMBS)*, Jul. 2019, pp. 886–889.

[74] J. Yang, N. C. Dvornek, F. Zhang, J. Chapiro, M. Lin, and J. S. Duncan, "Unsupervised domain adaptation via disentangled representations: Application to cross-modality liver segmentation," in *Proc. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*. Cham, Switzerland: Springer, 2019, pp. 255–263.

[75] A. Chartsias *et al.*, "Disentangled representation learning in cardiac image analysis," *Med. Image Anal.*, vol. 58, Dec. 2019, Art. no. 101535.

[76] Z. Li, H. Li, H. Han, G. Shi, J. Wang, and S. K. Zhou, "Encoding CT anatomy knowledge for unpaired chest X-Ray image decomposition," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*. Springer, 2019, pp. 275–283.

[77] H. Li *et al.*, "High-resolution chest X-ray bone suppression using unpaired CT structural priors," *IEEE Trans. Med. Imag.*, vol. 39, no. 10, pp. 3053–3063, Oct. 2020.

[78] O. Gozes and H. Greenspan, "Lung structures enhancement in chest radiographs via CT based FCNN training," in *Proc. Int. Workshop Reconstruct. Anal. Moving Body Organs*, in Lecture Notes in Computer Science, 2018, pp. 147–158.

[79] W.-A. Lin *et al.*, "DuDoNet: Dual domain network for CT metal artifact reduction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 10512–10521.

[80] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 2, pp. 1–19, 2019.

[81] J. Konečný, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, "Federated learning: Strategies for improving communication efficiency," 2016, *arXiv:1610.05492*. [Online]. Available: http://arxiv.org/abs/1610.05492

[82] Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra, "Federated learning with non-IID data," 2018, *arXiv:1806.00582*. [Online]. Available: http://arxiv.org/abs/1806.00582

[83] E. Bagdasaryan, A. Veit, Y. Hua, D. Estrin, and V. Shmatikov, "How to backdoor federated learning," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2020, pp. 2938–2948.

[84] M. J. Sheller, G. A. Reina, B. Edwards, J. Martin, and S. Bakas, "Multi-institutional deep learning modeling without sharing patient data: A feasibility study on brain tumor segmentation," in *Proc. Int. MICCAI BrainLesion Workshop*. Cham, Switzerland: Springer, 2018, pp. 92–104.

[85] W. Li *et al.*, "Privacy-preserving federated brain tumour segmentation," in *Proc. Int. Workshop Mach. Learn. Med. Imag.* Cham, Switzerland: Springer, 2019, pp. 133–141.

[86] X. Li, Y. Gu, N. Dvornek, L. Staib, P. Ventola, and J. S. Duncan, "Multi-site fMRI analysis using privacy-preserving federated learning and domain adaptation: ABIDE results," 2020, *arXiv:2001.05647*. [Online]. Available: http://arxiv.org/abs/2001.05647

[87] W. J. Murdoch, C. Singh, K. Kumbier, R. Abbasi-Asl, and B. Yu, "Interpretable machine learning: Definitions, methods, and applications," 2019, *arXiv:1901.04592*. [Online]. Available: http://arxiv.org/abs/1901.04592

[88] J. R. Clough, I. Oksuz, E. Puyol-Antón, B. Ruijsink, A. P. King, and J. A. Schnabel, "Global and local interpretability for cardiac MRI classification," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*. Cham, Switzerland: Springer, 2019, pp. 656–664.

[89] C. Biffi *et al.*, "Learning interpretable anatomical features through deep generative models: Application to cardiac remodeling," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent.* Cham, Switzerland: Springer, 2018, pp. 464–471.

[90] X. Li, N. C. Dvornek, Y. Zhou, J. Zhuang, P. Ventola, and J. S. Duncan, "Graph neural network for interpreting task-fMRI biomarkers," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*. Cham, Switzerland: Springer, 2019, pp. 485–493.

[91] X. Li, N. C. Dvornek, J. Zhuang, P. Ventola, and J. S. Duncan, "Brain biomarker interpretation in ASD using deep learning and fMRI," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*. Cham, Switzerland: Springer, 2018, pp. 206–214.

[92] K. Wickstrøm, M. Kampffmeyer, and R. Jenssen, "Uncertainty and interpretability in convolutional neural networks for semantic segmentation of colorectal polyps," *Med. Image Anal.*, vol. 60, Feb. 2020, Art. no. 101619.

[93] S. Pereira *et al.*, "Enhancing interpretability of automatically extracted machine learning features: Application to a RBM-random forest system on brain lesion segmentation," *Med. Image Anal.*, vol. 44, pp. 228–244, Feb. 2018.

[94] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 1050–1059.

[95] C. F. Baumgartner *et al.*, "Phiseg: Capturing uncertainty in medical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*. Springer, 2019, pp. 119–127.

[96] S. P. Awate, S. Garg, and R. Jena, "Estimating uncertainty in MRF-based image segmentation: A perfect-MCMC approach," *Med. Image Anal.*, vol. 55, pp. 181–196, Jul. 2019.

[97] A. Jungo and M. Reyes, "Assessing reliability and challenges of uncertainty estimations for medical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*. Cham, Switzerland: Springer, 2019, pp. 48–56.

[98] T. Nair, D. Precup, D. L. Arnold, and T. Arbel, "Exploring uncertainty measures in deep networks for multiple sclerosis lesion detection and segmentation," *Med. Image Anal.*, vol. 59, Jan. 2020, Art. no. 101557.

[99] F. C. Ghesu *et al.*, "Quantifying and leveraging classification uncertainty for chest radiograph assessment," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*. Cham, Switzerland: Springer, 2019, pp. 676–684.

[100] M. S. Ayhan, L. Kühlewein, G. Aliyeva, W. Inhoffen, F. Ziemssen, and P. Berens, "Expert-validated estimation of diagnostic uncertainty for deep neural networks in diabetic retinopathy detection," *Med. Image Anal.*, vol. 64, Aug. 2020, Art. no. 101724.

[101] T. Araújo *et al.*, "DR|GRADUATE: Uncertainty-aware deep learning-based diabetic retinopathy grading in eye fundus images," *Med. Image Anal.*, vol. 63, Jul. 2020, Art. no. 101715.

[102] Y. Dgani, H. Greenspan, and J. Goldberger, "Training a neural network based on unreliable human annotation of medical images," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 39–42.

[103] S. Roy *et al.*, "Deep learning for classification and localization of COVID-19 markers in point-of-care lung ultrasound," *IEEE Trans. Med. Imag.*, vol. 39, no. 8, pp. 2676–2687, Aug. 2020.

[104] S. E. Gerard and J. M. Reinhardt, "Pulmonary lobe segmentation using a sequence of convolutional neural networks for marginal learning," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2019, pp. 1207–1211.

[105] W. Xie, C. Jacobs, J.-P. Charbonnier, and B. van Ginneken, "Relational modeling for robust and efficient pulmonary lobe segmentation in CT scans," *IEEE Trans. Med. Imag.*, vol. 39, no. 8, pp. 2664–2675, Aug. 2020.

[106] E. A. Regan *et al.*, "Genetic epidemiology of COPD (COPDGene) study design," *COPD, J. Chronic Obstructive Pulmonary Disease*, vol. 7, no. 1, pp. 32–43, Feb. 2011.

[107] S. E. Gerard, T. J. Patton, G. E. Christensen, J. E. Bayouth, and J. M. Reinhardt, "FissureNet: A deep learning approach for pulmonary fissure detection in CT images," *IEEE Trans. Med. Imag.*, vol. 38, no. 1, pp. 156–166, Jan. 2019.

[108] J. Irvin *et al.*, "CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison," 2019, *arXiv:1901.07031*. [Online]. Available: http://arxiv.org/abs/1901.07031

[109] A. E. W. Johnson *et al.*, "MIMIC-CXR-JPG, a large publicly available database of labeled chest radiographs," 2019, *arXiv:1901.07042*. [Online]. Available: http://arxiv.org/abs/1901.07042

[110] A. Bustos, A. Pertusa, J.-M. Salinas, and M. de la Iglesia-Vayá, "PadChest: A large chest X-ray image dataset with multi-label annotated reports," 2019, *arXiv:1901.07441*. [Online]. Available: http://arxiv.org/abs/1901.07441

[111] American College of Radiology. (2019). *Lung CT Screening Reporting & Data System v1.1*. [Online]. Available: https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/Lung-Rads

[112] B. Liu *et al.*, "Evolving the pulmonary nodules diagnosis from classical approaches to deep learning-aided decision support: Three decades' development course and future prospect," *J. Cancer Res. Clin. Oncol.*, vol. 146, no. 1, pp. 153–185, Jan. 2020.

[113] F. Ciompi *et al.*, "Towards automatic pulmonary nodule management in lung cancer screening with deep learning," *Sci. Rep.*, vol. 7, no. 1, p. 46479, Apr. 2017.

[114] D. Ardila *et al.*, "End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography," *Nature Med.*, vol. 25, no. 6, pp. 954–961, Jun. 2019.

[115] J. Carreira and A. Zisserman, "Quo vadis, action recognition? A new model and the kinetics dataset," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4724–4733.

[116] K. Murphy *et al.*, "Computer aided detection of tuberculosis on chest radiographs: An evaluation of the CAD4TB v6 system," *Sci. Rep.*, vol. 10, no. 1, pp. 1–11, 2020. [Online]. Available: https://arxiv.org/abs/1903.03349

[117] K. Murphy *et al.*, "COVID-19 on the chest radiograph: A multi-reader evaluation of an AI system," *Radiology*, vol. 296, no. 1, 2020, Art. no. 201874.

[118] M. Prokop *et al.*, "CO-RADS—A categorical CT assessment scheme for patients with suspected COVID-19: Definition and evaluation," *Radiology*, vol. 296, no. 2, 2020, Art. no. 201473.

[119] F. Isensee, P. F. Jäger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "Automated design of deep learning methods for biomedical image segmentation," 2019, *arXiv:1904.08128*. [Online]. Available: http://arxiv.org/abs/1904.08128

[120] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Oct. 2016, pp. 565–571.

[121] S. Han, A. Carass, Y. He, and J. L. Prince, "Automatic cerebellum anatomical parcellation using U-Net with locally constrained optimization," *NeuroImage*, vol. 218, Sep. 2020, Art. no. 116819.

[122] B. Fischl, "FreeSurfer," *NeuroImage*, vol. 62, no. 2, pp. 774–781, Aug. 2012.

[123] J. Kleesiek *et al.*, "Deep MRI brain extraction: A 3D convolutional neural network for skull stripping," *NeuroImage*, vol. 129, pp. 460–469, Apr. 2016.

[124] J. Dolz, C. Desrosiers, and I. Ben Ayed, "3D fully convolutional networks for subcortical segmentation in MRI: A large-scale study," *NeuroImage*, vol. 170, pp. 456–470, Apr. 2018.

[125] H. Chen, Q. Dou, L. Yu, J. Qin, and P.-A. Heng, "VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images," *NeuroImage*, vol. 170, pp. 446–455, Apr. 2018.

[126] C. Wachinger, M. Reuter, and T. Klein, "DeepNAT: Deep convolutional neural network for segmenting neuroanatomy," *NeuroImage*, vol. 170, pp. 434–445, Apr. 2018.

[127] Y. Huo *et al.*, "3D whole brain segmentation using spatially localized atlas network tiles," *NeuroImage*, vol. 194, pp. 105–119, Jul. 2019.

[128] M. Havaei *et al.*, "Brain tumor segmentation with deep neural networks," *Med. Image Anal.*, vol. 35, pp. 18–31, Jan. 2017.

[129] K. Kamnitsas *et al.*, "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Med. Image Anal.*, vol. 36, pp. 61–78, Feb. 2017.

[130] S. Bakas *et al.*, "Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge," 2018, *arXiv:1811.02629*. [Online]. Available: http://arxiv.org/abs/1811.02629

[131] A. Carass *et al.*, "Comparing fully automated state-of-the-art cerebellum parcellation from magnetic resonance images," *NeuroImage*, vol. 183, pp. 150–172, Dec. 2018.

[132] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, "VoxelMorph: A learning framework for deformable medical image registration," *IEEE Trans. Med. Imag.*, vol. 38, no. 8, pp. 1788–1800, Aug. 2019.

[133] H. Li and Y. Fan, "Non-rigid image registration using fully convolutional networks with deep self-supervision," 2017, *arXiv:1709.00799*. [Online]. Available: http://arxiv.org/abs/1709.00799

[134] S. Miao, Z. J. Wang, and R. Liao, "A CNN regression approach for real-time 2D/3D registration," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1352–1363, May 2016.

[135] X. Yang, R. Kwitt, M. Styner, and M. Niethammer, "QuickSilver: Fast predictive image registration—A deep learning approach," *NeuroImage*, vol. 158, pp. 378–396, Sep. 2017.

[136] V. M. Bashyam *et al.*, "MRI signatures of brain age and disease over the lifespan based on a deep brain network and 14,468 individuals worldwide," *Brain*, vol. 143, no. 7, pp. 2312–2324, 2020.

[137] J. H. Cole *et al.*, "Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker," *NeuroImage*, vol. 163, pp. 115–124, Dec. 2017.

[138] B. A. Jónsson *et al.*, "Brain age prediction using deep learning uncovers associated sequence variants," *Nature Commun.*, vol. 10, no. 1, pp. 1–10, Dec. 2019.

[139] S. Liu, S. Liu, W. Cai, S. Pujol, R. Kikinis, and D. Feng, "Early diagnosis of Alzheimer's disease with deep learning," in *Proc. Int. Symp. Biomed. Imag. (ISBI)*, 2014, pp. 1015–1018.

[140] D. Lu, K. Popuri, G. W. Ding, R. Balachandar, and M. F. Beg, "Multimodal and multiscale deep neural networks for the early diagnosis of Alzheimer's disease using structural MR and FDG-PET images," *Sci. Rep.*, vol. 8, no. 1, pp. 1–13, Dec. 2018.

[141] J. Islam and Y. Zhang, "Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks," *Brain Informat.*, vol. 5, no. 2, p. 2, Dec. 2018.

[142] M. Liu, J. Zhang, E. Adeli, and D. Shen, "Landmark-based deep multi-instance learning for brain disease diagnosis," *Med. Image Anal.*, vol. 43, pp. 157–168, Jan. 2018.

[143] W. Yan *et al.*, "Discriminating schizophrenia using recurrent neural network applied on time courses of multi-site FMRI data," *EBioMedicine*, vol. 47, pp. 543–552, 2019.

[144] L.-L. Zeng *et al.*, "Multi-site diagnostic classification of schizophrenia using discriminant deep learning with functional connectivity MRI," *EBioMedicine*, vol. 30, pp. 74–85, Apr. 2018.

[145] N. Bhagwat, J. D. Viviano, A. N. Voineskos, and M. M. Chakravarty, "Modeling and prediction of clinical symptom trajectories in Alzheimer's disease using longitudinal data," *PLOS Comput. Biol.*, vol. 14, no. 9, Sep. 2018, Art. no. e1006376.

[146] J. M. Wolterink, A. M. Dinkla, M. H. Savenije, P. R. Seevinck, C. A. van den Berg, and I. Išgum, "Deep MR to CT synthesis using unpaired data," in *Proc. Int. Workshop Simulation Synth. Med. Imag.* Cham, Switzerland: Springer, 2017, pp. 14–23.

[147] D. Tamada, H. Onishi, and U. Motosugi, "Motion artifact reduction in abdominal MR imaging using the U-NET network," in *Proc. ICMRM Sci. Meeting KSMRM*, Seoul, South Korea, 2018.

[148] Y. Chen, Y. Xie, Z. Zhou, F. Shi, A. G. Christodoulou, and D. Li, "Brain MRI super resolution using 3D deep densely connected neural networks," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 739–742.

[149] C.-H. Pham, A. Ducournau, R. Fablet, and F. Rousseau, "Brain MRI super-resolution using deep 3D convolutional networks," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 197–200.

[150] T.-A. Song, S. R. Chowdhury, F. Yang, and J. Dutta, "Super-resolution PET imaging using convolutional neural networks," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 518–528, 2020.

[151] K. T. Chen *et al.*, "Ultra-low-dose 18F-florbetaben amyloid PET imaging using deep learning with multi-contrast MRI inputs," *Radiology*, vol. 290, no. 3, pp. 649–656, 2019.

[152] J. Xu, E. Gong, J. Pauly, and G. Zaharchuk, "200x low-dose PET reconstruction using deep learning," 2017, *arXiv:1712.04119*. [Online]. Available: http://arxiv.org/abs/1712.04119

[153] S. Kaplan and Y.-M. Zhu, "Full-dose PET image estimation from low-dose PET image using deep learning: A pilot study," *J. Digit. Imag.*, vol. 32, no. 5, pp. 773–778, Oct. 2019.

[154] V. Nath *et al.*, "Inter-scanner harmonization of high angular resolution DW-MRI using null space deep learning," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*. Cham, Switzerland: Springer, 2019, pp. 193–201.

[155] Z. Yang, J. Wen, and C. Davatzikos, "Smile-GANs: Semi-supervised clustering via GANs for dissecting brain disease heterogeneity from medical images," 2020, *arXiv:2006.15255*. [Online]. Available: http://arxiv.org/abs/2006.15255

[156] K. Armanious *et al.*, "MedGAN: Medical image translation using GANs," *Comput. Med. Imag. Graph.*, vol. 79, Jan. 2020, Art. no. 101684.

[157] H. Van Nguyen, S. K. Zhou, and R. Vemulapalli, "Cross-domain synthesis of medical images using efficient location-sensitive deep network," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 677–684.

[158] C. Zhao *et al.*, "A deep learning based anti-aliasing self super-resolution algorithm for MRI," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2018, pp. 100–108.

[159] A. Bustin, N. Fuin, R. M. Botnar, and C. Prieto, "From compressed-sensing to artificial intelligence-based cardiac MRI reconstruction," *Frontiers Cardiovascular Med.*, vol. 7, p. 17, Feb. 2020. [Online]. Available: https://www.frontiersin.org/article/10.3389/fcvm.2020.00017

[160] Q. Zheng, H. Delingette, and N. Ayache, "Explainable cardiac pathology classification on cine MRI with motion characterization by semi-supervised learning of apparent flow," *Med. Image Anal.*, vol. 56, pp. 80–95, Aug. 2019.

[161] A. de Marvao, T. Dawes, and D. O'Regan, "Artificial intelligence for cardiac imaging-genetics research," *Frontiers Cardiovascular Med.*, vol. 6, p. 195, Jan. 2020. [Online]. Available: https://www.frontiersin.org/article/10.3389/fcvm.2019.00195

[162] N. Parajuli *et al.*, "Flow network tracking for spatiotemporal and periodic point matching: Applied to cardiac motion analysis," *Med. Image Anal.*, vol. 55, pp. 116–135, Jul. 2019.

[163] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.

[164] P. Vu Tran, "A fully convolutional neural network for cardiac segmentation in short-axis MRI," 2016, *arXiv:1604.00494*. [Online]. Available: http://arxiv.org/abs/1604.00494

[165] W. Bai *et al.*, "Automated cardiovascular magnetic resonance image analysis with fully convolutional networks," *J. Cardiovascular Magn. Reson.*, vol. 20, no. 1, p. 65, Dec. 2018.

[166] F. Isensee, P. Jaeger, P. Full, I. Wolf, S. Engelhardt, and K. Meier-Klein, "Automatic cardiac disease assessment on cine-MRI via time-series segmentation and domain specific features," in *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges* (Lecture Notes in Computer Science). Cham, Switzerland: Springer, 2017, pp. 120–129.

[167] J. Wolterink, T. Leiner, M. Viergever, and I. Isgum, "Automatic segmentation and disease classification using cardiac cine MR images," in *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges* (Lecture Notes in Computer Science). Cham, Switzerland: Springer, 2017, pp. 101–110.

[168] X. Huang *et al.*, "Contour tracking in echocardiographic sequences via sparse representation and dictionary learning," *Med. Image Anal.*, vol. 18, no. 2, pp. 253–271, Feb. 2014.

[169] O. Oktay *et al.*, "Anatomically constrained neural networks (ACNNs): Application to cardiac image enhancement and segmentation," *IEEE Trans. Med. Imag.*, vol. 37, no. 2, pp. 384–395, Feb. 2018.

[170] Z. Xiong, V. V. Fedorov, X. Fu, E. Cheng, R. Macleod, and J. Zhao, "Fully automatic left atrium segmentation from late gadolinium enhanced magnetic resonance imaging using a dual fully convolutional neural network," *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 515–524, Feb. 2019.

[171] C. Ye, W. Wang, S. Zhang, and K. Wang, "Multi-depth fusion network for whole-heart CT image segmentation," *IEEE Access*, vol. 7, pp. 23421–23429, 2019.

[172] S. Dong *et al.*, "3D left ventricle segmentation on echocardiography with atlas guided generation and voxel-to-voxel discrimination," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, in Lecture Notes in Computer Science. Cham, Switzerland: Springer, 2018, pp. 622–629.

[173] G. Yang, T Hua, and W. Xue, "Left ventricle full quantification challenge," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2019.

[174] V. Campello *et al.*, "Multi-centre, multi-vendor & multi-disease cardiac image segmentation challenge (M&MS)," in *Proc. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*. Cham, Switzerland: Springer, 2020.

[175] K. Ta, S. Ahn, J. Stendahl, A. Sinusas, and J. Duncan, "A semi-supervised joint network for simultaneous left ventricular motion tracking and segmentation in 4D echocardiography," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, in Lecture Notes in Computer Science. Cham, Switzerland: Springer, 2020, pp. 1–10.

[176] E. Ferdian *et al.*, "Fully automated myocardial strain estimation from cardiovascular MRI–tagged images using a deep learning framework in the UK biobank," *Radiol., Cardiothoracic Imag.*, vol. 2, no. 1, Feb. 2020, Art. no. e190032, doi: 10.1148/ryct.2020190032.

[177] J. Cheng, Y.-H. Tsai, S. Wang, and M.-H. Yang, "SegFlow: Joint learning for video object segmentation and optical flow," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 686–695.

[178] G. A. Bello *et al.*, "Deep-learning cardiac motion analysis for human survival prediction," *Nature Mach. Intell.*, vol. 1, no. 2, pp. 95–104, Feb. 2019.

[179] W. Bai *et al.*, "Recurrent neural networks for aortic image sequence segmentation with sparse annotations," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger, Eds. Springer, 2018, pp. 586–594, doi: 10.1007/978-3-030-00937-3_67.

[180] Z. Guo *et al.*, "DeepCenterline: A multi-task fully convolutional network for centerline extraction," in *Proc. Inf. Process. Med. Imag. (IPMI)*, A. C. S. Chung, J. C. Gee, P. A. Yushkevich, and S. Bao, Eds. Cham, Switzerland: Springer, 2019, pp. 441–453.

[181] J. M. Wolterink, R. W. van Hamersvelt, M. A. Viergever, T. Leiner, and I. Išgum, "Coronary artery centerline extraction in cardiac CT angiography using a CNN-based orientation classifier," *Med. Image Anal.*, vol. 51, pp. 46–60, Jan. 2019.

[182] P. Moeskops *et al.*, "Deep learning for multi-task medical image segmentation in multiple modalities," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. B. Ünal, and W. Wells, Eds. Cham, Switzerland: Springer, 2016, pp. 478–486.

[183] M. C. H. Lee, K. Petersen, N. Pawlowski, B. Glocker, and M. Schaap, "TeTrIS: Template transformer networks for image segmentation with shape priors," *IEEE Trans. Med. Imag.*, vol. 38, no. 11, pp. 2596–2606, Nov. 2019.

[184] J. Wolterink, T. Leiner, and I. Išgum, "Graph convolutional networks for coronary artery segmentation in cardiac CT angiography," in *Proc. Workshop Graph Learn. Med. Imag.*, D. Zhang, L. Zhou, B. Jie, and M. Liu, Eds. Cham, Switzerland: Springer, 2019, pp. 62–69.

[185] A. L. Simpson *et al.*, "A large annotated medical image dataset for the development and evaluation of segmentation algorithms," 2019, *arXiv:1902.09063*. [Online]. Available: http://arxiv.org/abs/1902.09063

[186] K. Yan, X. Wang, L. Lu, and R. M. Summers, "Deeplesion: Automated mining of large-scale lesion annotations and universal lesion detection with deep learning," *J. Med. Imag.*, vol. 5, no. 3, p. 036501, 2018.

[187] J. J. Cerrolaza *et al.*, "Computational anatomy for multi-organ analysis in medical imaging: A review," *Med. Image Anal.*, vol. 56, pp. 44–67, Aug. 2019.

[188] H. Seo, C. Huang, M. Bassenne, R. Xiao, and L. Xing, "Modified U-net (mU-Net) with incorporation of object-dependent high level features for improved liver and liver-tumor segmentation in CT images," *IEEE Trans. Med. Imag.*, vol. 39, no. 5, pp. 1316–1325, May 2020.

[189] P. M. Graffy, V. Sandfort, R. M. Summers, and P. J. Pickhardt, "Automated liver fat quantification at nonenhanced abdominal CT for population-based steatosis assessment," *Radiology*, vol. 293, no. 2, pp. 334–342, Nov. 2019.

[190] Y.-B. Tang, Y.-X. Tang, Y. Zhu, J. Xiao, and R. M. Summers, "E^{2}Net: An edge enhanced network for accurate liver and tumor segmentation on CT scans," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, 2020, pp. 512–522.

[191] S. G. Armato *et al.*, "PROSTATEx challenges for computerized classification of prostate lesions from multiparametric magnetic resonance images," *J. Med. Imag.*, vol. 5, no. 4, 2018, Art. no. 044501.

[192] R. Cheng *et al.*, "Fully automated prostate whole gland and central gland segmentation on MRI using holistically nested networks with short connections," *J. Med. Imag.*, vol. 6, no. 2, 2019, Art. no. 024007.

[193] H. R. Roth *et al.*, "Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation," *Med. Image Anal.*, vol. 45, pp. 94–107, Apr. 2018.

[194] H. R. Roth, A. Farag, E. B. Turkbey, L. Lu, J. Liu, and R. M. Summers, "Data from pancreas-CT," in *Cancer Imag. Arch.*, 2015.

[195] H. R. Roth *et al.*, "An application of cascaded 3D fully convolutional networks for medical image segmentation," *Comput. Med. Imag. Graph.*, vol. 66, pp. 90–99, Jun. 2018.

[196] L. C. Chu *et al.*, "Application of deep learning to pancreatic cancer detection: Lessons learned from our initial experience," *J. Amer. College Radiol.*, vol. 16, no. 9, pp. 1338–1342, Sep. 2019.

[197] L. Zhang *et al.*, "Spatio-temporal convolutional LSTMs for tumor growth prediction by learning 4D longitudinal patient data," *IEEE Trans. Med. Imag.*, vol. 39, no. 4, pp. 1114–1126, Apr. 2020.

[198] G. E. Humpire-Mamani, J. Bukala, E. T. Scholten, M. Prokop, B. van Ginneken, and C. Jacobs, "Fully automatic volume measurement of the spleen at CT using deep learning," *Radiol., Artif. Intell.*, vol. 2, no. 4, Jul. 2020, Art. no. e190102.

[199] H. R. Roth *et al.*, "Improving computer-aided detection using convolutional neural networks and random view aggregation," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1170–1181, May 2016.

[200] Y.-B. Tang, Y. Ke, J. Xiao, and R. M. Summers, "One click lesion RECIST measurement and segmentation on CT scans," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, 2020, pp. 573–583.

[201] H. R. Roth *et al.*, "CT lymph nodes," *Cancer Imag. Arch.*, 2015.

[202] R. Tachibana *et al.*, "Deep learning electronic cleansing for single-and dual-energy CT colonography," *RadioGraphics*, vol. 38, no. 7, pp. 2034–2050, 2018.

[203] S. Shin, S. Lee, D. Elton, J. Gullery, and R. M. Summers, "Deep small bowel segmentation with cylindrical topological constraints," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, 2020, pp. 207–215.

[204] J. Liu *et al.*, "Detection and diagnosis of colitis on computed tomography using deep convolutional neural networks," *Med. Phys.*, vol. 44, no. 9, pp. 4630–4642, Sep. 2017.

[205] P. Rajpurkar *et al.*, "AppendiXNet: Deep learning for diagnosis of appendicitis from a small dataset of CT exams using video pretraining," *Sci. Rep.*, vol. 10, no. 1, pp. 1–7, Dec. 2020.

[206] P. M. Cheng, T. K. Tejura, K. N. Tran, and G. Whang, "Detection of high-grade small bowel obstruction on conventional radiography with convolutional neural networks," *Abdominal Radiol.*, vol. 43, no. 5, pp. 1120–1127, May 2018.

[207] C.-C. Kuo *et al.*, "Automation of the kidney function prediction and classification through ultrasound-based kidney imaging using deep learning," *NPJ Digit. Med.*, vol. 2, no. 1, pp. 1–9, Dec. 2019.

[208] S. Wang *et al.*, "Deep learning provides a new computed tomography-based prognostic biomarker for recurrence prediction in high-grade serous ovarian cancer," *Radiotherapy Oncol.*, vol. 132, pp. 171–177, Mar. 2019.

[209] D. Dreizin, Y. Zhou, Y. Zhang, N. Tirada, and A. L. Yuille, "Performance of a deep learning algorithm for automated segmentation and quantification of traumatic pelvic hematomas on CT," *J. Digit. Imag.*, vol. 33, no. 1, pp. 243–251, Feb. 2020.

[210] S. Kazemifar *et al.*, "Segmentation of the prostate and organs at risk in male pelvic CT images using deep learning," *Biomed. Phys. Eng. Exp.*, vol. 4, no. 5, Jul. 2018, Art. no. 055003.

[211] H. Li, H. Han, and S. Kevin Zhou, "Bounding maps for universal lesion detection," 2020, *arXiv:2007.09383*. [Online]. Available: http://arxiv.org/abs/2007.09383

[212] P. J. Pickhardt *et al.*, "Automated CT biomarkers for opportunistic prediction of future cardiovascular events and mortality in an asymptomatic screening population: A retrospective cohort study," *Lancet Digit. Health*, vol. 2, no. 4, pp. e192–e200, Apr. 2020.

[213] K. Bera, K. A. Schalper, D. L. Rimm, V. Velcheti, and A. Madabhushi, "Artificial intelligence in digital pathology—New tools for diagnosis and precision oncology," *Nature Rev. Clin. Oncol.*, vol. 16, no. 11, pp. 703–715, Nov. 2019.

[214] J. Xu *et al.*, "Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images," *IEEE Trans. Med. Imag.*, vol. 35, no. 1, pp. 119–130, Jan. 2016.

[215] A. Janowczyk and A. Madabhushi, "Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases," *J. Pathol. Informat.*, vol. 7, no. 1, p. 29, 2016.

[216] A. Cruz-Roa *et al.*, "Accurate and reproducible invasive breast cancer detection in whole-slide images: A deep learning approach for quantifying tumor extent," *Sci. Rep.*, vol. 7, no. 1, p. 46450, Jun. 2017.

[217] A. Cruz-Roa *et al.*, "High-throughput adaptive sampling for whole-slide histopathology image analysis (HASHI) via convolutional neural networks: Application to invasive breast cancer detection," *PLoS ONE*, vol. 13, no. 5, May 2018, Art. no. e0196828.

[218] B. E. Bejnordi *et al.*, "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," *J. Amer. Med. Assoc.*, vol. 318, no. 22, pp. 2199–2210, Dec. 2017.

[219] N. Coudray *et al.*, "Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning," *Nature Med.*, vol. 24, no. 10, pp. 1559–1567, 2018.

[220] G. Campanella *et al.*, "Clinical-grade computational pathology using weakly supervised deep learning on whole slide images," *Nature Med.*, vol. 25, no. 8, pp. 1301–1309, Aug. 2019.

[221] R. J. Babaian and W. A. Grunow, "Reliability of Gleason grading system in comparing prostate biopsies with total prostatectomy specimens," *Urology*, vol. 25, no. 6, pp. 564–567, Jun. 1985.

[222] W. Bulten *et al.*, "Automated deep-learning system for Gleason grading of prostate cancer using biopsies: A diagnostic study," *Lancet Oncol.*, vol. 21, no. 2, pp. 233–241, Feb. 2020.

[223] P. Ström *et al.*, "Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: A population-based, diagnostic study," *Lancet Oncol.*, vol. 21, no. 2, pp. 222–232, Feb. 2020.

[224] A. Madabhushi, M. D. Feldman, and P. Leo, "Deep-learning approaches for Gleason grading of prostate biopsies," *Lancet Oncol.*, vol. 21, no. 2, pp. 187–189, Feb. 2020.

[225] D. Romo-Bucheli, A. Janowczyk, H. Gilmore, E. Romero, and A. Madabhushi, "Automated tubule nuclei quantification and correlation with oncotype DX risk categories in ER+ breast cancer whole slide images," *Sci. Rep.*, vol. 6, no. 1, p. 32706, Dec. 2016.

[226] D. Romo-Bucheli, A. Janowczyk, H. Gilmore, E. Romero, and A. Madabhushi, "A deep learning based strategy for identifying and associating mitotic activity with gene expression derived risk categories in estrogen receptor positive breast cancers," *Cytometry A*, vol. 91, no. 6, pp. 566–573, Jun. 2017.

[227] J. N. Kather *et al.*, "Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer," *Nature Med.*, vol. 25, no. 7, pp. 1054–1056, Jul. 2019.

[228] O.-J. Skrede *et al.*, "Deep learning for prediction of colorectal cancer outcome: A discovery and validation study," *Lancet*, vol. 395, no. 10221, pp. 350–360, Feb. 2020.

[229] P. Courtiol *et al.*, "Deep learning-based classification of mesothelioma improves prediction of patient outcome," *Nature Med.*, vol. 25, no. 10, pp. 1519–1525, Oct. 2019.

[230] C. Saillard *et al.*, "Predicting survival after hepatocellular carcinoma resection using deep-learning on histological slides," *J. Hepatol.*, vol. 73, p. S381, Aug. 2020.

[231] N. Tomita, B. Abdollahi, J. Wei, B. Ren, A. Suriawinata, and S. Hassanpour, "Attention-based deep neural networks for detection of cancerous and precancerous esophagus tissue on histopathological slides," *JAMA Netw. Open*, vol. 2, no. 11, Nov. 2019, Art. no. e1914645.

[232] A. Heindl, I. Sestak, K. Naidoo, J. Cuzick, M. Dowsett, and Y. Yuan, "Relevance of spatial heterogeneity of immune infiltration for predicting risk of recurrence after endocrine therapy of ER+ breast cancer," *JNCI, J. Nat. Cancer Inst.*, vol. 110, no. 2, pp. 166–175, Feb. 2018.

[233] J. Saltz *et al.*, "Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images," *Cell Rep.*, vol. 23, no. 1, pp. 181–193, 2018.

[234] G. Corredor *et al.*, "Spatial architecture and arrangement of tumor-infiltrating lymphocytes for predicting likelihood of recurrence in early-stage non–small cell lung cancer," *Clin. Cancer Res.*, vol. 25, no. 5, pp. 1526–1534, Mar. 2019.

[235] Q. Yao, Z. He, Y. Lin, K. Ma, Y. Zheng, and S. Kevin Zhou, "A hierarchical feature constraint to camouflage medical adversarial attacks," 2020, *arXiv:2012.09501*. [Online]. Available: http://arxiv.org/abs/2012.09501

[236] D. Rueckert and J. A. Schnabel, "Model-based and data-driven strategies in medical image computing," *Proc. IEEE*, vol. 108, no. 1, pp. 110–124, Jan. 2020.

[237] M. Brown *et al.*, "Integration of chest CT CAD into the clinical workflow and impact on radiologist efficiency," *Acad. Radiol.*, vol. 26, no. 5, pp. 626–631, May 2019.

[238] B. de Hoop *et al.*, "Computer-aided detection of lung cancer on chest radiographs: Effect on observer performance," *Radiology*, vol. 257, no. 2, pp. 532–540, Nov. 2010.

[239] W. Jorritsma, F. Cnossen, and P. M. A. van Ooijen, "Improving the radiologist–CAD interaction: Designing for appropriate trust," *Clin. Radiol.*, vol. 70, no. 2, pp. 115–122, Feb. 2015.

[240] T. Drew, C. Cunningham, and J. M. Wolfe, "When and why might a computer-aided detection (CAD) system interfere with visual search? An eye-tracking study," *Acad. Radiol.*, vol. 19, no. 10, pp. 1260–1267, Oct. 2012.

[241] O. Gozes *et al.*, "Rapid AI development cycle for the coronavirus (COVID-19) pandemic: Initial results for automated detection & patient monitoring using deep learning CT image analysis," 2020, *arXiv:2003.05037*. [Online]. Available: https://arxiv.org/abs/2003.05037

[242] F. Shi *et al.*, "Review of artificial intelligence techniques in imaging data acquisition, segmentation, and diagnosis for COVID-19," *IEEE Rev. Biomed. Eng.*, vol. 14, pp. 4–15, 2021.

[243] H. Greenspan, R. S.-J. Estepar, W. Niessen, E. Siegel, and M. Nielsen, "Position paper on COVID-19 imaging and AI: From the clinical needs and technological challenges to initial ai solutions at the lab and national level towards a new era for ai in healthcare," *Med. Image Anal.*, vol. 66, Dec. 2020, Art. no. 101800.

[244] S. Rathore, M. Habes, M. A. Iftikhar, A. Shacklett, and C. Davatzikos, "A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages," *NeuroImage*, vol. 155, pp. 530–548, Jul. 2017.

[245] J. M. Mateos-Pérez, M. Dadar, M. Lacalle-Aurioles, Y. Iturria-Medina, Y. Zeighami, and A. C. Evans, "Structural neuroimaging as clinical predictor: A review of machine learning applications," *NeuroImage, Clin.*, vol. 20, pp. 506–522, Jan. 2018.

[246] A. N. Nielsen, D. M. Barch, S. E. Petersen, B. L. Schlaggar, and D. J. Greene, "Machine learning with neuroimaging: Evaluating its applications in psychiatry," *Biol. Psychiatry, Cognit. Neurosci. Neuroimag.*, vol. 5, no. 8, pp. 791–798, Aug. 2020.

[247] A. F. Kazerooni, S. Bakas, H. S. Rad, and C. Davatzikos, "Imaging signatures of glioblastoma molecular characteristics: A radiogenomics review," *J. Magn. Reson. Imag.*, vol. 52, no. 1, pp. 54–69, Jul. 2020.