Deep Learning Applications in Medical Image Analysis

Justin Ker, Lipo Wang, Jai Rao, and Tchoyoson Lim

Abstract—The tremendous success of machine learning algorithms at image recognition tasks in recent years intersects with a time of dramatically increased use of electronic medical records and diagnostic imaging. This review introduces machine learning algorithms as applied to medical image analysis, focusing on convolutional neural networks, and emphasizing clinical aspects of the field. The advantage of machine learning in an era of medical big data is that significant hierarchal relationships within the data can be discovered algorithmically without laborious hand-crafting of features. We cover key research areas and applications of medical image classification, localization, detection, segmentation and registration. We conclude by discussing research obstacles, emerging trends and possible future directions.

Index Terms—Convolutional neural networks, Medical image analysis, Machine learning, Deep Learning

I. Introduction[1]

Machine learning algorithms have the potential to be invested deeply in all fields of medicine, from drug discovery to clinical decision making, significantly altering the way medicine is practiced. The success of machine learning algorithms at computer vision tasks in recent years comes at an opportune time when medical records are increasingly digitalized. The use of electronic health records (EHR) quadrupled from 11.8% to 39.6% amongst office-based physicians in the US from 2007 to 2012 [1]. Medical images are an integral part of a patient's EHR and are currently analyzed by human radiologists, who are limited by speed, fatigue, and experience. It takes years and great financial cost to train a qualified radiologist, and some health-care systems outsource radiology reporting to lower-cost countries such as India via tele-radiology. A delayed or erroneous diagnosis causes harm to the patient. Therefore, it is ideal for medical image analysis to be carried out by an automated, accurate and efficient machine learning algorithm.

Medical image analysis is an active field of research for machine learning, partly because the data is relatively structured and labelled, and it is likely that this will be the area where patients first interact with functioning, practical artificial intelligence systems. This is significant for two reasons. Firstly,

in terms of actual patient metrics, medical image analysis is a litmus test as to whether artificial intelligence systems will actually improve patient outcomes and survival. Secondly, it provides a testbed for human-AI interaction, of how receptive patients will be towards health-altering choices being made, or assisted by a non-human actor.

A. Types of Medical Imaging

There is a myriad of imaging modalities, and the frequency of their use is increasing. Smith-Bindman *et al.* [2] looked at imaging use from 1996 to 2010 across six large integrated healthcare systems in the United States, involving 30.9 million imaging examinations. The authors found that over the study period, CT, MRI and PET usage increased 7.8%, 10% and 57% respectively.

Modalities of digital medical images include ultrasound (US), X-ray, computed tomography (CT) scans and magnetic-resonance imaging (MRI) scans, positron emission tomography (PET) scans, retinal photography, histology slides, and dermoscopy images. Fig. 1. shows some example medical images. Some of these modalities examine multiple organs (such as CT, MRI) while others are organ specific (retinal photography, dermoscopy). The amount of data generated from each study also varies. A histology slide is an image file of a few megabytes while a single MRI may be a few hundred megabytes. This has technical implications on the way the data is pre-processed, and on the design of an algorithm's architecture, in the context of processor and memory limitations.

B. History of medical image analysis

The symbolic AI paradigm of the 1970s led to the development of rule-based, expert systems. One early implementation in medicine was the MYCIN system by Shortliffe [3], which suggested different regimes of antibiotic therapies for patients. Parallel to these developments, AI algorithms moved from heuristics-based techniques to manual, handcrafted feature extraction techniques. and then to supervised learning techniques. Unsupervised machine learning methods are also being researched, but the majority of the algorithms from 2015-2017 in the published literature have employed supervised learning methods, namely Convolutional Neural Networks (CNN) [4]. Aside from the availability of large labelled data sets being available, hardware advancements in Graphical Processing Units (GPUs) have also led to improvements in CNN performance, and their widespread use in medical image analysis.

Justin Ker is with Neurosurgery, National Neuroscience Institute, 11 Jalan Tan Tock Seng, Singapore, e-mail: justin.ker@mohh.com.sg

Lipo Wang is with School of Electrical and Electronic Engineering, Nanyang Technological University, Block S2.1, 50 Nanyang Avenue, 639798, Singapore, e-mail: ELPWang@ntu.edu.sg (Corresponding author).

Jai Rao is with Neurosurgery, National Neuroscience Institute

Tchoyoson Lim is with Neuro-Radiology, National Neuroscience Institute

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2788044, IEEE Access

2

McCulloch and Pitts [5] described the first artificial neuron in 1943, which developed into the perceptron posited by Rosenblatt in 1958 [6]. In essence, an artificial neural network is a layer of connected perceptrons linking inputs and outputs, and deep neural networks are multiple layers of artificial neural networks. The advantage of a deep neural network is its ability to automatically learn significant low level features (such as lines or edges), and amalgamate them to higher level features (such as shapes) in the subsequent layers. Interestingly, this is how the mammalian and human visual cortices are thought to process visual information and recognize objects [7]. CNNs may have their origins in the Neocognitron concept proposed by Fukushima in 1982 [8], but it was Lecun *et al.* [9] who formalized CNNs and used the error backpropagation described by Rumelhart, Hinton and Williams [10], to successfully perform the automatic recognition of hand-written digits. The widespread use of CNNs in image recognition came about after Krizhevsky *et al.* [11] won the 2012 Imagenet Large Scale Visual Recognition Challenge (ILSVRC) with a CNN that had a 15% error rate. The runner up had almost double the error rate at 26%. Krizhevsky *et al.* introduced significant concepts that are used in CNNs today, including the use of Rectified Linear Unit (RELU) functions in CNNs, data augmentation and dropout. Since then, CNNs have featured as the most used architecture in every ILSVRC competition, surpassing human performance at recognizing images in 2015. Correspondingly, there has been a dramatic increase in the number of research papers published on CNN architecture and applications, such that CNNs have become the dominant architecture in medical image analysis.

C. Convolutional neural networks

Both the 2-dimensional and 3-dimensional structures of an organ being studied are crucial in order to identify what is normal versus abnormal. By maintaining these local spatial relationships, CNNs are well-suited to perform image recognition tasks. CNNs have been put to work in many ways, including image classification, localization, detection, segmentation and registration. CNNs are the most popular machine learning algorithm in image recognition and visual learning tasks, due to its unique characteristic of preserving local image relations, while performing dimensionality reduction. This captures important feature relationships in an image (such as how pixels on an edge join to form a line), and reduces the number of parameters the algorithm has to compute, increasing computational efficiency. CNNs are able to take as inputs and process both 2-dimensional images, as well as 3-dimensional images with minor modifications. This is a useful advantage in designing a system for hospital use, as some modalities like X-rays are 2-dimensional while others like CT or MRI scans are 3-dimensional volumes.

CNNs and Recurrent Neural Networks (RNNs) are examples of supervised machine learning algorithms, which require significant amounts of training data. Unsupervised learning algorithms have also been studied for use in medical image analysis. These include Autoencoders, Restricted Boltzmann Machines (RBMs), Deep Belief Networks (DBNs), and Generative Adversarial Networks (GANs).

D. Resources

Four reviews are highly recommended; Litjens *et al.* [4] provides a thorough list of papers published in the field, Shen *et al.* [12] and Suzuki *et al.* [13] summarize many of the advances, while Greenspan *et al.* [14] gives a succinct overview of recent important papers. These review articles and a list of relevant books can be found in Table 1. This was collated by searching for books in the Elsevier, IEEE Xplore and Springer databases. We generated a list of the 200 most highly-cited papers from Google Scholar, using the query terms 'deep learning' and 'medical image analysis' in October 2017 using citation software [15]. These were manually vetted to ensure that the returned results were relevant and significant in the field. We limited the papers to those published or prepublished in the last 3 years, although older significant papers are mentioned in this article. Table 2 shows the top 20 papers from this list, and the full list of 200 papers can be found in the Supplementary Data as Table S1.

Where available, the datasets used by the authors of various papers in this article are described. The website Grand Challenges in Biomedical Image Analysis (https://grand-challenge.org/all_challenges) aggregates and links to numerous competitions and their respective image datasets. The Cancer Imaging Archive [16] contains numerous datasets across many organ systems, and the National Institute of Health recently released a tranche of over 100,000 anonymized chest x-rays for research use [17] called "ChestX-ray 8". Of note, Nifty-Net (www.niftynet.io) [18] is a useful open source framework that contains many machine learning algorithms, released under an Apache License. It allows researchers to explore CNNs and published machine algorithms, such as V-net, U-net, DeepMedic [19], [20], [21], and to share pre-trained models.

The aim of this report is to provide an overview on the state of machine learning algorithms as applied to medical imaging, with an emphasis on which aspects are most useful to the clinician, as some of the authors are practicing surgeons and radiologists. It is hoped that this perspective aids researchers in moving from being trapped in the local minima of speculative research, to designing implementable systems that will impact medical science and patient care.

Section II describes various machine learning architectures used in medical image analysis, with an emphasis on CNNs. Machine learning is broadly classified into Supervised, Unsupervised, Semi-supervised and Reinforcement learning methods; it is the first two which are currently most applicable to image analysis. **Section III** dives into different application areas. **Section IV** concludes with obstacles that the field of medical image analysis faces, and some of the future possible directions.

II. Machine learning Architectures

A. Supervised learning models

1) Convolutional neural networks:

Currently, CNNs are the most researched machine learning algorithms in medical image analysis [4]. The reason for this is that CNNs preserve spatial relationships when filtering input

images. As mentioned, spatial relationships are of crucial importance in radiology, for example, in how the edge of a bone joins with muscle, or where normal lung tissue interfaces with cancerous tissue. As shown in Fig. 2., a CNN takes an input image of raw pixels, and transforms it via Convolutional Layers, Rectified Linear Unit (RELU) Layers and Pooling Layers. This feeds into a final Fully Connected Layer which assigns class scores or probabilities, thus classifying the input into the class with the highest probability.

a) Convolution Layer: A convolution is defined as an operation on two functions. In image analysis, one function consists of input values (e.g. pixel values) at a position in the image, and the second function is a filter (or kernel); each can be represented as array of numbers. Computing the dot

product between the two functions gives an output. The filter is then shifted to the next position in the image as defined by the *stride* length. The computation is repeated until the entire image is covered, producing a feature (or activation) map. This is a map of where the filter is strongly activated and 'sees' a feature such as a straight line, a dot, or a curved edge. If a photograph of a face was fed into a CNN, initially low-level features such as lines and edges are discovered by the filters. These build up to progressively higher features in subsequent layers, such as a nose, eye or ear, as the feature maps become inputs for the next layer in the CNN architecture.

Convolution exploits three ideas intrinsic to perform computationally efficient machine learning: sparse connections, parameter sharing (or weights sharing) and equivariant (or invariant) representation [22]. Unlike some neural networks where every input neuron is connected to every output neuron in the subsequent layer, CNN neurons have sparse connections, meaning that only some inputs are connected to the next layer. By having a small, local receptive field (i.e., the area covered by the filter per stride), meaningful features can be gradually learnt, and the number of weights to be calculated can be drastically reduced, increasing the algorithm's efficiency. In using each filter with its fixed weights across different positions of the entire image, CNNs reduce memory storage requirements. This is known as parameter sharing. This is in contrast to a fully connected neural network where the weights between layers are more numerous, used once and then discarded. Parameter sharing results in the quality of equivariant representation to arise. This means that input translations result in a corresponding feature map translation. The convolution operation is defined by the * symbol. An output (or feature map) *s(t)* is defined below when input *I(t)* is convolved with a filter or kernel *K(a)*.

$$s(t) = (I * K)(t). \qquad (1)$$

If *t* can only take integer values, the **discretized convolution** is given by:

$$s(t) = \sum_a I(a) \cdot K(t - a). \qquad (2)$$

The above assumes a one-dimensional convolutional operation. A two dimension convolution operation with input I

(m, n) and a kernel K (a, b) is defined as:

$$s(t) = \sum_a \sum_b I(a, b) \cdot K(m - a, n - b). \qquad (3)$$

By the commutative law, the kernel is flipped and the above is equivalent to:

$$s(t) = \sum_a \sum_b I(m - a, n - b) \cdot K(a, b). \qquad (4)$$

Neural networks implement the **cross-correlation** function, which is the same as convolution but without flipping the kernel.

$$s(t) = \sum_a \sum_b I(m + a, n + b) \cdot K(a, b). \qquad (5)$$

b) Rectified Linear Unit (RELU) Layer: The RELU layer is an activation function that sets negative input values to zero. This simplifies and accelerates calculations and training, and helps to avoid the vanishing gradient problem. Mathematically it is defined as:

$$f(x) = max(0, x). \qquad (6)$$

where *x* is the input to the neuron. Other activation functions include the sigmoid, tanh, leaky RELUs, Randomized RELUs and parametric RELUs.

c) Pooling Layer: The Pooling layer is inserted between the Convolution and RELU layers to reduce the number of parameters to be calculated, as well as the size of the image (width and height, but not depth). Max-pooling is most commonly used; other pooling layers include Average pooling and L2-normalization pooling. Max-pooling simply takes the largest input value within a filter and discards the other values; effectively it summarizes the strongest activations over a neighborhood. The rationale is that the relative location of a strongly activated feature to another is more important than its exact location.

d) Fully Connected Layer: The final layer in a CNN is the Fully Connected Layer, meaning that every neuron in the preceding layer is connected to every neuron in the Fully Connected Layer. Like the convolution, RELU and pooling layers, there can be 1 or more fully connected layers depending on the level of feature abstraction desired. This layer takes the output from the preceding layer (Convolutional, RELU or Pooling) as its input, and computes a probability score for classification into the different available classes. In essence, this layer looks at the combination of the most strongly activated features that would indicate the image belongs to a particular class. For example, on histology glass slides, cancer cells have a high DNA to cytoplasm ratio compared to normal cells. If features of DNA were strongly detected from the preceding layer, the CNN would be more likely to predict the presence of cancer cells. Standard neural network training methods with backpropagation [10] and stochastic gradient descent help the CNN learn important associations from training images.

2) Transfer learning with CNNs: Unlike general natural image recognition tasks, medical image analysis lacks large labelled training datasets. As a comparison, the Kaggle 2017 Data Science Bowl to detect tumors in CT lung scans had a dataset of approximately 2000 patient scans, while ILSVRC 2017 had over 1 million images across 1000 object classes [23]. Transfer learning involves training a machine learning algorithm on a partially-related or un-related dataset, as well as a labelled training dataset, to circumvent the obstacle of insufficient training data. Essentially the weights learned or pre-trained during the training of a CNN on one (partially-related or un-related) dataset are transferred to a second CNN, which is then trained on labelled medical data using these weights. The weights can be applied to some or all layers of the CNN, except the last fully connected layer. Although transfer learning techniques are commonly used in medical image analysis in conjunctions with CNNs, it is worth noting that they can be applied to other general machine learning algorithms as well.

Shin *et al.* [24] explored the impact of CNN architectures and transfer learning on detecting the presence of enlarged thoraco-abdominal lymph nodes, and in classifying interstitial lung disease on CT scans, and found transfer learning to be beneficial, despite natural images being disparate from medical images. Ravishankar *et al.* [25] looked at the task of automatically localizing the presence of a kidney on ultrasound images. Using a CNN pre-trained on Imagenet, they showed that the greater the degree of transfer learning, the better the CNN performed. Tajbakhsh *et al.* [26] studied the effectiveness of transfer learning in 4 different applications across 3 imaging modalities: polyp detection on colonoscopy videos, colonoscopy video frame classification, pulmonary embolus detection on CT pulmonary angiograms, and segmentation of the layers of the walls of the carotid artery on ultrasound scans. They transferred pre-trained weights from Alexnet [11] to either a few ('shallow tuning') or many ('deep tuning') layers in a CNN. Overall, they found that transfer learning more layers improved the CNN performance, compared to training a CNN from scratch. In contrast to many computer vision tasks where shallow tuning of the last few layers is adequate, medical image analysis requires a deep tuning of more layers. They also noted that the number of optimal layers trained varied between different applications.

3) Recurrent neural networks (RNNs): RNNs have traditionally been used in analyzing sequential data, such as the words in a sentence. Due to their ability to generate text [27], RNNs have been employed in text analysis tasks, like machine translation, speech recognition, language modelling, text prediction and image caption generation [28]. In a plain RNN, the output of a layer is added to the next input, and this is fed back into the layer, resulting in a capacity for contextual 'memory'. To avoid vanishing gradient problems with backpropagation through time, plain RNNs have evolved into Long Short Term Memory (LSTM) networks and Gated-Recurrent Units (GRUs). These are modifications of RNNs to hold long term dependencies, and to discard or forget some of the accumulated information.

In the medical image analysis space, RNNs have been used mainly in segmentation. Chen *et al.* [29] combined CNN and RNN to segment neuronal and fungal structures from three-dimensional electron microscope images. Using a multi-dimensional LSTM, Stollenga *et al.* [30] segmented both three-dimensional electron microscope images of neurons as well as MRI brain scans. Shin *et al.* [31] describe annotating X-ray images with captions trained on radiology reports.

B. Unsupervised learning models

1) Autoencoders: Autoencoders learn feature representations of input data (called *codings*) in an unsupervised manner without labelled data. It is a model that takes input data, gleans codings from this, and then uses these codings to reconstruct output data (called *reconstructions*). The rationale behind autoencoders is that the output data must be as similar to the input data as possible, i.e., autoencoder models contain a cost function which penalizes the model when inputs and outputs are different. Autoencoders have several useful features. Firstly, they are employed as feature detectors that can learn codings in an unsupervised manner, without training labels. Secondly, they reduce the model dimensionality and complexity as codings often exist in a lower dimension. Thirdly, by having to reconstruct outputs, autoencoders generate new data that is similar to the input training data. These features are an advantage in medical image analysis, where labelled training data is scarce. Various network architectures are depicted in Fig. 3.

The unique architectural feature of autoencoders is that the number of neurons in the input and output layers must be equal. Autoencoders have hidden layers that can be stacked, like CNNs. **Stacked autoencoders** (SAEs) have a typically symmetrical architecture, with a line of reflection through the middle =hidden layer. Some techniques for optimizing autoencoder performance include tying weights of the decoder layer to the encoder layer, training different subsets of autoencoders separately before stacking them together, and transfer learning [32]. Simply stacking more layers may not aid in the accuracy of the model, as the model may end up performing the trivial task of simply copying the input to the output. That is, the model performs well during training but it has not learned any useful feature representations that allow the model to be generalized and applied outside of the training data.

To force models to learn useful representations, constraints need to be added. One example is the **Denoising Autoencoder** reported by Vincent *et al.* [32], where Gaussian noise is added to the early hidden layers. Applying dropout i.e., randomly turning off some of the neurons in the early hidden layers, accomplishes the same goal, by forcing the model to learn useful codings to generate back the noise-free inputs in the output layer. A second example are **Sparse Autoencoders** [33], whereby a defined proportion of the neurons in the hidden layers are deactivated or set to zero. This is accomplished by having a cost function that penalizes the model when there

are active neurons beyond a defined threshold. The rationale behind this, is as Bengio states, for a given observation, only a small fraction of the possible factors are relevant, meaning that much of the features extracted from the data could be represented by being set to zero [34]. Kallenberg *et al.* [35] combined unsupervised convolution layers trained as autoencoders, and supervised layers to classify mammograms into different densities and textures. The texture classification task was used to impute if a mammogram was normal or depicted breast cancer. They used 2700 mammograms from the Dutch Breast Cancer Screening Program, the Dutch Breast Cancer Screening dataset, and the Mayo Mammography Health Study. Interestingly, they used a sparse autoencoder to learn the parameters of the feature-extracting convolution layers, before this input was fed into a softmax classifier. For the cancer classification task, this Convolution stacked autoencoder (CSAE) model obtained an AUC score of 0.57, which the authors reported as state of the art.

Variational Autoencoders (VAEs) are an emerging and popular unsupervised learning architecture described by Kingma et al [36]. VAEs are a generative model, consisting of a Bayesian inference encoder network and a decoder network, that can be trained with stochastic gradient descent. The encoder network projects input data into latent space variables, whose true distribution is approximated using a Gaussian distribution. The decoder network then maps the latent space back into output data, trained and guided by two cost functions: a reconstruction loss function and the Kullback–Leibler divergence.

2) Restricted Boltzmann Machines and Deep Belief Networks: Boltzmann machines were invented by Ackley *et al.* [37] in 1985, and were modified as **Restricted Boltzmann Machines** (RBMs) a year later by Smolensky [38]. RBMs are generative, stochastic, probabilistic, bidirectional graphical models consisting of visible and hidden layers [22]. These layers are connected to each other but there are no connections within the layers themselves. RBMs use the backward pass of input data to generate a reconstruction, and estimate the probability distribution of the original in- put. van Tulder *et al.* [39] modified RBMs into what they described as convolutional RBMs to classify lung tissue into normal, emphysematous, fibrosed, micronodular, or ground glass tissue. For this task, they used the CT chest scans of 128 patients with interstitial lung disease from the ILD database. Convolutional RBMs were trained with either purely discriminative, purely generative, or mixed-discriminative and generative learning objectives to learn filters. These filters were then used to perform feature extraction and create feature activation maps, before classification using a random forest classifier. Classification accuracies of between 41% to 68% were obtained, depending on the proportion of generative learning and the input patch size. They also found that filters generated from mixed-discriminative and generative learning performed the best, concluding that discriminative learning could help unsupervised feature extractors learn filters optimized for classification tasks.

RBMs can be efficiently trained with Contrast-Divergence algorithms [40] and stacked into **Deep Belief Networks**

(DBNs), where the hidden layer output of a RBM becomes the input for the visible layer of a second RBM stacked on it. DBNs were described by Hinton, Osindero and Teh in 2006 in a seminal paper [41], which was largely responsible for the renaissance in deep learning. The insight from Hinton *et al.* was that DBNs could be trained in a greedy, layer- by-layer fashion [42], with lower layers learning low level features, and progressively higher layers learning high level features, mirroring real world data hierarchy. DBNs can also be coupled to layers of supervised RBMs to produce a semi-supervised deep learning architecture. An application of RBMs was reported by Khatami *et al.* [43], who used DBNs to classify x-ray images into 5 classes of anatomic areas and orientations.

3) Generative Adversarial Networks: Generative Adversarial Networks (GANs) [44] represent a type of unsupervised learning which holds promise for medical image analysis tasks. As its name suggests, a GAN is a generative model, and is similar to a VAE in that respect. GANs comprise of two simultaneously-trained, competing models, which may be multilayer perceptrons such as CNNs. The models may be described as two players competing in a zero-sum game. One CNN is a generator that generates artificial training images. The other CNN is called a discriminator, which classifies if images are real training images or artificial ones from the generator. The desired end-point of this adversarial arrangement is one where the discriminator is unable to tell the difference between a real and a generated image i.e., the probability of assigning an image to either data distribution is 1/2. An advantage is that both generator and discriminator can be trained with backpropagation and dropout, without unwieldly inference and Markov chains. GANs are relatively new but some applications in brain MRI segmentation and synthetic medical data generation are discussed in Section 4.1 below.

III. APPLICATIONS IN MEDICAL IMAGE ANALYSIS

To the researcher, CNNs have been put to task for classification, localization, detection, segmentation and registration in image analysis. Machine learning research draws a distinction between localization (draw a bounding box around a single object in the image), and detection (draw bounding boxes around multiple objects, which may be from different classes). Segmentation draws outlines around the edges of target objects, and labels them (semantic segmentation). Registration refers to fitting one image (which may be 2 or 3 dimensional) onto another. This separation of tasks is based on different machine learning techniques and is maintained below.

To the clinician this separation of tasks is not that crucial, and it is the authors' opinion that a pragmatic machine learning system will incorporate some or all of the tasks into a unified system. It would be ideal to, in a single workflow, detect a lung tumor on a CT chest scan, and then localize and segment it away from normal tissue, and to prognosticate various treatment options, such as chemotherapy or surgery. Indeed, some of these tasks blur into one another in the papers discussed here. From the clinician's perspective, classification ascertains if a disease state is present or not, i.e., is blood present on this

MRI brain scan signifying a hemorrhagic stroke? Localization implies the identification of normal anatomy, for example, where is the kidney in this ultrasound image? This is in contrast to detection, which implies an abnormal, pathological state, for example, where are all the lung tumors in this CT scan of the lung? Segmenting the outline of a lung tumor helps the clinician determine its distance from major anatomical structures, and helps to answer a question such as, should this patient be operated on, and if so, what should be the extent of resection?

A. Classification

Classification is sometimes also known as Computer-Aided Diagnosis (CADx). Lo et al. described a CNN to detect lung nodules on chest X-rays as far back as 1995 [45]. They used 55 chest x-rays and a CNN with 2 hidden layers to output whether or not a region had a lung nodule. The relative availability of chest x-ray images has likely accelerated deep learning progress in this modality. Rajkomar et al. [46] augmented 1850 chest x-ray images into 150,000 training samples. Using a modified pre-trained GoogLeNet CNN [47], they classified the orientation of the images into frontal or lateral views with near 100% accuracy. Although this task of identifying the orientation of the chest x-ray is of limited clinical use, it does demonstrate the effectiveness of pre-training, and data augmentation in learning the relevant image metadata, as part of an eventually fully-automated diagnostic work-flow. Pneumonia or chest infection is a common health-problem world-wide that is eminently treatable. Rajpurkar et al. [48] employed a modified DenseNet [49] with 121 convolutional layers called CheXNet to classify 14 different diseases seen on the chest x-rays, using 112,000 images from the ChestX-ray14 [17] dataset. CheXNet achieved state of the art performance in classifying the 14 diseases; pneumonia classification in particular achieved an Area Under Curve (AUC) score of 0.7632 with Receiver Operating Characteristics (ROC) analysis. Moreover, on a test set of 420 images, CheXNet matched or bettered the performance of 4 individual radiologists, and also the performance of a panel comprising of 3 radiologists. Shen et al. [50] used CNNs combined with Support Vector Machine (SVM) and Random Forest (RF) classifiers to classify lung nodules into benign or malignant, based on 1010 labelled CT lung scans from the Lung Image Database Consortium (LIDC-IDRI) dataset. They used 3 parallel CNNs with 2 convolution layers each, with each CNN taking image patches at different scales to extract features. The learned features were used to construct an output feature vector, which was then classified using either a SVM with radial basis function (RBF) filter or RF classifier into benign or malignant. Their method classified nodules with 86% accuracy and they also found that it was robust against different levels of noise inputs. Li et al. [51] used 3-dimensional CNNs to interpolate missing imaging data between MRI and PET images. 830 patients with MRI and PET scans from the Alzheimer Disease Neuroimaging Initiative (ADNI) database were studied. 3-D CNNs were trained with MRI and PET images as input and output respectively, and used to reconstruct PET images from patients who

did not have them. Their reconstructed PET images almost matched ground truth results of disease classification, but one caveat is that issues of overfitting were not addressed, limiting the potential generalizability of their technique. Hosseini-Asl et al. [52] achieved state of the art results in diagnosing patients with Alzheimer's Disease versus normal, with an accuracy of 99%. They employed 3-D CNNs in an auto-encoder architecture, pre-trained on the CADDementia dataset to learn generic brain structural features. The learned feature outputs were then connected to higher layers where deep supervision techniques fine-tuned the algorithm's ability to discriminate between scans of patients with normal brains, mild cognitive impairment, or Alzheimer's Disease from the ADNI database. Korolev et al. [53] evaluated the performance of their VOXCNN and ResNet, which was based on the VGGNet [54] and Residual neural network architectures [55] respectively. They also used the ADNI database to discriminate between normal and Alzheimer Disease patients. Although their accuracy of 79% for Voxnet and 80% for ResNet was lower than what Hosseini-Asl achieved, Korolev states that their algorithms did not need hand-crafting of features and were simpler to implement.

Diabetic retinopathy (DR) can also be diagnosed using CNNs. Using digital photographs of the fundus of the eye, Pratt et al. [56] trained a CNN with 10 convolutional layers and 3 fully connected layers on approximately 90,000 fundus images. They classified DR into 5 clinically used classifications of DR severity, with 75% accuracy. Abramoff et al. [57] evaluated a commercial device, the IDx-DR version X2.1 (IDx LLC, Iowa City, Iowa, USA) to detect DR. The author does not disclose the CNN architectures but states they are inspired by Alexnet and VGGNet. The device, trained on up to 1.2 million DR images, obtained an AUC score of 0.98.

Unsupervised learning methods are also an active area of research. Pils et al. [58] used Deep Belief Networks to extract features from functional fMRI (fMRI) images, and MRI scans of patients with Huntington Disease and Schizophrenia. Suk et al. [59] classified fMRI images into diagnoses of Healthy or Mild Cognitive Impairment, using a stacked architecture of RBMs to learn hierarchal functional relationships between different brain regions. Looking outside the usual CNN models, Kumar et al. [60] compared the performance of the well-known CNNs Alexnet and VGGNet to other techniques, namely Bag of Visual Words (BOVV) and Local Binary Patterns (LBP). Interestingly, the BOVV technique performed the best at classifying histopathological images into 20 different tissue types.

B. Localization

Localization of normal anatomy is less likely to interest the practicing clinician although applications may arise in anatomy education. Alternatively, localization may find use in fully automated end-to-end applications, whereby the radiological image is autonomously analyzed and reported without any human intervention. Yan et al. [61] looked at transverse CT image slices and constructed a two stage CNN where the first stage identified local patches, and the second

stage discriminated the local patches by various body organs, achieving better results than a standard CNN. Roth *et al.* [62] trained a CNN with 5 convolution layers to discriminate approximately 4000 transverse axial CT images into one of 5 categories: neck, lung, liver, pelvis, legs. He was able to achieve a 5.9% classification error rate and an AUC score of 0.998, after data augmentation techniques. Shin *et al.* [63] used stacked autoencoders on 78 contrast-enhanced MRI scans of the abdominal region containing liver or kidney metastatic tumors, to detect the locations of the liver, heart, kidney and spleen. Hierarchal features were learned over the spatial and temporal domains, giving detection accuracies of between 62% and 79%, depending on the organ.

C. Detection

Detection, sometimes known as Computer-Aided Detection (CADe) is a keen area of study as missing a lesion on a scan can have drastic consequences for both the patient and the clinician. The task for the Kaggle Data Science Bowl of 2017 [64] involved the detection of cancerous lung nodules on CT lung scans. Approximately 2000 CT scans were released for the competition and the winner Liao *et al.* achieved a logarithmic loss score of 0.399 [65]. Their solution used a 3-D CNN inspired by U-Net architecture [19] to isolate local patches first for nodule detection. Then this output was fed into a second stage consisting of 2 fully connected layers for classification of cancer probability. Shin *et al.* [24] evaluated five well-known CNN architectures in detecting thoraco-abdominal lymph nodes and Interstitial lung disease on CT scans. Detecting lymph nodes is important as they can be a marker of infection or cancer. They achieved a mediastinal lymph node detection AUC score of 0.95 with a sensitivity of 85% using GoogLeNet, which was state of the art. They also documented the benefits of transfer learning, and the use of deep learning architectures of up to 22 layers, as opposed to fewer layers which was the norm in medical image analysis. Overfeat was a CNN pre-trained on natural images that won the ILSVRC 2013 localization task [66]. Ciompi *et al.* [67] applied Overfeat to 2-dimensional slices of CT lung scans oriented in the coronal, axial and sagittal planes, to predict the presence of nodules within and around lung fissures. They combined this approach with simple SVM and RF binary classifiers, as well as a Bag of Frequencies [68], a novel 3-dimensional descriptor of their own invention.

Other than lung lesions, there are also a myriad of other applications, including detecting malignant skin cells. Esteva *et al.* [69] used 130,000 dermatological photographs and dermoscopic images to train a GoogLeNet Inception V3 CNN, with no hand-crafting of features. The CNN outperformed human dermatologists in classifying the images as benign, malignant or non-neoplastic lesions, reaching an accuracy of 72% compared to the 65% and 66% accuracies obtained by 2 human dermatologists. The CNN again bettered 21 human dermatologists at deciding treatment plans for two types of skins cancers: carcinoma and melanoma. This task involved 376 biopsy-proven images, and the CNN achieved AUC scores of between 0.91 to 0.96.

Histopathological images are increasingly digitized and this has led to numerous papers in this field. Currently these images are laboriously read by human pathologists who look for markers of malignancy such as: increased nucleus to cytoplasm ratios, increased numbers of mitotic figures indicating increased cell replication, atypical cellular architecture, signs of cellular necrosis, high cell proliferation index from molecular markers like Ki-67. A histopathological slide can contain hundreds to thousands of cells, and wading through them at high magnification carries the risk of missing aberrant neoplastic areas. Ciresan *et al.* [70] used 11-13 layer CNNs to identify mitotic figures in 50 breast histology images from the MITOS dataset. Their approach achieved precision and recall scores of 0.88 and 0.70 respectively. More recently, Yang *et al* [71]. achieved accuracies of 97-98% in classifying kidney cancer histopathological images into tumor or non-tumor, using CNNs that were 5-7 layers deep. Sirinukunwattana *et al.* [72] also used CNNs, but to detect cell nuclei in 100 colorectal adenocarcinoma histology images. For training, almost 30,000 nuclei had to be hand-labelled. The novelty in their approach was the use of their Spatially-Constrained CNN, which used spatial regression and the surrounding spatial context to identify the centers of nuclei. Nuclei in breast cancer histological slides were also identified by Xu *et al.* [73], although he used a Stacked Sparse Autoencoder (SSAE) instead. Their model obtained precision and recall scores of 0.89 and 0.83 respectively, showing that unsupervised learning methods can also be successfully employed in this field. Albarquoni et al. [74], deal with the issue of insufficient labelling of medical images, by "crowd-sourcing" the ground-truth labelling of mitoses in breast cancer histology images to non-experts on the internet. The crowd-sourced input labels were fed into a CNN, and this represents an interesting proof-of-concept work that may solve the perennial problem of insufficient labelling in medical image analysis.

D. Segmentation

CT and MRI mage segmentation research covers a variety of organs such as liver, prostate and knee cartilage, but a large amount of work has focused on brain segmentation, including tumor segmentation. The latter is especially important in surgical planning to determine the exact boundaries of the tumor in order to direct surgical resection. Sacrificing too much of eloquent brain areas during surgery would cause neurological deficits such as limb weakness, numbness and cognitive impairment. Traditionally, medical anatomical segmentation was done by hand, with a clinician drawing outlines slice by slice through an entire MRI or CT volume stack, therefore it is ideal to implement a solution that automates this laborious task. An excellent review of brain MRI segmentation was written by Akkus *et al.* [75], who reviewed various CNN architectures and metrics used in segmentation. Additionally, he also detailed the numerous competitions and their datasets, such as Brain Tumor Segmentation (BRATS), Mild traumatic brain injury outcome prediction (MTOP) and Ischemic Stroke Lesion Segmentation (ISLES).

Moeskops *et al.* [76] used 3 CNNs, each with a different 2-dimensional input patch size, running in parallel to classify and segment MRI brain images of 22 pre-term infants and 35 adults into different tissue classes such as white matter, grey matter and cerebrospinal fluid. The advantage of using 3 different input patch sizes is that each focuses on capturing different aspects of the image, with the smallest patch focused on local textures while the larger patch sizes assimilated spatial features. Overall, the algorithm achieved good accuracy, with Dice coefficients between 0.82 and 0.87. Most segmentation research has been on 2-dimensional image slices, but Milleterai *et al.* [20] applied 3-dimensional CNN to segment MRI prostate images from the PROMISE2012 challenge dataset. Their proposed V-net was inspired by Ronnerberger's U-Net architecture [19], and was trained on 50 MRI prostate scans and tested on 30 similar scans. V-net achieved a dice similarity coefficient score of 0.869, which was similar to that of the top placed teams in the challenge. Pereira *et al.* [77] applied deliberately small filters of 3 x 3 size, to allow the design of a deeper 11 convolution layer CNN, and to reduce overfitting. Their CNN was trained on 274 MRI brain tumor scans of gliomas, a type of brain tumor with significant malignant potential, obtaining first place in the BRATS 2013 and second place in the BRATS 2015 challenge. Havaei *et al.* [78] also looked at gliomas, and explored various 2-dimensional CNN architectures on the BRATS 2013 dataset. Their algorithm performed better than the BRATS 2013 winner, and took 3 minutes to run, compared to 100 minutes. Their InputCascadeCNN had a cascaded architecture, with the output of a first CNN being fed into a second CNN. Chen *et al.* [79] proposed using up-sampled filters, atrous spatial pyramid pooling, and fully connected Conditional Random Fields (CRFs). These aid in enlarging the field of each filter's view at multiple scales and improve localization accuracy. With this architecture which they called DeepLab, Chen *et al.* achieved state-of-the-art performance in the PASCAL VOC-2012 Image segmentation task, reaching 79.7% mean Intersection over Union (mIOU). There is some overlap with Moeskops' [76] use of input patches at different scales, and it would be interesting to see how this work in image segmentation can be advanced further. A more recent study by Casamitjana *et al* [80], compared various 3-dimensional CNN architectures. Training on the BRATS 2015 brain tumor dataset, they found that their 3D-CNN, modified from the DeepMedic CNN by Kamnitsas *et al.* [21], performed the best, and advocated using smaller receptive fields with multi-scale architecture. Brosch *et al.* [81] also exploited the analysis of multi-scale architecture, in segmenting MRI brain lesions of multiple sclerosis. They employed a novel approach but using both an encoder convolutional pathway consisting of pre-trained RBMs, and a deconvolutional pathway similar to a U-Net architecture.

E. Registration

Although the registration of medical images has many potential applications, which were reviewed by El-Zahraa *et*

al. [82], their actual clinical use is encountered in niche areas. Image registration is employed in neurosurgery or spinal surgery, to localize a tumor or spinal bony landmark, in order to facilitate surgical tumor removal or spinal screw implant placement. A reference image is aligned to a second image, called a sense image and various similarity measures and reference points are calculated to align the images, which can be 2 or 3-dimensional. The reference image may be a pre-operative MRI brain scan and the sense image may be an intraoperative MRI brain scan done after a first-pass resection, to determine if there is remnant tumor and if further resection is required. Using MRI brain scans from the OASIS dataset, Yang *et al.* [83] stacked convolution layers in an encoder-decoder fashion, to predict how an input pixel would morph into its final configuration. They invoked the use of a Large deformation diffeomorphic metric mapping (LDDMM) registration model and achieved dramatic improvements in computational time. Miao *et al.* [84] trained a 5 layer CNN on synthetic X-ray images in order to register 3-dimensional models of a knee implant, a hand implant, and a trans-esophageal probe onto 2-dimensional X-ray images, in order to estimate their pose. Their method obtained successful registrations 79-99% of the time, and took 0.1 seconds, a significant improvement over traditional intensity-based registration methods.

IV. CONCLUSION

A. Challenges

A recurring theme in machine learning is the limit imposed by the lack of labelled datasets, which hampers training and task performance. Conversely, it is acknowledged that more data improves performance, as Sun *et al.* [85] shows using an internal Google dataset of 300 million images. In general computer vision tasks, attempts have been made to circumvent limited data by using smaller filters on deeper layers [47], with novel CNN architecture combinations [86], or hyperparameter optimization [87].

In medical image analysis, the lack of data is two-fold and more acute: there is general lack of publicly available data, and high quality labelled data is even more scarce. Most of the datasets presented in this review involve fewer than 100 patients. Yet the situation may not be as dire as it seems, as despite the small training datasets, the papers in this review report relatively satisfactory performance in the various tasks. The question of how many images are necessary for training in medical image analysis was partially answered by Cho *et al.* [88]. He ascertained the accuracy of a CNN with GoogLeNet architecture in classifying individual axial CT images into one of 6 body regions: brain, neck, shoulder, chest, abdomen, pelvis. With 200 training images, accuracies of 88-98% were achieved on a test set of 6000 images. While categorization into various body regions is not a realistic medical image analysis task, his report does suggest that the problem may be surmountable. Being able to accomplish classification with a small dataset is possibly due to the general intrinsic image homogeneity across different patients, as opposed to the near-infinite variety of natural images, such as a dog in various breeds, colors and poses.

VAEs and GANS, being generative models, may sidestep the data paucity problem, by creating synthetic medical data. This was done by Guibas and Virdi, who used a 2 stage GAN to segment and then generate retinal fundus images successfully [89]. Their work was built on the research of Costa *et al.*, which first described using GANs to generate retinal fundus images [90]. Aside from synthetic data generation, GANs have been used in brain MRI segmentation as well by Moeskops *et al.* [91], Kamnitsas *et al.* [92] and Varghese *et al.* [93].

Data or class imbalance in the training set is also a significant issue in medical image analysis [94]. This refers to the number of images in the training data being skewed towards normal and non-pathological images. Rare diseases are an extreme example of this and can be missed without adequate training examples. This data imbalance effect can be ameliorated by using data augmentation to generate more training images of rare or abnormal data, though there is risk of overfitting. Aside from data-level strategies, algorithmic modification strategies and cost sensitive learning have also been studied [95], [96].

An important, non-technical challenge is the public reception towards their health results being studied by a non-human actor. This situation is not helped by the apocalyptic artificial intelligence scenarios painted by some. Machine learning algorithms have surpassed human performance in image recognition tasks, and it is likely that they will perform better than humans in medical image analysis as well. Indeed, some of the papers in this review report that dermatologists and radiologists have already been bested by machine learning. Yet the question regarding legal and moral culpability arises when a patient is misdiagnosed, or suffers morbidity as a result of AI or AI-assisted medical management. This is accentuated by our inability to fully explain how the black-box of machine algorithms work. However, it is likely that our relationship will continue evolve and recalibrate as AI-based technologies mature and inexorably permeate different facets of our lives.

B. Future Applications

The traditional applications for medical image analysis were discussed in Section 3. New areas of research include prognostication [97], content-based image retrieval [98], [99], image report or caption generation [100], [101], and manipulation of physical objects with LSTMs and reinforcement learning [102], [103] involving surgical robots [104], [105]. A few innovative applications that span across traditional medical image analysis categories are described below.

An interesting application was reported by Nie *et al.* [106], [107], in which GANs were used to generate CT brain images from MRI images. This is remarkable, as it means that patients can potentially avoid the ionizing radiation from a CT scanner altogether, lowering cost and improving patient safety. Nie also exploited the ability of GANs to generate improved, higher resolution images from native images [108] and reduced the blurriness in the CT images. A useful extension of resolution-improvement techniques would be applying them to generate MRI images of higher quality. High quality MRI images require high tesla (and correspondingly costlier) MRI scanners.

Algorithmically-generated high quality MRI images on a lower field-strength scanner would thus lower healthcare costs.

Chang [109] demonstrated a novel application in the nascent area of radio-genomics, which uses radiological images to predict the underlying molecular origin of a tissue. He first used an autoencoder to learn latent features from MRI images of glioblastoma multiforme (GBM), a malignant brain tumor, from The Cancer Genome Atlas Glioblastoma Multiforme (TCGA-GBM) data collection [110]. The learned features were then fed into a fully connected classifier layer to classify a MRI scan into one of 4 known molecular sub-types of GBM. Although still early, Chang's work could potentially diagnose a GBM sub-type and obviate the need for surgical biopsy and molecular assays. The generalizability of this technique to tumors elsewhere in the body is also promising.

Coudray *et al.* [111] accomplished an analogous task, but used histopathological images to classify lung cancer sub-types, and to predict common genetic mutations. Knowing the genetic mutations is helpful in prognosticating length of survival and guiding the choice of chemotherapy. Their method outperforms a human pathologist, and the prediction of genetic mutations had AUC scores of between 0.73 to 0.86.

Tsochatzidis *et al.* [112] described an original work combining content-based image retrieval (CBIR) and computer-aided diagnosis (CADx). In essence, their model segmented a lesion on a query image, and compared this to the segmented lesions in their database, consisting of 400 Regions of interest derived from the Digital Database for Screening Mammography (DDSM). The basis of comparison were the Euclidean distances between the representation vectors of the query lesion and database lesions. The model then outputs both reference images and a likelihood of a lesion being benign or malignant. They reported that their combined CBIR and CADx method resulted in state of the art prediction accuracy of 81%. These examples highlight how the field of machine learning in medical image analysis is changing rapidly, and that there may still be numerous applications which have not been conceived of yet.

Fig. 1. A collage of images depicting medical images, from left to right, top to bottom: an axial CT brain scan with a left-sided hemorrhagic stroke, an axial MRI brain scan with a left-sided brain tumor (a schwannoma), a normal chest X-ray, a normal axial CT lung scan, and a histology slide with high grade glioma (a brain tumor)

Fig. 2. In this example disease classification task, an input image of an abnormal axial slice of a T2-weighted MRI brain is run through a schematic depiction of a CNN. Feature extraction of the input image is performed via the Convolution, RELU and pooling layers, before classification by the fully connected layer

Fig. 3. Various neural network architectures. A. Recurrent neural network, B. Autoencoder, C. Restricted Boltzmann Machine, D. Deep Belief Network, E. Generative Adversarial

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2788044, IEEE Access

10

Network. x, y, h represent input, output and hidden layers respectively

TABLE I REVIEW ARTICLES AND BOOKS IN DEEP LEARNING FOR MEDICAL IMAGE ANALYSIS

TABLE II HIGHLY-CITED ARTICLES IN DEEP LEARNING FOR MEDICAL IMAGE ANALYSIS

REFERENCES

[1] C. J. Hsiao, E. Hing, and J. Ashman, "Trends in electronic health record system use among office-based physicians: United states, 2007-2012," *Natl Health Stat Report*, no. 75, pp. 1–18, 2014.

[2] R. Smith-Bindman, D. L. Miglioretti, E. Johnson, C. Lee, H. S. Feigelson, M. Flynn, R. T. Greenlee, R. L. Kruger, M. C. Hornbrook, D. Roblin, L. I. Solberg, N. Vanneman, S. Weinmann, and A. E. Williams, "Use of diagnostic imaging studies and associated radiation exposure for patients enrolled in large integrated health care systems, 1996-2010," *JAMA*, vol. 307, no. 22, pp. 2400–2409, 2012.

[3] E. Shortliffe, *Computer-based medical consultations: MYCIN*. Elsevier, 1976, vol. 2.

[4] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *arXiv preprint arXiv:1702.05747*, 2017.

[5] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity. 1943," *Bull. Math. Biol.*, vol. 52, no. 1-2, pp. 99–115, 1990.

[6] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain." *Psychological review*, vol. 65, no. 6, p. 386, 1958.

[7] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *The Journal of physiology*, vol. 160, no. 1, pp. 106–154, 1962.

[8] K. Fukushima and S. Miyake, "Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition," *Competition and cooperation in neural nets*, pp. 267–285, 1982.

[9] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural computation*, vol. 1, no. 4, pp. 541–551, 1989.

[10] D. E. Rumelhart, G. E. Hinton, R. J. Williams *et al.*, "Learning representations by back-propagating errors," *Cognitive modeling*, vol. 5, no. 3, p. 1, 1988.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[12] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annual Review of Biomedical Engineering*, no. 0, 2017.

[13] K. Suzuki, "Overview of deep learning in medical imaging," *Radiological Physics and Technology*, vol. 10, no. 3, pp. 257–273, 2017.

[14] 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 Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1153–1159, 2016.

[15] D. Adams, "Publish or perish 5," 2017, accessed:30/10/2017; https://harzing.com/resources/publish-or-perish/windows.

[16] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle *et al.*, "The cancer imaging archive (tcia): maintaining and operating a public information repository," *Journal of digital imaging*, vol. 26, no. 6, pp. 1045–1057, 2013.

[17] 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," *arXiv preprint arXiv:1705.02315*, 2017.

[18] G. Eli, L. Wenqi, S. Carole, F. Lucas, I. S. Dzhoshkun, W. Guotai, E.-R. Zach, G. Robert, D. Tom, H. Yipeng, W. Tom, N. Parashkev, M. Marc, C. B. Dean, O. Sébastien, J. C. M., and V. Tom, "Niftynet: a deep-learning platform for medical imaging," 2017, http://www.niftynet.io/.

[19] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 234–241.

[20] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *3D Vision (3DV), 2016 Fourth International Conference on*, 2016, pp. 565–571.

[21] K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, "Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation," *Medical image analysis*, vol. 36, pp. 61–78, 2017.

[22] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[23] R. Olga, D. Jia, S. Hao, K. Jonathan, S. Sanjeev, M. Sean, H. Zhiheng, K. Andrej, K. Aditya, B. Michael, C. B. Alexander, and F.-F. Li, "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015. [Online]. Available: 10.1007/s11263-015-0816-y

[24] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.

[25] H. Ravishankar, P. Sudhakar, R. Venkataramani, S. Thiruvenkadam, P. Annangi, N. Babu, and V. Vaidya, "Understanding the mechanisms of deep transfer learning for medical images," in *International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis*, 2016, pp. 188–196.

[26] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.

[27] I. Sutskever, J. Martens, and G. E. Hinton, "Generating text with recurrent neural networks," in *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, 2011, pp. 1017–1024.

[28] A. Karpathy and L. Fei-Fei, "Deep visual-semantic alignments for generating image descriptions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3128–3137.

[29] J. Chen, L. Yang, Y. Zhang, M. Alber, and D. Z. Chen, "Combining fully convolutional and recurrent neural networks for 3d biomedical image segmentation," in *Advances in Neural Information Processing Systems*, 2016, pp. 3036–3044.

[30] M. F. Stollenga, W. Byeon, M. Liwicki, and J. Schmidhuber, "Parallel multi-dimensional lstm, with application to fast biomedical volumetric image segmentation," in *Advances in Neural Information Processing Systems*, 2015, pp. 2998–3006.

[31] H.-C. Shin, K. Roberts, L. Lu, D. Demner-Fushman, J. Yao, and R. M. Summers, "Learning to read chest x-rays: recurrent neural cascade model for automated image annotation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2497–2506.

[32] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *Journal of Machine Learning Research*, vol. 11, no. Dec, pp. 3371–3408, 2010.

[33] C. Poultney, S. Chopra, Y. L. Cun *et al.*, "Efficient learning of sparse representations with an energy-based model," in *Advances in neural information processing systems*, 2007, pp. 1137–1144.

[34] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.

[35] M. Kallenberg, K. Petersen, M. Nielsen, A. Y. Ng, P. Diao, C. Igel, C. M. Vachon, K. Holland, R. R. Winkel, N. Karssemeijer *et al.*, "Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1322–1331, 2016.

[36] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[37] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, "A learning algorithm for boltzmann machines," *Cognitive science*, vol. 9, no. 1, pp. 147–169, 1985.

[38] P. Smolensky and P. Smolensky, "Information processing in dynamical systems: Foundations of harmony theory," ser. Parallel Distributed Processing. Explorations in the Microstructure of Cognition: Foundations, D. E. Rumelhart, Ed., vol. 1. Cambridge, MA: MIT Press, 1986, pp. 194–281.

[39] G. van Tulder and M. de Bruijne, "Combining generative and discriminative representation learning for lung ct analysis with convolutional restricted boltzmann machines," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1262–1272, 2016.

TABLE I
TABLE I REVIEW ARTICLES AND BOOKS IN DEEP LEARNING FOR MEDICAL IMAGE ANALYSIS

Type	Author	Title	Journal or Publisher	Year Published
Review	Greenspan H., et al. [14]	Guest Editorial Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique	IEEE Transactions on Medical Imaging	2016
Review	Litjens G.,et al. [4]	A survey on deep learning in medical image analysis	Medical Image Analysis	2017
Review	Shen D., et al. [12]	Deep Learning in Medical Image Analysis	Annual Review of Biomedical Engineering	2017
Review	Suzuki K. [13]	Overview of deep learning in medical imaging	Radiological Physics and Technology	2017
Book	Zhou K., et al. [113]	Medical image recognition, segmentation and parsing: machine learning and multiple object approaches	Academic Press	2015
Book	Carneiro G., et al. [114]	Deep Learning and Data Labeling for Medical Applications	Springer	2016
Book	Wu G., et al. [115]	Machine Learning and Medical Imaging	Academic Press	2016
Book	Cardoso MJ., et al. [116]	Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support	Springer	2017
Book	Hernández MV., et al. [117]	Medical Image Understanding and Analysis	Springer	2017
Book	Lu L., et al. [118]	Deep learning and convolutional neural networks for medical image computing: precision medicine, high performance and large-scale datasets	Springer	2017
Book	Zhou K., et al. [119]	Deep Learning for Medical Image Analysis	Academic Press	2017

TABLE II
HIGHLY-CITED ARTICLES IN DEEP LEARNING FOR MEDICAL IMAGE ANALYSIS

| S/n	Author	Year Published	Title	Citations
1	Shin et al. [24]	2016	Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning	225
2	Esteva et al. [69]	2017	Dermatologist-level classification of skin cancer with deep neural networks	214
3	Havaei et al. [78]	2017	Brain tumor segmentation with deep neural networks	198
4	Kamnitsas et al. [21]	2017	Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation	136
5	Zhang et al. [120]	2015	Deep convolutional neural networks for multi-modality isointense infant brain image segmentation	133
6	Iglesias et al. [121]	2015	Multi-atlas segmentation of biomedical images: a survey	131
7	Greenspan et al. [14]	2016	Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique	128
8	Tajbakhsh et al. [26]	2016	Convolutional neural networks for medical image analysis: Full training or fine tuning?	113
9	Pereira et al. [77]	2016	Brain tumor segmentation using convolutional neural networks in MRI images	109
10	Milletari et al. [20]	2016	V-net: Fully convolutional neural networks for volumetric medical image segmentation	89
11	Roth et al. [122]	2016	Improving computer-aided detection using convolutional neural networks and random view aggregation	89
12	Çiçek et al. [123]	2016	3D U-Net: learning dense volumetric segmentation from sparse annotation	82
13	Sirinukunwattana et al. [72]	2016	Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images	81
14	Anthimopoulos et al. [124]	2016	Lung pattern classification for interstitial lung diseases using a deep convolutional neural network	79
15	Setio et al. [125]	2016	Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks	79
16	Xu et al. [73]	2016	Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images	78
17	Roth et al. [126]	2015	Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation	77
18	Moeskops et al. [76]	2016	Automatic segmentation of MR brain images with a convolutional neural network	76
19	Suk et al. [97]	2015	Latent feature representation with stacked auto-encoder for AD/MCI diagnosis	74
20	Bar et al. [127]	2015	Deep learning with non-medical training used for chest pathology identification	73
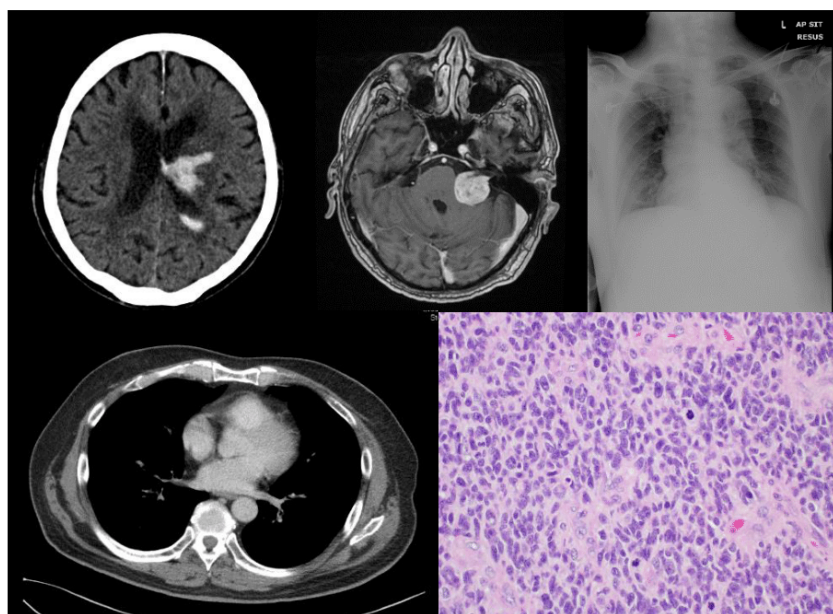
Fig. 1. A collage of images depicting medical images, from left to right, top to bottom: an axial CT brain scan with a left-sided hemorrhagic stroke, an axial MRI brain scan with a left-sided brain tumor, a normal chest X-ray, a normal axial CT lung scan, and a histology slide with high grade glioma (a brain tumor).
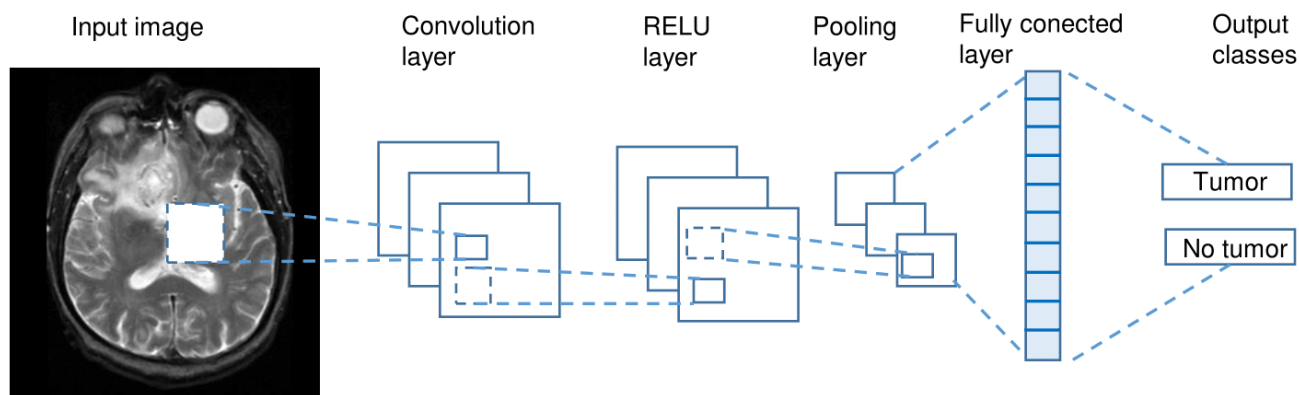

Fig. 2. In this example disease classification task, an input image of an abnormal axial slice of a T2-weighted MRI brain is run through a schematic depiction of a CNN. Feature extraction of the input image is performed via the Convolution, RELU and pooling layers, before classification by the fully connected layer

[40] M. A. Carreira-Perpinan and G. E. Hinton, "On contrastive divergence learning." in *Aistats*, vol. 10, 2005, pp. 33–40.

[41] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[42] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," in *Advances in neural information processing systems*, 2007, pp. 153–160.

[43] A. Khatami, A. Khosravi, T. Nguyen, C. P. Lim, and S. Nahavandi, "Medical image analysis using wavelet transform and deep belief networks," *Expert Systems with Applications*, 2017.

[44] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[45] S. B. Lo, S. A. Lou, J. S. Lin, M. T. Freedman, M. V. Chien, and S. K. Mun, "Artificial convolution neural network techniques and applications for lung nodule detection," *IEEE Trans Med Imaging*, vol. 14, no. 4, pp. 711–718, 1995.

[46] A. Rajkomar, S. Lingam, A. G. Taylor, M. Blum, and J. Mongan, "High-throughput classification of radiographs using deep convolutional neural networks," *Journal of digital imaging*, vol. 30, no. 1, pp. 95–101, 2017.

[47] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer*

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2788044, IEEE Access
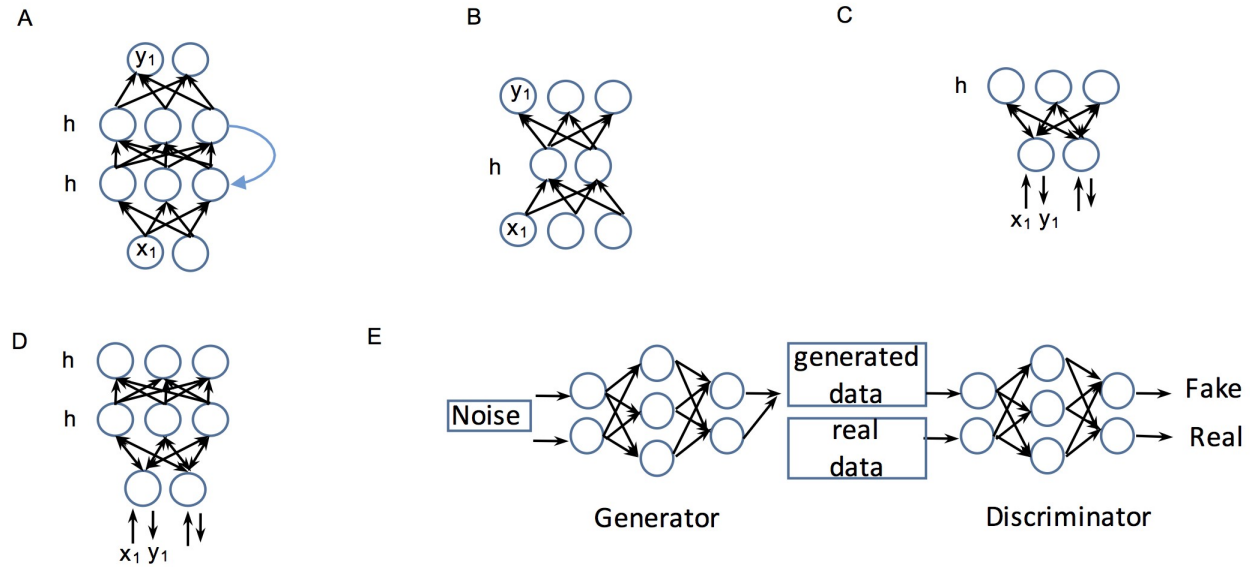
13

Fig. 3. Various neural network architectures. A. Recurrent neural network, B. Autoencoder, C. Restricted Boltzmann Machine, D. Deep Belief Network, E. Generative Adversarial Network. x, y, h represent input, output and hidden layers respectively

vision and pattern recognition, 2015, pp. 1–9.

[48] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya *et al.*, "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning," *arXiv preprint arXiv:1711.05225*, 2017.

[49] G. Huang, Z. Liu, K. Q. Weinberger, and L. van der Maaten, "Densely connected convolutional networks," *arXiv preprint arXiv:1608.06993*, 2016.

[50] W. Shen, M. Zhou, F. Yang, C. Yang, and J. Tian, "Multi-scale convolutional neural networks for lung nodule classification," *Inf Process Med Imaging*, vol. 24, pp. 588–599, 2015.

[51] R. Li, W. Zhang, H. I. Suk, L. Wang, J. Li, D. Shen, and S. Ji, "Deep learning based imaging data completion for improved brain disease diagnosis," *Med Image Comput Comput Assist Interv*, vol. 17, no. Pt 3, pp. 305–312, 2014.

[52] E. Hosseini-Asl, M. Ghazal, A. Mahmoud, A. Aslantas, A. M. Shalaby, M. F. Casanova, G. N. Barnes, G. Gimel'farb, R. Keynton, and A. El-Baz, "Alzheimer's disease diagnostics by a 3d deeply supervised adaptable convolutional network," *Front Biosci (Landmark Ed)*, vol. 23, pp. 584–596, 2018.

[53] S. Korolev, A. Safiullin, M. Belyaev, and Y. Dodonova, "Residual and plain convolutional neural networks for 3d brain mri classification," *arXiv preprint arXiv:1701.06643*, 2017.

[54] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[55] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[56] H. Pratt, F. Coenen, D. M. Broadbent, S. P. Harding, and Y. Zheng, "Convolutional neural networks for diabetic retinopathy," *Procedia Computer Science*, vol. 90, pp. 200–205, 2016.

[57] M. D. Abramoff, Y. Lou, A. Erginay, W. Clarida, R. Amelon, J. C. Folk, and M. Niemeijer, "Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learningdeep learning detection of diabetic retinopathy," *Investigative ophthalmology & visual science*, vol. 57, no. 13, pp. 5200–5206, 2016.

[58] S. M. Plis, D. R. Hjelm, R. Salakhutdinov, E. A. Allen, H. J. Bockholt, J. D. Long, H. J. Johnson, J. S. Paulsen, J. A. Turner, and V. D. Calhoun, "Deep learning for neuroimaging: a validation study," *Front Neurosci*, vol. 8, p. 229, 2014.

[59] H. I. Suk, C. Y. Wee, S. W. Lee, and D. Shen, "State-space model with deep learning for functional dynamics estimation in resting-state fmri," *Neuroimage*, vol. 129, pp. 292–307, 2016.

[60] M. D. Kumar, M. Babaie, S. Zhu, S. Kalra, and H. R. Tizhoosh, "A comparative study of cnn, bovw and lbp for classification of histopathological images," *arXiv preprint arXiv:1710.01249*, 2017.

[61] Z. Yan, Y. Zhan, Z. Peng, S. Liao, Y. Shinagawa, D. N. Metaxas, and X. S. Zhou, "Bodypart recognition using multi-stage deep learning," *Inf Process Med Imaging*, vol. 24, pp. 449–461, 2015.

[62] H. R. Roth, C. T. Lee, H.-C. Shin, A. Seff, L. Kim, J. Yao, L. Lu, and R. M. Summers, "Anatomy-specific classification of medical images using deep convolutional nets," in *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on*, 2015, pp. 101–104.

[63] H.-C. Shin, M. R. Orton, D. J. Collins, S. J. Doran, and M. O. Leach, "Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4d patient data," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 8, pp. 1930–1943, 2013.

[64] B. A. H. I. Kaggle, "Kaggle data science bowl 2017," 2017, https://www.kaggle.com/c/data-science-bowl-2017.

[65] L. Fangzhou, "Solution of the grt123 team," 2017, explanation by winner of 2017 Kaggle Data Science Bowl; https://github.com/lfz/DSB2017/blob/master/solution-grt123-team.pdf.

[66] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "Overfeat: Integrated recognition, localization and detection using convolutional networks," *arXiv preprint arXiv:1312.6229*, 2013.

[67] F. Ciompi, B. de Hoop, S. J. van Riel, K. Chung, E. T. Scholten, M. Oudkerk, P. A. de Jong, M. Prokop, and B. van Ginneken, "Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2d views and a convolutional neural network out-of-the-box," *Med Image Anal*, vol. 26, no. 1, pp. 195–202, 2015.

[68] F. Ciompi, C. Jacobs, E. T. Scholten, M. M. Wille, P. A. de Jong, M. Prokop, and B. van Ginneken, "Bag-of-frequencies: a descriptor of pulmonary nodules in computed tomography images," *IEEE Trans Med Imaging*, vol. 34, no. 4, pp. 962–973, 2015.

[69] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[70] D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks," in *International Conference on Medical Image Computing and Computer-assisted Intervention*, 2013, pp. 411–418.

[71] X. Yang, S. Y. Yeo, J. M. Hong, S. T. Wong, W. T. Tang, Z. Z. Wu, G. Lee, S. Chen, V. Ding, B. Pang *et al.*, "A deep learning approach for tumor tissue image classification," *Biomedical Engineering*, 2016.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2788044, IEEE Access

14

[72] K. Sirinukunwattana, S. E. A. Raza, Y.-W. Tsang, D. R. Snead, I. A. Cree, and N. M. Rajpoot, "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1196–1206, 2016.

[73] J. Xu, L. Xiang, Q. Liu, H. Gilmore, J. Wu, J. Tang, and A. Madabhushi, "Stacked sparse autoencoder (ssae) for nuclei detection on breast cancer histopathology images," *IEEE transactions on medical imaging*, vol. 35, no. 1, pp. 119–130, 2016.

[74] S. Albarqouni, C. Baur, F. Achilles, V. Belagiannis, S. Demirci, and N. Navab, "Aggnet: deep learning from crowds for mitosis detection in breast cancer histology images," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1313–1321, 2016.

[75] Z. Akkus, A. Galimzianova, A. Hoogi, D. L. Rubin, and B. J. Erickson, "Deep learning for brain mri segmentation: State of the art and future directions," *J Digit Imaging*, vol. 30, no. 4, pp. 449–459, 2017.

[76] P. Moeskops, M. A. Viergever, A. M. Mendrik, L. S. de Vries, M. J. Benders, and I. Isgum, "Automatic segmentation of mr brain images with a convolutional neural network," *IEEE Trans Med Imaging*, vol. 35, no. 5, pp. 1252–1261, 2016.

[77] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in mri images," *IEEE Trans Med Imaging*, vol. 35, no. 5, pp. 1240–1251, 2016.

[78] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P. M. Jodoin, and H. Larochelle, "Brain tumor segmentation with deep neural networks," *Med Image Anal*, vol. 35, pp. 18–31, 2017.

[79] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *arXiv preprint arXiv:1606.00915*, 2016.

[80] A. Casamitjana, S. Puch, A. Aduriz, E. Sayrol, and V. Vilaplana, "3d convolutional networks for brain tumor segmentation," *Proceedings of the MICCAI Challenge on Multimodal Brain Tumor Image Segmentation (BRATS)*, pp. 65–68, 2016.

[81] T. Brosch, L. Y. Tang, Y. Yoo, D. K. Li, A. Traboulsee, and R. Tam, "Deep 3d convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1229–1239, 2016.

[82] F. E.-Z. A. El-Gamal, M. Elmogy, and A. Atwan, "Current trends in medical image registration and fusion," *Egyptian Informatics Journal*, vol. 17, no. 1, pp. 99–124, 2016.

[83] X. Yang, R. Kwitt, M. Styner, and M. Niethammer, "Quicksilver: Fast predictive image registration - a deep learning approach," *Neuroimage*, vol. 158, pp. 378–396, 2017.

[84] S. Miao, Z. J. Wang, and R. Liao, "A cnn regression approach for real-time 2d/3d registration," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1352–1363, 2016.

[85] C. Sun, A. Shrivastava, S. Singh, and A. Gupta, "Revisiting unreasonable effectiveness of data in deep learning era," *arXiv preprint arXiv:1707.02968*, 2017.

[86] R. Socher, B. Huval, B. Bath, C. D. Manning, and A. Y. Ng, "Convolutional-recursive deep learning for 3d object classification," in *Advances in Neural Information Processing Systems*, 2012, pp. 656–664.

[87] J. Snoek, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," in *Advances in neural information processing systems*, 2012, pp. 2951–2959.

[88] J. Cho, K. Lee, E. Shin, G. Choy, and S. Do, "How much data is needed to train a medical image deep learning system to achieve necessary high accuracy?" *arXiv preprint arXiv:1511.06348*, 2015.

[89] J. T. Guibas, T. S. Virdi, and P. S. Li, "Synthetic medical images from dual generative adversarial networks," *arXiv preprint arXiv:1709.01872*, 2017.

[90] P. Costa, A. Galdran, M. I. Meyer, M. D. Abràmoff, M. Niemeijer, A. M. Mendonça, and A. Campilho, "Towards adversarial retinal image synthesis," *arXiv preprint arXiv:1701.08974*, 2017.

[91] P. Moeskops, M. Veta, M. W. Lafarge, K. A. Eppenhof, and J. P. Pluim, "Adversarial training and dilated convolutions for brain mri segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, 2017, pp. 56–64.

[92] K. Kamnitsas, C. Baumgartner, C. Ledig, V. Newcombe, J. Simpson, A. Kane, D. Menon, A. Nori, A. Criminisi, D. Rueckert *et al.*, "Unsupervised domain adaptation in brain lesion segmentation with adversarial networks," in *International Conference on Information Processing in Medical Imaging*, 2017, pp. 597–609.

[93] V. Alex, M. S. KP, S. S. Chennamsetty, and G. Krishnamurthi, "Generative adversarial networks for brain lesion detection," in *SPIE Medical Imaging*, 2017, pp. 101 330G–101 330G.

[94] M. A. Mazurowski, P. A. Habas, J. M. Zurada, J. Y. Lo, J. A. Baker, and G. D. Tourassi, "Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance," *Neural networks*, vol. 21, no. 2, pp. 427–436, 2008.

[95] V. López, A. Fernández, S. García, V. Palade, and F. Herrera, "An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics," *Information Sciences*, vol. 250, pp. 113–141, 2013.

[96] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, "A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 4, pp. 463–484, 2012.

[97] H.-I. Suk, S.-W. Lee, D. Shen, D. N. I. Alzheimer's *et al.*, "Latent feature representation with stacked auto-encoder for ad/mci diagnosis," *Brain Structure and Function*, vol. 220, no. 2, pp. 841–859, 2015.

[98] X. Liu, H. R. Tizhoosh, and J. Kofman, "Generating binary tags for fast medical image retrieval based on convolutional nets and radon transform," in *Neural Networks (IJCNN), 2016 International Joint Conference on*, 2016, pp. 2872–2878.

[99] Y. Anavi, I. Kogan, E. Gelbart, O. Geva, and H. Greenspan, "Visualizing and enhancing a deep learning framework using patients age and gender for chest x-ray image retrieval," in *Medical Imaging 2016: Computer-Aided Diagnosis*, vol. 9785, 2016, p. 978510.

[100] X. Wang, L. Lu, H. chang Shin, L. Kim, I. Nogues, J. Yao, and R. Summers, "Unsupervised category discovery via looped deep pseudo-task optimization using a large scale radiology image database," *arXiv preprint arXiv:1603.07965*, 2016.

[101] H.-C. Shin, L. Lu, L. Kim, A. Seff, J. Yao, and R. M. Summers, "Interleaved text/image deep mining on a very large-scale radiology database," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1090–1099.

[102] Y. Duan, M. Andrychowicz, B. Stadie, J. Ho, J. Schneider, I. Sutskever, P. Abbeel, and W. Zaremba, "One-shot imitation learning," *arXiv preprint arXiv:1703.07326*, 2017.

[103] S. Levine, C. Finn, T. Darrell, and P. Abbeel, "End-to-end training of deep visuomotor policies," *Journal of Machine Learning Research*, vol. 17, no. 39, pp. 1–40, 2016.

[104] B. Thananjeyan, A. Garg, S. Krishnan, C. Chen, L. Miller, and K. Goldberg, "Multilateral surgical pattern cutting in 2d orthotropic gauze with deep reinforcement learning policies for tensioning," in *Robotics and Automation (ICRA), 2017 IEEE International Conference on*, 2017, pp. 2371–2378.

[105] D. Seita, S. Krishnan, R. Fox, S. McKinley, J. Canny, and K. Goldberg, "Fast and reliable autonomous surgical debridement with cable-driven robots using a two-phase calibration procedure," *arXiv preprint arXiv:1709.06668*, 2017.

[106] D. Nie, R. Trullo, J. Lian, C. Petitjean, S. Ruan, Q. Wang, and D. Shen, "Medical image synthesis with context-aware generative adversarial networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2017, pp. 417–425.

[107] D. Nie, X. Cao, Y. Gao, L. Wang, and D. Shen, "Estimating ct image from mri data using 3d fully convolutional networks," in *International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis*, 2016, pp. 170–178.

[108] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," *arXiv preprint arXiv:1609.04802*, 2016.

[109] P. Chang, "Deep learning for predicting glioblastoma subtypes from mri," 2016, http://cai2r.net/i2i.

[110] "Radiology data from the cancer genome atlas glioblastoma multiforme [tcga-gbm] collection. the cancer imaging archive." 2016, https://wiki.cancerimagingarchive.net/display/Public/TCGA-GBM.

[111] N. Coudray, A. L. Moreira, T. Sakellaropoulos, D. Fenyo, N. Razavian, and A. Tsirigos, "Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning," *bioRxiv*, p. 197574, 2017.

[112] L. Tsochatzidis, K. Zagoris, N. Arikidis, A. Karahaliou, L. Costaridou, and I. Pratikakis, "Computer-aided diagnosis of mammographic masses based on a supervised content-based image retrieval approach," *Pattern Recognition*, 2017.

[113] S. K. Zhou, *Medical image recognition, segmentation and parsing: machine learning and multiple object approaches*. Academic Press, 2015.

[114] G. Carneiro, D. Mateus, P. Lo ́ ʻ, A. Bradley, J. M. R. Tavares, V. Belagiannis, J. P. Papa, J. C. Nascimento, M. Loog, Z. Lu *et al.*, *Deep Learning and Data Labeling for Medical Applications: First International Workshop, LABELS 2016, and Second International Workshop, DLMIA 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 21, 2016, Proceedings*. Springer, 2016, vol. 10008.

[115] G. Wu, D. Shen, and M. Sabuncu, *Machine Learning and Medical Imaging*. Academic Press, 2016.

[116] M. J. Cardoso, T. Arbel, G. Carneiro, T. Syeda-Mahmood, J. M. R. Tavares, M. Moradi, A. Bradley, H. Greenspan, J. P. Papa, A. Madabhushi *et al.*, *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, Proceedings*. Springer, 2017, vol. 10553.

[117] M. V. Hernández and V. González-Castro, *Medical Image Understanding and Analysis: 21st Annual Conference, MIUA 2017, Edinburgh, UK, July 11-13, 2017, Proceedings*. Springer, 2017, vol. 723.

[118] L. Lu, Y. Zheng, G. Carneiro, and L. Yang, *Deep learning and convolutional neural networks for medical image computing: precision medicine, high performance and large-scale datasets*. Springer, 2017.

[119] S. K. Zhou, H. Greenspan, and D. Shen, *Deep Learning for Medical Image Analysis*. Academic Press, 2017.

[120] W. Zhang, R. Li, H. Deng, L. Wang, W. Lin, S. Ji, and D. Shen, "Deep convolutional neural networks for multi-modality isointense infant brain image segmentation," *NeuroImage*, vol. 108, pp. 214–224, 2015.

[121] J. E. Iglesias and M. R. Sabuncu, "Multi-atlas segmentation of biomedical images: a survey," *Medical image analysis*, vol. 24, no. 1, pp. 205–219, 2015.

[122] H. R. Roth, L. Lu, J. Liu, J. Yao, A. Seff, K. Cherry, L. Kim, and R. M. Summers, "Improving computer-aided detection using convolutional neural networks and random view aggregation," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1170–1181, 2016.

[123] O. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: learning dense volumetric segmentation from sparse annotation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2016, pp. 424–432.

[124] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou, "Lung pattern classification for interstitial lung diseases using a deep convolutional neural network," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1207–1216, 2016.

[125] A. A. A. Setio, F. Ciompi, G. Litjens, P. Gerke, C. Jacobs, S. J. van Riel, M. M. W. Wille, M. Naqibullah, C. I. Sánchez, and B. van Ginneken, "Pulmonary nodule detection in ct images: false positive reduction using multi-view convolutional networks," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1160–1169, 2016.

[126] H. R. Roth, L. Lu, A. Farag, H.-C. Shin, J. Liu, E. B. Turkbey, and R. M. Summers, "Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 556–564.

[127] Y. Bar, I. Diamant, L. Wolf, and H. Greenspan, "Deep learning with non-medical training used for chest pathology identification," in *Proc. SPIE*, vol. 9414, 2015, p. 94140V.