Author Manuscript

1

6   Deep learning in medical imaging and radiation therapy

7   Berkman Sahiner, Aria Pezeshk, Lubomir M. Hadjiiski, Xiaosong Wang, Karen Drukker, Kenny

8   H. Cha, Ronald M. Summers, Maryellen L. Giger

9

10   Berkman Sahiner: DIDSR/OSEL/CDRH, U.S. Food and Drug Administration, Silver Spring, MD
11   20993 USA

12   Aria Pezeshk: DIDSR/OSEL/CDRH, U.S. Food and Drug Administration, Silver
13   Spring, MD 20993 USA

14   Lubomir M. Hadjiiski: Department of Radiology, University of Michigan, Ann Arbor, MI
15   48109

16   Xiaosong Wang: Imaging Biomarkers and Computer-aided Diagnosis Lab, Radiology and
17   Imaging Sciences, NIH Clinical Center, Bethesda, MD 20892-1182

18   Karen Drukker: Department of Radiology, University of Chicago, Chicago, IL 60637

19   Kenny H. Cha: DIDSR/OSEL/CDRH, U.S. Food and Drug Administration, Silver
20   Spring, MD 20993 USA

21   Ronald M. Summers: Imaging Biomarkers and Computer-aided Diagnosis Lab, Radiology and
22   Imaging Sciences, NIH Clinical Center, Bethesda, MD 20892-1182

23   Maryellen L. Giger: Department of Radiology, University of Chicago, Chicago, IL 60637

24   **Corresponding Author**: Berkman Sahiner, berkman.sahiner @fda.hhs.gov,

25   DIDSR/OSEL/CDRH, U.S. Food and Drug Administration, Silver Spring, MD 20993 USA

26   # Deep learning in medical imaging and radiation therapy

27   Berkman Sahiner[1], Aria Pezeshk[1], Lubomir M. Hadjiiski[2], Xiaosong Wang[3], Karen Drukker[4],

28   Kenny H. Cha[1], Ronald M. Summers[3], Maryellen L. Giger[4]

29   **Abstract**

30   The goals of this review paper on deep learning (DL) in medical imaging and radiation therapy

31   are to: 1) summarize what has been achieved to date; 2) identify common and unique challenges,

32   and strategies that researchers have taken to address these challenges; and 3) identify some of the

33   promising avenues for the future both in terms of applications as well as technical innovations.

34   We introduce the general principles of DL and convolutional neural networks, survey five major

35   areas of application of DL in medical imaging and radiation therapy, identify common themes,

36   discuss methods for data set expansion, and conclude by summarizing lessons learned, remaining

37   challenges, and future directions.

38   ## 1. INTRODUCTION

39   In the last few years, artificial intelligence (AI) has been rapidly expanding and permeating both

40   industry and academia. Many applications such as object classification, natural language

---

[1] DIDSR/OSEL/CDRH, U.S. Food and Drug Administration, Silver Spring, MD 20993 USA
[2] Department of Radiology, University of Michigan, Ann Arbor, MI 48109
[3] Imaging Biomarkers and Computer-aided Diagnosis Lab, Radiology and Imaging Sciences, NIH Clinical Center, Bethesda, MD 20892-1182
[4] Department of Radiology, University of Chicago, Chicago, IL 60637

41 processing and speech recognition, which until recently seemed to be many years away from

42 being able to achieve human levels of performance, have suddenly become viable.[1-3] Every

43 week, there is a news story about an AI system that has surpassed humans at various tasks

44 ranging from playing board games[4] to flying autonomous drones.[5] One report shows that

45 revenues from AI will increase by around 55% annually in the 2016-2020 time period from

46 roughly $8 billion to $47 billion.[6] Together with breakthroughs in other areas such as

47 biotechnology and nanotechnology, the advances in AI are leading to what the World Economic

48 Forum refers to as the fourth industrial revolution.[7] The disruptive changes associated with AI

49 and automation are already being seriously discussed among economists and other experts as

50 both having the potential to positively improve our everyday lives, e.g., by reducing healthcare

51 costs, as well as to negatively affect society, e.g., by causing large scale unemployment and

52 rising income inequality[8, 9] (according to one estimate, half of all working activities can be

53 automated by existing technologies[10]). The advances in AI discussed above have been almost

54 entirely based on the groundbreaking performance of systems that are based on deep learning

55 (DL). We now use DL-based systems on a daily basis when we use search engines to find images

56 on the web or talk to digital assistants on smart phones and home entertainment systems. Given

57 its widespread success in various computer vision applications (among other areas), DL is now

58 poised to dominate medical image analysis and has already transformed the field in terms of

59 performance levels that have been achieved across various tasks as well as its application areas.

60   **1.A. Deep learning, history, and techniques**

61 DL is a subfield of machine learning, which in turn is a field within AI. In general, DL consists

62 of massive multi-layer networks of artificial neurons that can automatically discover useful

63   features, i.e. representations of input data (in our case images) needed for tasks such as detection

64   and classification, given large amounts of unlabeled or labeled data.[11, 12]

65   Traditional applications of machine learning using techniques such as support vector machines

66   (SVM) or random forests (RF) took as input hand-crafted features, which are often developed

67   with a reliance on domain expertise, for each separate application such as object classification or

68   speech recognition. In imaging, hand-crafted features are extracted from the image input data

69   and reduce the dimensionality by summarizing the input into what is deemed to be the most

70   relevant information that helps with distinguishing one class of input data from another. On the

71   other hand, using the image pixels as the input, the image data can be flattened into a high-

72   dimensional vector; for example, in mammographic mass classification, a 500x500 pixel region

73   of interest will result in a vector with 250,000 elements. Given all the possible variations of a

74   mass's appearance due to differences in breast type, dose, type and size of a mass, etc., finding

75   the hyperplane that separates the high dimensional vectors of malignant and benign masses

76   would require a very large number of examples if the original pixel values are used. However,

77   each image can be summarized into a vector consisting of a few dozen or a few hundred

78   elements (as opposed to over a million elements in the original format) by extracting specialized

79   features that for instance describe the shape of the mass. This lower dimensional representation

80   is more easily separable using fewer examples if the features are relevant. A key problem with

81   this general approach is that useful features are difficult to design, often taking the collective

82   efforts of many researchers over years or even decades to optimize. The other issue is that the

83   features are domain or problem specific. One would not generally expect that features developed

84   for image recognition should be relevant for speech recognition, but even within image

85   recognition different types of problems such as lesion classification and texture identification

86 require separate sets of features. The impact of these limitations has been well demonstrated in

87 experiments that show performance of top machine learning algorithms to be very similar when

88 they are used to perform the same task using the same set of input features.[13] In other words,

89 traditional machine learning algorithms were heavily dependent on having access to good feature

90 representations, otherwise it was very difficult to improve the state-of-the-art results on a given

91 data set.

92 The key difference between DL and traditional machine learning techniques is that the former

93 can automatically learn useful representations of the data, thereby eliminating the need for hand-

94 crafted features. What is more interesting is that the representations learned from one data set can

95 be useful even when they are applied to a different set of data. This property, referred to as

96 transfer learning[14, 15], is not unique to DL but the large training data requirements of DL make it

97 particularly useful in cases where relevant data for a particular task is scarce. For instance, in

98 medical imaging, a DL system can be trained on a large number of natural images or those in a

99 different modality to learn proper feature representations that allow it to "see". The pre-trained

100 system can subsequently use these representations to produce an encoding of a medical image

101 that is used for classification.[16-18] Systems using transfer learning often outperform the state-of-

102 the-art methods based on traditional hand-crafted features that were developed over many years

103 with a great deal of expertise.

104 The success of DL compared to traditional machine learning methods is primarily based on two

105 inter-related factors: depth and compositionality.[11, 12, 19] A function is said to have a compact

106 expression if it has few computational elements used to represent it ("few" here is a relative term

107 that depends on the complexity of the function). An architecture with sufficient depth can

108 produce a compact representation, whereas an insufficiently deep one may require an

109   exponentially larger architecture (in terms of the number of computational elements that need to

110   be learned) to represent the same function. A compact representation requires fewer training

111   examples to tune the parameters and produces better generalization to unseen examples. This is

112   critically important in complex tasks such as computer vision where each object class can exhibit

113   many variations in appearance which would potentially require several examples per type of

114   variation in the training set if a compact representation is not used. The second advantage of

115   deep architectures has to do with how successive layers of the network can utilize the

116   representations from previous layers to compose more complex representations that better

117   capture critical characteristics of the input data and suppress the irrelevant variations (for

118   instance, simple translations of an object in the image should result in the same classification). In

119   image recognition, deep networks have been shown to capture simple information such as

120   presence or absence of edges at different locations and orientations in the first layer. Successive

121   layers of the network assemble the edges into compound edges and corners of shapes, and then

122   into more and more complex shapes that resemble object parts. Hierarchical representation

123   learning is very useful in complicated tasks such as computer vision where adjacent pixels and

124   object parts are correlated with each other and their relative locations provide clues about each

125   class of object, or speech recognition and natural language processing where the sequence of

126   words follow contextual and grammatical rules that can be learned from the data. This

127   distributed hierarchical representation has similarities with the function of the visual and auditory

128   cortexes in the human brain where basic features are integrated into more complex

129   representations that are used for perception.[20, 21]

130   As discussed earlier, DL is not a completely new concept, but rather mostly an extension of

131   previously existing forms of artificial neural networks (ANNs) to larger number of hidden layers

132 and nodes in each layer. In the late 1990s until early 2000s, ANNs started to lose popularity in

133 favor of SVMs and decision-tree-based methods such as random forests and gradient boosting

134 trees that seemed to be more consistently outperforming other learning methods.[22] The reason for

135 this was that ANNs were found to be both slow and difficult to train aside from shallow

136 networks with one to two hidden layers, as well as prone to getting stuck in local minima.

137 However, starting around 2006 a combination of several factors led to faster and more reliable

138 training of deep networks. One of the first influential papers was a method for efficient

139 unsupervised (i.e. using unlabeled data, as opposed to supervised training that uses data labeled

140 based on the ground truth) layer by layer training of deep restricted Boltzmann machines.[23] As

141 larger data sets became more commonplace, and with availability of commercial gaming

142 graphical processing units (GPUs) it became possible to explore training of larger deeper

143 architectures faster. At the same time, several innovations and best practices in network

144 architecture and training led to faster training of deep networks with excellent generalization

145 performance using stochastic gradient descent. Some examples include improved methods for

146 network initialization and weight updates,[24] new neuron activation functions,[25] randomly cutting

147 connections or zeroing of weights during training,[26, 27] and data augmentation strategies that

148 render the network invariant to simple transformations of the input data. Attention to these

149 improvements was still mostly concentrated within the machine learning community and not

150 being seriously considered in other fields such as computer vision. This changed in 2012 in the

151 ImageNet[28] competition in which more than a million training images with 1000 different object

152 classes were made available to the challenge participants. A DL architecture that has since been

153 dubbed AlexNet outperformed the state-of-the-art results from the computer vision community

154   by a large margin and convinced the general community that traditional methods were on their

155   way out.[29]

156   The most successful and popular DL architecture in imaging is the convolutional neural network

157   (CNN).[30] Nearby pixels in an image are correlated with one another both in areas that exhibit

158   local smoothness and areas consisting of structures (e.g. edges of objects or textured regions).

159   These correlations typically manifest themselves in different parts of the same image.

160   Accordingly, instead of having a fully connected network where every pixel is processed by a

161   different weight, every location can be processed using the same set of weights to extract various

162   repeating patterns across the entire image. These sets of trainable weights, referred to as kernels

163   or filters, are applied to the image using a dot product or convolution and then processed by a

164   non-linearity (e.g. a sigmoid or tanh function). Each of these convolution layers can consist of

165   many such filters resulting in the extraction of multiple sets of patterns at each layer. A pooling

166   layer (e.g. max-pooling where the output is the maximum value within a window) often follows

167   each convolution layer to both reduce the dimensionality as well as impose translation invariance

168   so that the network becomes immune to small shifts in location of patterns in the input image.

169   These convolution and pooling layers can be stacked to form a multi-layer network often ending

170   in one or more fully connected layers as shown in Fig **Error! Reference source not found.**,

171   followed by a softmax layer. The same concepts can be applied in 1D and 3D to accommodate

172   time-series and volumetric data, respectively. Compared to a fully connected network, CNNs

173   contain far fewer trainable parameters and therefore require less training time and fewer training

174   examples. Moreover, since their architecture is specifically designed to take advantage of

175   presence of local structures in images they are a natural choice for imaging applications and a

176   regular winner of various imaging challenges.

177 Another very interesting type of network is the recurrent neural network (RNN) which is ideal

178 for analyzing sequential data (e.g. text or speech) due to having an internal memory state that can

179 store information about previous data points. A variant of RNNs, referred to as long short term

180 memory (LSTM),[31] has improved memory retention compared to a regular RNN and has

181 demonstrated great success across a range of tasks from image captioning[32, 33] to speech

182 recognition[1, 34] and machine translation.[35]

183

184 Generative adversarial networks (GANs) and its different variants (e.g. WGAN[36], CycleGAN[37],

185 etc.) are another promising class of DL architectures that consist of two networks: a generator

186 and a discriminator.[38] The generator network produces new data instances that try to mimic the

187 data used in training, while the discriminator network tries to determine the probability of

188 whether the generated candidates belong to the training samples or not. The two networks are

189 trained jointly with backpropagation, with the generative network becoming better at generating

190 more realistic samples and the discriminator becoming better at detecting artificially generated

191 samples. GANs have recently demonstrated great potential in medical imaging applications such

192 as image reconstruction for compressed sensing in magnetic resonance imaging (MRI).[39]

193   **1.B. Deep learning in medical imaging**

194 In medical imaging, machine learning algorithms have been used for decades, starting with

195 algorithms to analyze or help interpret radiographic images in the mid-1960's.[40-42] Computer-

196 aided detection/diagnosis (CAD) algorithms started to make advances in the mid 1980's, first

197 with algorithms dedicated to cancer detection and diagnosis on chest radiographs and

198 mammograms,[43, 44] and then widening in scope to other modalities such as computed

199   tomography (CT) and ultrasound.[45, 46] CAD algorithms in the early days predominantly used a

200   data-driven approach as most DL algorithms do today. However, unlike most DL algorithms,

201   most of these early CAD methods heavily depended on feature engineering. A typical workflow

202   for developing an algorithm for a new task consisted of understanding what types of imaging and

203   clinical evidence clinicians use for the interpretation task, translating that knowledge into

204   computer code to automatically extract relevant features, and then using machine learning

205   algorithms to combine the features into a computer score. There were, however, some notable

206   exceptions. Inspired by the neocognitron architecture,[47] a number of researchers investigated the

207   use of CNNs[48-51] or shift-invariant ANNs[52, 53] in the early and mid-1990's, and massively-trained

208   artificial neural networks (MTANNs)[54, 55] in the 2000's for detection and characterization tasks

209   in medical imaging. These methods all shared common properties with current deep CNNs

210   (DCNNs): Data propagated through the networks via convolutions, the networks learned filter

211   kernels, and the methods did not require feature engineering, i.e., the inputs into the networks

212   were image pixel values. However, severely restricted by computational requirements of the

213   time, most of these networks were not deep, i.e., they mostly consisted of only one or two hidden

214   layers. In addition, they were trained using much smaller data sets compared to a number of

215   high-profile DCNNs that were trained using millions of natural images. Concepts such as

216   transfer learning,[14] residual learning,[56] and fully convolutional networks with skip connections[57]

217   were generally not well-developed. Thus, these earlier CNNs in medical imaging, as competitive

218   as they were compared to other methods, did not result in a massive transformation in machine

219   learning for medical imaging.

220   With the advent of DL, applications of machine learning in medical imaging have dramatically

221   increased, paralleling other scientific domains such as natural image and speech processing.

222 Investigations accelerated not only in traditional machine learning topics such as segmentation,

223 lesion detection and classification,[58] but also in other areas such as image reconstruction and

224 artifact reduction that were previously not considered as data driven topics of investigation. Fig.

225 2 [22] shows the number of peer-reviewed publications in the last six years in the areas of focus

226 for this paper, DL for radiological images, and shows a very strong trend: For example, in the

227 first three months of 2018, more papers were published on this topic than the whole year of

228 2016.

229

230 Using DL involves making a very large number of design decisions such as number of layers,

231 number of nodes in each layer (or number and size of kernels in the case of CNNs), type of

232 activation function, type and level of regularization, type of network initialization, whether to

233 include pooling layers and if so what type of pooling, type of loss function, and so on. One way

234 to avoid using trial and error for devising the best architecture is to follow the same exact

235 architectures that have shown to be successful in natural image analysis such as AlexNet,[29]

236 VGGNet,[59] ResNet,[56] DenseNet,[60] Xception,[61] or Inception V3.[62] These networks can be trained

237 from scratch for the new task.[63-67] Alternatively, they can be pre-trained on natural images that

238 are more plentiful compared to medical images so that the weights in the feature extraction

239 layers are properly set during training (see Sec 3.B for more details). The weights only in the last

240 fully-connected layer or last few layers (including some of the convolutional layers) can then be

241 retrained using medical images to learn the class associations for the desired task.

242   **1.C. Existing platforms and resources**

243   A large number of training examples are required to estimate the large number of parameters of a

244   DL system. One needs to perform backpropagation throughout many iterations using stochastic

245   gradient descent over mini-batches consisting of a small subset of samples at any given time to

246   train hundreds of thousands to hundreds of millions or even billions of parameters. A single or

247   multi-core central processing unit (CPU) or a cluster of CPU nodes in a high-performance

248   computing (HPC) environment could be used for training, but the former approach would take an

249   extremely long amount of time while the latter requires access to costly infrastructure.

250   Fortunately, in the last ten years gaming GPUs have become cheaper, increasingly powerful, and

251   easier to program. This has resulted in simultaneously far cheaper hardware requirements for

252   running DL (compared to HPC solutions) and training times that are several orders of magnitude

253   shorter compared to a solution run on a CPU.[27, 68] The most common setup for training DL

254   models is therefore to train networks on a desktop workstation containing one or more powerful

255   gaming GPUs that can be easily configured for a reasonable price. There are also several cloud-

256   based solutions including Amazon Web Services (AWS)[69] and Nvidia GPU cloud[70] that allow

257   users to train or deploy their models remotely. Recently, Google has developed Application-

258   Specific Integrated Circuit (ASIC) for neural networks to run its wide variety of applications that

259   utilize DL. These accelerators, referred to as Tensor Processing Units (TPUs), are several times

260   faster than CPU or GPU solutions and have recently been made available to general users via

261   Google Cloud.[71]

262   In line with the rapid improvements in performance of GPUs, several open-source DL libraries

263   have been developed and made public that free the user from directly programming GPUs. These

264   frameworks allow the users to focus on how to set up a particular network and explore different

265   training strategies. The most popular DL libraries are TensorFlow,[72] Caffe,[73] Torch,[74] and

266   Theano.[75] They all have Application Programming Interfaces (APIs) in different programming

267   languages, with the most popular language being Python.

268   **1.D. Organization of the paper**

269   Throughout the paper, we strived to refer to published journal articles as much as we could.

270   However, DL is a very fast-changing field, and reports of many excellent and new studies either

271   appear as a conference proceeding paper only, or as a pre-print in online resources such as arxiv.

272   We did not refrain from citing articles from these resources whenever necessary. In sections

273   other than Section 2, to better summarize the state-of-the-art, we have included publications from

274   many different medical imaging, and natural imaging. However, to keep the length of the paper

275   reasonable, in Section 2, we focused only on applications in radiological imaging and radiation

276   therapy, although there are other areas in medical imaging that have seen influx of DL

277   applications, such as digital pathology and optical imaging. This paper is organized as follows:

278   In Section 2, we summarize applications of DL to radiological imaging and radiation therapy. In

279   Section 3, we describe some of the common themes among DL applications, which include

280   training and testing with small data set sizes, pretraining and fine tuning, combining DL with

281   radiomics applications, and different types of training, such as supervised, unsupervised and

282   weakly supervised. Since data set size is a major bottleneck for DL applications in medical

283   imaging, we have devoted Section 4 to special methods for data set expansion. In Section 5, we

284   summarize some of the perceived challenges, lessons learned, and possible trends for the future

285   of DL in medical imaging and radiation therapy.

2. APPLICATION AREAS IN RADIOLOGICAL IMAGING AND RADIATION

THERAPY

2.A. Image Segmentation

DL has been used to segment many different organs in different imaging modalities, including single-view radiographic images, CT, MR, and ultrasound images.

Image segmentation in medical imaging based on DL generally uses two different input methods: 1) patches of an input image, and 2) the entire image. Both methods generate an output map that provides the likelihood that a given region is part of the object being segmented. While patch-based segmentation methods were initially used, most recent studies use the entire input image to give contextual information and reduce redundant calculations. Multiple works subsequently refine these likelihood maps using classic segmentation methods, such as level sets,[76-79], graph cuts,[80] and model-based methods,[81, 82] to achieve a more accurate segmentation than using the likelihood maps alone. Popular deep-learning frameworks used for segmentation tasks include Caffe, Matlab™ and cuda-convnet.

2.A.1.   **Organ and substructure segmentation**

Segmentation of organs and their substructures may be used to calculate clinical parameters such as volume, as well as to define the search region for computer-aided detection tasks to improve their performance. Patch-based segmentation methods, with refinements using traditional segmentation methods, have been shown to perform well for different segmentation tasks.[76, 83] Table **Error! Reference source not found.** briefly summarizes published performance of DL methods in organ and substructure segmentation tasks using either Dice coefficient or Jaccard index, if given, as the performance metric.

308 A popular network architecture for segmentation is the U-net.[84] It was originally developed for

309 segmentation of neuronal structures in electron microscope stacks. U-nets consist of several

310 convolution layers, followed by deconvolution layers, with connections between the opposing

311 convolution and deconvolution layers (skip connections), which allow for the network to analyze

312 the entire image during training, and allow for obtaining segmentation likelihood maps directly,

313 unlike the patch-based methods. Derivatives of U-net have been used for multiple tasks,

314 including segmenting breast and fibroglandular tissue[85] and craniomaxillofacial bony

315 structures.[86]

316 Another DL structure that is being used for segmentation of organs is holistically nested

317 networks (HNN). HNN uses side-outputs of the convolutional layers, which are multi-scale and

318 multi-level, and produce a corresponding edge map at different scale levels. A weighted average

319 of the side-outputs is used to generate the final output, and the weights for the average are

320 learned during the training of the network. HNN has been successfully implemented in

321 segmentation of the prostate[87] and brain tumors.[88]

322   2.A.2. **Lesion segmentation**

323 Lesion segmentation is a similar task to organ segmentation; however, lesion segmentation is

324 generally more difficult than organ segmentation, as the object being segmented can have

325 varying shapes and sizes. Multiple papers covering many different lesion types have been

326 published for DL lesion segmentation (Table **Error! Reference source not found.**). A common

327 task is the segmentation of brain tumors, which could be attributed to the availability of a public

328 database with dedicated training and test sets for use with the brain tumor segmentation

329 challenge held by the Medical Image Computing and Computer Assisted Intervention (MICCAI)

330   conference from 2014 to 2016, and continuing in 2017 and 2018. Methods evaluated on this data

331   set include patch-based auto-encoders,[115, 116] U-net-based structures,[117] as well as HNN.[88]

332

2.B. Detection

334   2.B.1.   **Organ detection**

335   Anatomical structure detection is a fundamental task in medical image analysis, which involves

336   computing the location information of organs and landmarks in 2D/3D image data. Localized

337   anatomical information can guide more advanced analysis of specific body parts or pathologies

338   present in the images, e.g. organ segmentation, lesion detection, and radio-therapy planning. In a

339   similar fashion to counterparts using traditional machine learning techniques, DL based organ /

340   landmark detection approaches can be mainly divided into two groups, i.e. classification based

341   methods and regression based ones. While classification based methods focus on discriminating

342   body parts / organs on the image or patch level, regression based methods target at recovering

343   more detailed location information, e.g., coordinates of landmarks. Table **Error! Reference**

344   **source not found.** illustrates a list of the DL based anatomical structure detection methods

345   together with their performance on different evaluation settings.

346   Early classification based approaches often utilized off-the-shelf CNN features to classify image

347   or image patches that contain anatomical structures. Yang et al.[135] adopted a CNN classifier to

348   locate 2D image patches (extracted from 3D MR volumes) that contain possible landmarks as an

349   initialization of the follow-up segmentation process for the femur bone. Chen et al.[136] adopted an

350   ImageNet pre-trained model and fine-tuned the model using fetal ultrasound frames from

351   recorded scan videos to classify the fetal abdominal standard plane images.

352   A variety of information in addition to original images could also be included to help the

353   detection task. For the same standard plane detection task in fetal ultrasound, Baumgartner et

354   al.[137] proposed a joint CNN framework to classify 12 standard scan planes and also localize the

355   fetal anatomy using a series of ultrasound fetal mid-pregnancy scans. The final bounding boxes

356   were generated based on the saliency maps computed as the visualization of network activation

357   for each plane class.

358   Improvements were also achieved by adapting the CNN network with more advanced

359   architecture and components. Kumar et al.[138] composed a two-path CNN network with features

360   computed from both original images and pre-generated saliency maps in each path. The final

361   standard plane classification was performed using SVM on a set of selected features.

362   Another category of methods tackle the anatomy detection problems with regression analysis

363   techniques. Ghesu et al.[139] formulated the 3D heart detection task as a regression problem,

364   targeting at the 3D bounding box coordinates and affine transform parameters in transesophageal

365   echocardiogram images. This approach integrated marginal space learning into the DL

366   framework and learned sparse sampling to reduce computational cost in the volumetric data

367   setting.[140]

368   Yan et al.[141, 142] formulated body part localization using DL. The system was developed using an

369   unsupervised learning method with two inter-sample CNN loss functions. The unsupervised

370   body part regression built a coordinate system for the body and output a continuous score for

371   each axial slice, representing the normalized position of the body part in the slice.

372   Besides the two common categories of methods discussed above, modern techniques (e.g.,

373   reinforcement learning) are also adopted to tackle the problem from a different direction. Ghesu

374   et al.[143] present a good example of combining reinforcement learning and DL in anatomical

375   detection task. With the application in multiple image data sets across a number of different

376   modalities, the method could search the optimal paths from a random starting point to the

377   predefined anatomical landmark via reinforcement learning with the help of effective

378   hierarchical features extracted via DCNN models. Furthermore, the system was further extended

379   to search 3D landmark positions with 3D volumetric CNN features.[144, 145] Later on, Xu et al.[146]

380   further extended this approach by turning the optimal action path searching problem into an

381   image partitioning problem, in which a global action map across the whole image was

382   constructed and learned by a DCNN network to guide the searching action.

383

384   2.B.2.   **Lesion detection**

385 Detection of abnormalities (including tumors and other suspicious growths) in medical images is

386 a common but costly and time-consuming part of the daily routine of physicians, especially

387 radiologists and pathologists. Given that the location is often not known a priori, the physician

388 should search across the 2D image or 3D volume to find deviations compared to surrounding

389 tissue and then to determine whether that deviation constitutes an abnormality that requires

390 follow-up procedures or something that can be dismissed from further investigation. This is often

391 a difficult task that can lead to errors in many situations either due to the vast amount of data that

392 needs to be searched to find the abnormality (e.g. in the case of volumetric data or whole-slide

393 images) or because of the visual similarity of the abnormal tissue with normal tissue (e.g. in the

394 case of low-contrast lesions in mammography). Automated computer detection algorithms have

395 therefore been of great interest in the research community for many years due to their potential

396 for reducing reading costs, shortening reading times and thereby streamlining the clinical

397   workflow, and providing quality care for those living in remote areas who have limited access to

398   specialists.

399   Traditional lesion detection systems often consist of long processing pipelines with many

400   different steps.[158, 159] Some of the typical steps include pre-processing the input data e.g. by

401   rescaling the pixel values or removing irrelevant parts of the image, identification of locations in

402   the image that are similar to the object of interest according to rule-based methods, extraction of

403   hand-crafted features, and classification of the candidate locations using a classifier such as SVM

404   or RF. In comparison, DL approaches for lesion detection are able to avoid the time-consuming

405   pipeline design approach. Table 4 presents a list of studies that used DL for lesion detection,

406   along with some details about the DL architecture.

407   Many of the papers focused on detection tasks use transfer learning with architectures from

408   computer vision.[160] Examples of this approach can be found in many publications, including

409   those for lesion detection in breast ultrasound,[161] for detection of bowel obstructions in

410   radiography,[162] and for detection of the third lumbar vertebra slice in a CT scan.[163] Usage of

411   CNNs in lesion detection is not limited to architectures taken directly from computer vision, but

412   also includes some applications where custom architectures are used.[164-167]

413   Most of the early applications used 2D CNNs, even if the data was 3D. Due to prior experience

414   with 2D architectures, limitations in the amount of available memory of GPUs, and higher

415   number of samples needed for training the larger number of parameters in a 3D architecture,

416   many DL systems used multi-view 2D CNNs for analysis of CT and MRI data sets in what is

417   referred to as 2.5D analysis. In these methods, orthogonal views of a lesion or multiple views at

418   different angles through the lesion were used to train an ensemble of 2D CNNs whose scores

419   would be merged together to obtain the final classification score.[166, 168] More recently, 3D CNNs

420   that use three-dimensional convolution kernels are successfully replacing 2D CNNs for

421   volumetric data. A common approach to deal with the small number of available cases is to train

422   the 3D CNNs on 3D patches extracted from each case. This way, each case can be used to extract

423   hundreds or thousands of 3D patches. Combined with various data augmentation methods, it is

424   possible to generate sufficient number of samples to train 3D CNNs. Examples of using 3D

425   patches can be found for detection of pulmonary nodules in chest CT,[169] and for detection of

426   lacunar strokes in brain MRI.[170] Due to the large size of volumetric data, it would be very

427   inefficient to apply the CNN in a sliding window fashion across the entire volume. Instead, once

428   the model is trained on patches the entire network can be converted into a fully convolutional

429   network[171] so that the whole network acts as a convolution kernel that can be efficiently applied

430   to an input of arbitrary size. Since convolution operations are highly optimized, this results in

431   fast processing of the entire volume when using a 3D CNN on volumetric data.[172]

432

433   **2.C. Characterization**

434   Over the past decades, characterization of diseases has been attempted with machine learning

435   leading to computer-aided diagnosis (CADx) systems. Radiomics, the –omics of images, is an

436   expansion of CADx to other tasks such as prognosis and cancer sub-typing. Radiomic features

437   can be described as (a) "hand-crafted"/"engineered"/"intuitive" features or (b) deep-learned

438   features. Characterization of disease types will depend on the specific disease types and the

439   clinical question. With hand-crafted radiomic features, the features are devised based on imaging

440   characteristics typically used by radiologists in their interpretation of a medical image. Such

441   features might include tumor size, shape, texture, and/or kinetics (for dynamic contrast-enhanced

442 imaging). Various review papers have already been written about these hand-crafted radiomic

443 features that are merged with classifiers to output estimates of, for example, the likelihood of

444 malignancy, tumor aggressiveness, or risk of developing cancer in the future.[158, 159]

445 DL characterization methods, on the other hand, may take as input a region of the image around

446 the potential disease site, such as a region of interest (ROI) around a suspect lesion. How that

447 ROI is determined will likely affect the training and performance of the DL. Thinking of how a

448 radiologist is trained during residency will lend understanding of how a DL system needs to be

449 trained. For example, an ROI that is cropped tightly around a tumor will provide different

450 information to a DL system than an ROI that is much larger than the encompassing tumor since

451 with the latter more anatomical background is also included in the ROI.

452 More and more DL imaging papers are published each year but there are still only a few methods

453 that are able to characterized among the vast range of radiological presentations across subtle

454 disease states. Table 5 presents a list of published DL characterization studies in radiological

455 imaging.

456

457   2.C.1.   **Lesion characterization**

458 When it comes to computer algorithms and specific radiological interpretation tasks, there is no

459 one-size-fits-all for either conventional radiomic machine learning methods or DL approaches.

460 Each computerized image analysis method requires customizations specific to the task as well as

461 the imaging modality.

462 Lesion characterization is mainly being conducted using conventional CAD/radiomics computer

463 algorithms, especially when the need is to characterize (i.e., describe) a lesion rather than

464   conduct further machine learning for disease assessment. For example, characterization of lung

465   nodules as well as characterization of the change in lung nodules over time, are used to track the

466   growth of lung nodules in order to eliminate false positive diagnoses of lung cancer.

467   Other examples involving computer characterization of tumors occurs in research in imaging-

468   genomics. Here, radiomic features of tumors are used as image-based phenotypes for correlative

469   and association analysis with genomics as well as histopathology. A well-documented, multi-

470   institutional collaboration on such was conducted through the TCGA/TCIA Breast Phenotype

471   Group.[220-224]

472   Use of DL methods as feature extractors can lend itself to tumor characterization; however, the

473   extracted descriptors (e.g., CNN-based features) are not intuitive. Similar to 'conventional'

474   methods that use hand-crafted features, DL-extracted features could characterize a tumor relative

475   to some known trait – such as receptor status – during supervised training, and that subsequent

476   output could be used in imaging-genomics discovery studies.

477   Additional preprocessing and data use methods can further improve characterization such as in

478   the past use of unlabeled data with conventional features to enhance the machine learning

479   training.[225, 226] Here, the overall system can learn aspects of the data structure without knowledge

480   of the disease state, leaving the labeled information for the final classification step.

481   2.C.2.   **Tissue characterization**

482   Tissue characterization is sought when specific tumor regions are not relevant. Here we focus on

483   analysis of non-diseased tissue to predict a future disease state (such as texture analysis on

484   mammograms in order to assess the parenchyma with the goal to predict breast cancer risk[159])

485   and characterization of tissue that includes diffuse disease, such as in various types of interstitial

486   lung disease and liver disease.[227, 228]

487   In breast cancer risk assessment, computer-extracted characteristics of breast density and/or

488   breast parenchymal patterns are computed and related to breast cancer risk factors. Using

489   radiomic texture analysis, Li et al. have found that women at high risk for breast cancer have

490   dense breasts with parenchymal patterns that are coarse and low in contrast.[229] DL is now being

491   used to assess breast density.[194, 195] In addition, parenchymal characterization is being conducted

492   using DL, in which the parenchymal patterns are related through the CNN architecture to groups

493   of women using surrogate markers of risk. One example is shown by Li et al. assessing the

494   performance of DL in the distinction between women at normal risk of breast cancer and those at

495   high risk based on their BRCA1/2 status.[192]

496   Lung tissue has been analyzed with conventional texture analysis and DL for a variety of

497   diseases. Here, characterization of the lung pattern lends itself to DL as patches of the lung may

498   be informative of the underlying disease, commonly interpreted by the radiologist's eye-brain

499   system. Various investigators have developed CNNs, including those to classify interstitial lung

500   diseases characterized by inflammation of the lung tissue.[207-209] These disease characterizations

501   can include healthy tissue, ground glass opacity, micronodules, consolidation, reticulation, and

502   honeycombing patterns.[179]

503   Assessing liver tissue lends itself to DCNNs in the task of staging liver fibrosis on MRI by

504   Yasaka et al.[216] and ultrasonic fatty liver disease characterization by Bharath et al.[217]

505   2.C.3.   **Diagnosis**

506 Computer-Aided Diagnosis (CADx) involves the characterization of a region or tumor, initially

507 indicated by either a radiologist or a computer, after which the computer characterizes the

508 suspicious region or lesion and/or estimates the likelihood of being diseased (e.g., cancerous) or

509 non-diseased (e.g., non-cancerous), leaving the patient management to the physician.[158, 159] Note

510 that CADx is not a localization task but rather a characterization/classification task. The subtle

511 difference between this section and the preceding two sections, is that here the output of the

512 machine learning system is related to the likelihood of disease and not just a characteristic

513 feature of the disease presentation.

514 Many review papers have been written over the past two decades on CADx, radiomic features,

515 and machine learning,[158, 159] and thus details will not be presented in this paper.

516 An active area of DL is CADx of breast cancer. Training CNNs "from scratch" is often not

517 possible for CAD and other medical image interpretation tasks, and thus methods to use CNNs

518 trained on other data (transfer learning) are considered. Given the initial limited data sets and

519 variations in tumor presentations, investigators explored the use of transfer learning to extract

520 tumor characteristics using CNNs trained on nonmedical tasks. The outputs from layers can be

521 considered as characteristic features of the lesion and serve as input to classifiers, such as linear

522 discriminant analysis and support vector machines. Fig. 3a shows an example in which AlexNet

523 is used as a feature extractor for an SVM, and Fig. 3b shows the performance of the SVM based

524 on features from each layer of AlexNet.

525 Researchers have found that performance of the conventional radiomics CADx and that of the

526 CNN-based CADx yielded similar levels of diagnostic performance in the task of distinguishing

527 between malignant and benign breast lesions, and thus when combined, via a deep feature fusion

528 methodology, gave a statistically significant level of performance.[196, 197] Fig. 4 shows one

529 possible method for combining CNN-extracted and conventional radiomics features.

530

531 In an effort to augment, under limited data set constraints, CNN performance with dynamic

532 contrast-enhanced MRI, investigators have looked to vary the image types input to the CNN. For

533 example, instead of replicating a single image region to the three RGB channels of VGG19Net,

534 investigators have used the temporal images obtained from DCE-MRI, inputting the pre-contrast,

535 the first post-contrast, and the second post-contrast MR images to the RGB channels,

536 respectively. In addition, to exploit the 4D nature of DCE-MRI (3D and temporal), Antropova et

537 al. have input MIP (maximum intensity projections) images to the CNN.[200] Incorporation of

538 temporal information into the DL efforts has resulted in the use of recurrent neural network, such

539 as a long short-term memory (LSTM) recurrent networks.[201, 230]

540 Instead of using transfer learning for feature extraction, investigators have used transfer learning

541 for fine tuning by either (i) freezing the earlier layers of a pre-trained CNN and training the later

542 layers, i.e., fine tuning or (ii) training on one modality, such as digitized screen/film

543 mammography (dSFM), for use on a related modality, such as full-field digital mammography

544 (FFDM). The latter has been shown by Samala et al.[199] to be useful in the training of CNN-based

545 CADx for lesion diagnosis on FFDMs.

546 Investigations on DL for CADx are continuing across other cancers, e.g.., lung cancer, and other

547 disease types, and similar methods can be used.[204-219] The comparison to more conventional

548 radiomics-based CADx is also demonstrated further, which is potentially useful for both

549 understanding the CNN outputs as well as for providing additional decision support.

550   2.C.4.   **Prognosis and staging**

551 Once a cancer is identified, further workup through biopsies gives information on stage,

552 molecular subtype, and/or genomics to yield information on prognosis and potential treatment

553 options. Cancers are spatially heterogeneous, and therefore, investigators are interested whether

554 imaging can provide information on that spatial variation. Currently, many imaging biomarkers

555 of cancerous tumors include only size and simple enhancement measures (if dynamic imaging is

556 employed) and, thus, there is interest in expanding, through radiomic features, the knowledge

557 that can be obtained from images. Various investigators have used radiomics and machine

558 learning in assessing the stage and prognosis of cancerous tumors.[220, 231] Now, those analyses are

559 being investigated further with DL. It is important to note that when using DL to assess

560 prognosis, one can analyze the tumor from medical imaging, such as MRI or ultrasound, or from

561 pathological images. Also, in the evaluation, one needs to determine the appropriate comparison

562 – a radiologist, a pathologist, or some other histopathological/genomics test.

563 The goal is to better understand the imaging presentation of cancer, i.e., to obtain prognostic

564 biomarkers from image-based phenotypes, including size, shape, margin morphology,

565 enhancement texture, kinetics, and variance kinetic phenotypes. For example, enhancement

566 texture phenotypes can characterize the tumor texture pattern of contrast-enhanced tumors on

567 DCE-MRI though analysis the first post-contrast images, and thus quantitatively characterize the

568 heterogeneous nature of contrast uptake within the breast tumor.[220] Here, the larger the

569 enhancement texture entropy, the more heterogeneous is the vascular uptake pattern within the

570 tumor, which potentially reflects the heterogeneous nature of angiogenesis and treatment

571 susceptibility, and serves as a location-specific "virtual digital biopsy". Understanding the

572 relationships between image-based phenotypes and the corresponding biopsy information could

573 potentially lead to discoveries useful for assessing images obtained during screening as well as

574 during treatment follow-up, i.e., when an actual biopsy is not practical.

575 Shi et al.[203] demonstrated the prediction of prognostic markers using DL on mammography in

576 distinguishing between DCIS with occult invasion from pure DCIS. Staging on thoracic CTs is

577 being investigated by Masood et al. through DL by relating CNN output to metastasis

578 information for pulmonary nodules.[210] In addition, Gonzalez et al. evaluated DL on thoracic CTs

579 in the detection and staging of chronic obstructive pulmonary disease (COPD) and acute

580 respiratory disease (ARD).[211] While use of DL in the evaluation of thoracic CTs is promising,

581 more development is needed to reach clinical applicability.

582 2.C.5. **Quantification**

583 Use of DL in quantification requires a CNN output that correlates significantly with a known

584 quantitative medical measurement. For example, DL has been used in automatic calcium scoring

585 in low-dose CTs by Lessmann et al.[212] and in cardiac left ventricle quantification by Xue et al.[213]

586 Similar to cancer workup, in cardiovascular imaging, use of DL is expected to augment clinical

587 assessment of cardiac defect/function or uncover new clinical insights.[232] Larson et al. turned to

588 DL to assess skeletal maturity on pediatric hand radiographs with performance levels rivaling

589 that of an expert radiologist.[219] DL has been used to predict growth rates for pancreatic

590 neuroendocrine tumors[233] on PET-CT scans.

591 **2.D. Processing and reconstruction**

592 In the previous parts of this section, we focused on applications in which image pixels or ROIs

593 are classified into multiple classes (e.g., segmentation, lesion detection and characterization), the

594 subject is classified into multiple classes (e.g., prognosis, staging), or a feature in the image (or

595 the ROI) is quantified. In this part, we focus on applications in which the output of the machine

596 learning algorithm is also an image (or a transformation) that potentially has a quantifiable

597 advantage over no processing or traditional processing methods.

598

599 ### 2.D.1. **Filtering, noise/artifact reduction, and reconstruction**

600 Filtering: Going back to the early days of application of CNNs to medical images, one can find

601 examples of CNNs that produced output images for further processing. Zhang et al.[52] trained a

602 shift-invariant ANN that aimed at having a high or low pixel value in an output image depending

603 on whether the pixel was determined to be the center of a microcalcification by an expert

604 mammographer. Suzuki et al.[234] trained an MTANN as a supervised filter for the enhancement

605 lung nodules on thoracic CT scans. More recently, Yang et al.[235] used a cascade of CNNs for

606 bone suppression in chest radiography. Using ground-truth images extracted from dual-energy

607 subtraction chest X-rays, the authors trained a set of multi-scale networks to predict bone

608 gradients at different scales and fuse these results to obtain a bone image from a standard chest

609 x-ray. Another advantage of CNNs for image filtering is speed: Mori[236] investigated several

610 types of residual convolutional autoencoders and residual CNNs for contrast-limited adaptive

611 histogram equalization filtering and denoising of X-ray fluoroscopic imaging during treatment,

612 without specialized hardware.

613 Noise reduction: The past couple of years have seen a proliferation of applications of DL to

614 improve the noise quality of reconstruction medical images. One application area is low-dose

615 image reconstruction. This is important in modalities with ionizing radiation such as CT or PET

616 for limiting patient dose,[237-239, 241] or for limiting damage to samples in synchrotron-based X-ray

617   CT.[240] Chen et al.[237] designed a DL algorithm for noise reduction in reconstructed CT images.

618   They used the mean-squared pixelwise error between the ideal image and the denoised image as

619   the loss function, and synthesized noisy projections based on patient images to generate training

620   data.[238] They later combined a residual autoencoder with a CNN in an architecture called the

621   RED-CNN[238], which has a stack of encoders and a symmetrical stack of decoders that are

622   connected with shortcuts for the matching layers. Kang et al.[239] applied a DCNN to the wavelet

623   transform coefficients of low-dose CT images, and similar to the work of Chen et al.[238], used a

624   residual learning architecture for faster network training and better performance. Their method

625   won the second-best place at the 2016 "Low-Dose CT Grand Challenge.[266] Xiang et al used low-

626   dose PET images combined with T1-weighted images acquired on a PET/MRI scanner to obtain

627   standard acquisition quality PET images. In comparison to the papers above that started

628   denoising with reconstructed images, Yang et al. aimed at improving the quality of recorded

629   projections. They used a CNN-based approach for learning the mapping between a number of

630   pairs of low- and high-dose projections. After training with a limited number of high-dose

631   training examples, they used the trained network to predict high-dose projections from low-dose

632   projections, and then used the predicted projections for reconstruction.

633   Artifact reduction: Techniques similar to those described for denoising have been applied to

634   artifact reduction. Jin et al.[242] described a general framework for the utilization of CNNs for

635   inverse problems, applied the framework to reduce streaking artifacts in sparse-view

636   reconstruction on parallel beam CT, and compared their approach to filtered-backprojection

637   (FBP) and total variation (TV) techniques. Han et al.[244] used DL to reduce streak artifacts

638   resulting from limited number of radial lines in radial k-space sampling in MRI. Zhang et al.[243]

639   used a CNN-based approach to reduce metal artifacts on CT images. They combined the original

640   uncorrected image with images corrected with the linear interpolation and beam hardening

641   correction methods to obtain a three-channel input. This input was fed into a CNN, whose output

642   was further processed to obtain "replacement" projections for the metal-affected projections.

643   Reconstruction: Several studies indicated that DL may be useful in directly attacking the image

644   reconstruction problem. In one of the early publications in this area, Golkov et al.[245] applied a

645   deep-learning approach to diffusion-weighted MR images (DWI) to derive rotationally invariant

646   scalar measures for each pixel. Hammernik at al.[246] designed a variational network to learn a

647   complete reconstruction procedure for multi-channel MR data, including all free parameters

648   which would otherwise have to be set empirically. To obtain a reconstruction, the undersampled

649   k-space data, coil sensitivity maps and the zero-filling solution are fed into the network.

650   Schlemper et al.[247] evaluated the applicability of CNNs for reconstructing undersampled

651   dynamic cardiac MR data.. Zhu et al.[248] introduced an automated transform by manifold

652   approximation approach to replace the conventional image reconstruction with a unified image

653   reconstruction framework that learns the reconstruction relationship between sensor and image

654   domain without expert knowledge. They showed examples in which their approach resulted in

655   superior immunity to noise and a reduction in reconstruction artifacts compared with

656   conventional reconstruction methods.

657   2.D.2.   **Image registration**

658   To establish accurate anatomical correspondences between two medical images, both hand-

659   crafted features and features selected based on a supervised method are frequently employed in

660   deformable image registration. However, both types of features have drawbacks.[249] Wu et al.[249]

661   designed an unsupervised DL approach to directly learn the basis filters that can effectively

662   represent all observed image patches, and used the coefficients by these filters for

663 correspondence detection during image registration. They subsequently further refined the

664 registration performance by using a more advanced convolutional stacked autoencoder, and

665 comprehensively evaluated the registration results with respect to current state-of-the-art

666 deformable registration methods.[250] A deep encoder-decoder network was used for predictions

667 for the large deformation diffeomorphic metric mapping model by Yang et al.[251] for fast

668 deformable image registration. In a feasibility study, Lv et al.[252] trained a CNN for respiratory

669 motion correction for free-breathing 3D abdominal MRI. For the problem of 2D/3D registration,

670 Miao et al.[253] used a supervised CNN regression approach to find a rigid transformation from the

671 object coordinate system to the x-ray imaging coordinate system. The CNNs were trained using

672 synthetic data only. The authors compared their method with for intensity-based 2-D/3-D

673 registration methods and a linear regression- based method, and showed that their approach

674 achieved higher robustness and larger capture range, as well as higher computational efficiency.

675 A later study by the same research group identified a performance gap when the model trained

676 with synthetic data is tested on clinical data.[254] To narrow the gap, the authors proposed a

677 domain adaptation method by learning domain invariant features with only a few paired real and

678 synthetic data.

679   2.D.3. **Synthesis of one modality from another**

680 A number of studies have recently investigated using DL to generate synthetic CT (sCT) images

681 from MRI. This is important for at least two applications: First, for accurate PET image

682 reconstruction and uptake quantification, tissue attenuation coefficients can be readily estimated

683 from CT images. Thus, estimation of sCT from MRI in PET/MRI imaging is desirable. Second,

684 there is an interest in replacing CT with MRI in the treatment planning process mainly because

685 MRI is free of ionizing radiation. Nie et al.[255] used a 3D CNN to learn an end-to-end nonlinear

686 mapping from an MR image to a CT image. The same research group in their later research

687 added a context-aware GAN for improved results.[259] Han et al.[256] adopted and modified the U-

688 net architecture for sCT generation from MRI. Current commercially available MR attenuation

689 correction (MRAC) methods for body PET imaging use a fat/water map derived from a two-echo

690 Dixon MRI sequence. Leynes et al.[257] used multi-parametric MRI consisting of Dixon MRI and

691 proton-density-weighted zero (ZTE) echo-time MRI to generate sCT images with the use of a

692 DL model that also adopted the U-net architecture.[267] Liu et al.[258] trained a deep network (deep

693 MRAC) to generate sCT from T1-weighted MRI, and compared deep MRAC with Dixon

694 MRAC. Their results showed that significantly lower PET reconstruction errors were realized

695 with deep MRAC. Choi et al.[260] investigated a different type of synthetic image generation. They

696 noted that although PET combined with MRI is useful for precise quantitative analysis, not all

697 subjects have both PET and MR images in the clinical setting, and used a GAN-based method to

698 generate realistic structural MR images from amyloid PET images. Ben-Cohen et al.[261] aimed at

699 developing a system that can generate PET images from CT, to be used in applications such as

700 evaluation of drug therapies and detection of malignant tumors that require PET imaging, and

701 found that a conditional GAN is able to create realistic looking PET images from CT.

702   2.D.4.   **Quality assessment**

703 In addition to traditional characterization tasks in medical imaging, such as classification of ROIs

704 as normal or abnormal, DL has been applied to image quality assessment. Wu et al.[262] proposed a

705 DCNN for computerized fetal US image quality assessment to assist the implementation of US

706 image quality control in the clinical obstetric examination. The proposed system has two

707 components: The L-CNN that locates the ROI of the fetal abdominal region in the US image, and

708 the C-CNN evaluates the image quality by assessing the goodness of depiction for the key

709   structures of stomach bubble and umbilical vein. Neylon et al.[263] used a deep neural network as

710   an alternative to image similarity metrics to quantify deformable image registration performance.

711   Since the image quality strongly depends on both the characteristics of the patient as well as the

712   imager, both of which are highly variable, using simplistic parameters like noise to determine the

713   quality threshold is challenging. Lee et al.[264] showed that DL using fine-tuning of a pre-trained

714   VGG19 CNN was able to predict whether CT scans meet the minimal image quality threshold

715   for diagnosis, as deemed by a chest radiologist.

716   Esses et al.[265] used a DCNN for automated task-based image quality evaluation of T2-weighted

717   liver MRI acquisition, and compared this automated approach to image quality evaluation by two

718   radiologists. Both the CNN and the readers classified a set of test images as diagnostic or non-

719   diagnostic. The concordance between the CNN and reader 1 was 0.79, that between the CNN and

720   reader 2 was 0.73, and that between the two readers was 0.88. The relatively lower concordance

721   of the CNN with the readers was mostly due to cases that the readers agreed to be diagnostic, but

722   the CNN did not agree with readers. The authors concluded that although the accuracy of the

723   algorithm needs to be improved, the algorithm could be utilized to flag cases as low-quality

724   images for technologist review.

725   **2.E. Tasks involving imaging and treatment**

726   Radiotherapy and assessment of response to treatment are not areas that are traditionally

727   addressed using neural networks or data-driven approaches. However, these areas have recently

728   seen a strong increase in the application of deep learning techniques. Table 7 summarizes

729   studies in this fast-developing DL application area.

730

731 ### 2.E.1.   **Discovery: Imaging-genomics (Radiogenomics)**

732 A major need in breast cancer research is the elucidation of the relationship between the

733 macroscopic image-based presentation of the tumor and its environment and cancer biology

734 indicators of risk, diagnosis, prognosis, or treatment response. Imaging-genomics, i.e.,

735 "radiogenomics", aims to find these relationships between imaging data and clinical data,

736 molecular data, genomic data, and outcome data.[222, 224] Of interest is whether DL can provide

737 sufficient detailed information to relate to genetic data as have hand-crafted radiomic

738 phenotypes.[285]

739 ### 2.E.2.   **Radiotherapy**

740 The goals of DL in radiation oncology are to assist in treatment planning, assess response to

741 therapy, and provide automated adaptation in treatments over time. Deep reinforcement learning

742 using both prior treatment plans and methods for assessing tumor local control were used to

743 automatically estimate dose protocols.[278] Such adaptive radiotherapy methods may provide

744 clinical decision support for dose adaptation.

745 Much of the needs in treatment planning relate to the segmentation of organs (discussed earlier)

746 and in the prediction of dose distributions from contours. Nguyen et al.[280] used a U-net to predict

747 dose from patient image contours on prostate intensity-modulated radiation therapy (IMRT)

748 patients, and demonstrated desired radiation dose distributions. Foote et al.[279] combined a DCNN

749 with motion tracking to recover anatomical positions from a single projection radiographic image

750 in real time in order to achieve dynamic tracking of a lung tumor volume.

751 As discussed earlier, DL can be used to convert between modalities (Section 2.D.3), which can

752 benefit both diagnosis and therapy. Maspero et al.[282] have developed a DL method for creating

753 synthetic CTs from MR-only radiotherapy, leading to online adaptive replanning. Such methods,

754 in order to allow for real time changes, need to rapidly generate synthetic CTs, thus modeling the

755 radiation attenuation and dose calculations.

756 While DL methods are being developed to plan and predict radiation therapy to specific tumor

757 sites, they are also being investigated to assess toxicity to normal organs and tissue. Zhen et al.[283]

758 used a transfer learning strategy to predict rectum dose toxicity for cervical cancer radiotherapy.

759 Segmentation methods to aid in the assessment of treatment plans have been developed as well;

760 Tong et al. developed a CNN-based method for multi-organ segmentation for use in head and

761 neck cancer radiotherapy[274], Men et al developed a target tumor volume segmentation for rectal

762 cancer[272] and breast cancer,[286] while Jackson et al. focused on renal segmentation for automated

763 radiation dose estimation.[275] Dose estimation was also the aim of Kajikawa et al. who

764 investigated the feasibility of DL in the automated determination of dosimetric eligibility of

765 prostate cancer patients undergoing intensity modulated radiation therapy.[281]

766 Just as with imaging-genomics, as discussed earlier, incorporation of both image-based

767 phenotypes and genomics in treatment planning and response assessment may yield new

768 relationships and improved therapeutics.[273]

769 Overall, however, use of DL in radiation planning is still at a very early stage in development.

770   2.E.3. **Response to treatment**

771 Just as DL is used to extract tumor characteristics for diagnosis and prognosis, it can also be used

772 in decision making for assessing response to therapy. In machine learning, various classifiers can

773 be used to merge the tumor image-based phenotypes into a response prediction. Thus, DL can

774 also be used to analyze medical image(s) over time to predict response. For example, CNNs were

775   used with breast DCE-MRI to assess response to neoadjuvant chemotherapy, where the inputs

776   varied over contrast time points as well as treatment exam times.[270]

777   Cha et al.[268] have explored the feasibility of DL through CNNs on pre- and post-treatment CT of

778   bladder cancer patients to assist in assessment of treatment response. In addition, assessing

779   prognosis of a tumor contributes to decision making on treatment options and predicting

780   survival. Lao et al.[218] investigated MRI radiomic features and DL as a means to predict survival

781   in glioblastoma multiforme. Bibault et al. used DL to predict pathologic complete response after

782   chemoradiation in locally advanced rectal cancer,[284] while Ibramov et al. predicted hepatobiliary

783   toxicity after liver stereotactic body radiotherapy.[277] In research unrelated to oncology the

784   interest in using DL to assess response to treatment has increased as well. Shehata et al.[276] used

785   autoencoders for early detection/prediction of acute renal rejection after kidney transplant.

786   Nielsen et al. used DL to predict outcome and to assess the effect of treatment with recombinant

787   tissue-type plasminogen activator in ischemic stroke patients.[269]

788   **3. COMMON THEMES**

789   **3.A. Training and testing with size-limited data sets**

790   The rapid and immense success of DCNNs in many challenging computer vision problems is

791   achieved through accessibility to large-scale well-annotated data sets, e.g., PASCAL VOC,[287]

792   ImageNet[28] and MS COCO.[288] ImageNet pre-trained DCNN models[29, 73] serve as the foundation

793   in many higher level tasks, e.g. image captioning,[289] visual question answering,[290] and instance

794   relationship extraction.[291] Compared to natural image data sets, existing medical image data sets

795   are typically smaller in size. This is because the collection of medical image data sets is often a

796   challenging, time consuming process, which involves multiple steps, such as searching in large

797 hospital PACS systems with moderately structured clinical information, selection of a relatively

798 small number of useful clinical cases, and further data annotation by expert physicians. In this

799 sub-section, we explore some of the challenges for applying DL on relatively small data sets.

800 The concepts and principles discussed below, such as overfitting, the need for independent

801 training and test data sets, and dependence of performance on training data set size, apply to

802 most machine learning algorithms, including traditional (shallow) neural networks. However,

803 some aspects may be exacerbated due to the large number of tunable parameters in DL networks.

804 Overfitting: It has long been recognized that training a complex classifier with a small data set

805 invites the risk of overfitting (also termed overtraining). According to the Oxford English

806 dictionary overfitting is "the production of an analysis that corresponds too closely or exactly to

807 a particular set of data, and may therefore fail to fit additional data or predict future observations

808 reliably". In other words, overfitting occurs when a classifier models the training data too well,

809 resulting in it failing to generalize and performing poorly on new unseen data. John von

810 Neumann famously said 'With four parameters I can fit an elephant, and with five I can make

811 him wiggle his trunk'.[292] Both shallow neural networks and DL exhibit overtraining.

812 Surprisingly, compared to the huge number of tunable parameters in DL networks, they may

813 exhibit a more limited amount of overfitting compared to a shallow network designed to achieve

814 the same functionality. One possible explanation for this, as discussed in the introduction, is that

815 DL learns a hierarchical representation that matches the composition of the individual

816 components that the data consists of.[293] Another possible explanation, using concepts from

817 information theory, contends that a deep networks helps better compress the irrelevant

818 information in the input data and thus can achieve better generalization.[294]

819 A number of ways have been suggested in the literature to reduce overfitting, including

820 regularization,[295] early stopping,[296] and drop-out.[11, 26] Regularization involves the addition of an

821 extra term to the loss function during training akin to the use of a Lagrange multiplier to satisfy

822 certain boundary conditions. The regularization term is typically chosen to penalize overly

823 complex solutions and for example imposes rules for the smoothness of the solution. Early

824 stopping can be seen as regularization in time. The longer a network is trained, the more complex

825 its solutions become, so by regularizing on time (through early stopping) the complexity will be

826 reduced and generalizability improved. When to stop training is usually determined by

827 monitoring the loss on a validation set (see next paragraph). Dropout is another very efficient

828 way to prevent overfitting and the term "dropout" refers to dropping out units in a neural

829 network.

830 Training, validation and testing: Ideally, one has access to three large independent data sets to

831 serve as training, validation, and test set for the training and evaluation of any machine learning

832 approach. Although the terms 'validation set' and 'test set' may not be defined consistently

833 among all communities, here we use the term 'validation set' for the set used for fine-tuning as

834 part of training and 'test set' for the set used for final performance evaluation. Fig. 5 shows how

835 the training, validation, and test sets can be used in a supervised machine learning system in an

836 ideal scenario with a large number of available cases. However, when the total number of

837 available cases is small, such a scenario may be inadequate to make full use of the limited-size

838 data set. For example, if a total of a hundred cases is available, then it may not be reasonable to

839 randomly assign 20% as a test set and divide the remaining 80 cases into training and validation.

840 The statistical variability of the classification performance for 20 cases will typically be large,

841 limiting the usefulness of the reported performance. Instead, it may be clinically more useful to

842   use a cross-validation approach (with multiple training/validation and testing data splits) for

843   obtaining a more realistic performance estimate. Using a cross-validation training/validation and

844   testing approach is a way to obtain a realistic performance estimate for the entire data set when

845   done correctly but does not result in a single model. Care must be taken to perform all training

846   and validation steps only within the training fold of the cross-validation, so that there is no

847   leakage of information from the different folds into each other that might bias the cross-

848   validation performance estimate. In Section 4, methods to help overcome problems related to

849   training DL on a small data set are discussed, but one should keep in mind that these methods do

850   not overcome the most important limitation of having a small data set, i.e., that the small sample

851   may not accurately represent the population of interest.

852   Dependence of test performance on training set size: A number of studies in the literature have

853   investigated the effect of training size on the performance of the machine learning system.[297-301]

854   The general trend is that as the number of training cases increases, overtraining decreases and the

855   performance on the targeted population improves. There is also a number beyond which

856   increasing the training set size only marginally improves the test performance. However, this

857   number is believed to be a function of the machine learning system architecture, the task, and the

858   system inputs. A few papers studied the effect of varying the training set size on the performance

859   of their DL network.[16, 63, 193, 302, 303] Mohamed et al.[193] found that for breast density classification,

860   there is a small increase in test performance (the area under the receiver operating characteristic

861   curve increases from 0.95 to 0.99, $p < 0.001$) when their training set size increased from 2000

862   images to 6000 images. Azizi et al.[16] also found that increasing the training data set increased the

863   performance of a DL model used for prostate cancer detection in ultrasound Gulshan et al.[63]

864   showed that for their detection algorithm of diabetic retinopathy in retinal fundus photographs,

865   the relative specificity at a given sensitivity on their validation set consistently increased as the

866   number of training samples increased from around 200 samples to around 60,000 samples, at

867   which point the performance plateaued. Using natural images data sets, where the available

868   labeled data are much more abundant compared to medical images, Sun et al.[303] demonstrated

869   that the test performance of the DL network continued to increase when going from 10 million

870   training samples up to 300 million training samples for both object detection and semantic

871   segmentation tasks. While it is difficult to obtain data sets of annotated medical images similar in

872   size to data sets for natural images, the trend that increasing the training data set size increases

873   the performance of the DL network on a target population still applies.

874

875   **3.B. Transfer learning and fine tuning**

876   Transfer learning is a technique in which a DL network trained on a large data set from one

877   domain is used to retrain or fine-tune the DL network with a smaller data set associated with

878   another domain.[160] The limited size of the annotated medical image data sets, and the current

879   trend of using deeper and larger structures increase the risk of overtraining and makes transfer

880   learning more appealing in medical imaging.

881   Transfer learning in medical imaging commonly starts with a CNN that was already trained on

882   natural images, i.e., a pre-trained model. The limited medical image data set is then used to fine-

883   tune the pre-trained model or, in some applications, no fine-tuning is performed at all. During

884   fine-tuning, the DL architecture typically remains fixed, and only a subset of the weights may be

885   re-trained.

886   A commonly used data set for pre-training of DL structures is ImageNet[28] composed of natural

887   scene images. It has been used in more than 75% of the reported transfer learning studies.

888   Different data sets also used for pre-training include CIFAR-10,[204] Places205,[304] and texture data

889   sets, such as ALOT, DTD, FMD, and KTH-TIPS-2b, as discussed in the literature.[209]

890   Transfer learning within the same domain of the target task has also been performed. Kooi et

891   al.[202] pre-trained DCNN on a large mammogram data set and then re-trained the DCNN on a

892   different smaller mammogram data set for the task of discriminating benign solitary cysts from

893   malignant masses in digital mammography. Samala et al. first pre-trained a DCNN on

894   ImageNet[198] or a larger mammogram data set[17] and then fine-tuned on a digital breast

895   tomosynthesis (DBT) data set for classification and detection of masses on DBT. Zheng et al.[254]

896   pre-trained on synthetic data and retrained on clinical data for two-dimensional to three-

897   dimensional (2D/3D) registration of preoperative 3D image data. Azizi et al.[16] used

898   radiofrequency (RF) ultrasound images as a source domain to pre-train the DCNN and fine-tuned

899   it on B-mode images as a target domain for prostate cancer detection.

900   A number of studies used pre-trained CNNs for extracting features, which are sometimes

901   referred to as the off-the-shelf CNN features.[305] A relatively small labeled data set can then be

902   used to train a classifier such as an SVM for the problem at hand. A number of studies[173, 181, 192,

903   196, 306-308] extracted the outputs of the fully-connected layers of a DL network that has been pre-

904   trained ImageNet, and used those features as input to SVMs to build classification models, which

905   suggests that a network pre-trained on natural images is useful for extracting features for medical

906   image analysis purposes.

907 Many of the studies that use transfer learning fine-tune their models by performing additional

908 training on all the network layers, thus using transfer learning like a weight-initialization step.

909 With the assumption that the earlier layers perform more common filtering tasks and later layers

910 (usually fully connected layers) focus more on semantic and high-level features for specific

911 purposes, others have fine-tuned only a few of the last layers within the network.[110] Samala et

912 al.[199] studied the effects of fine-tuning different layers of the AlexNet architecture, and found

913 that fine-tuning different layer combinations resulted in different performance. For their task,

914 they found that freezing the weights of just the first convolution layer achieved higher

915 performance compared to freezing additional layers, or fine-tuning all the convolution layers.

916 Similar trends were observed by Lee et al.[309]. However, the data set size for the fine-tuning may

917 also need to be taken into consideration when using transfer learning, as Samala et al.[310] saw a

918 trend where the performance of the fine-tuned network increased with increasing data set size of

919 the target task domain used for fine-tuning.

920 **3.C. Combining deep learning with radiomics approaches**

921 Before DL was applied to medical imaging, hand-crafted-features-based approaches were

922 generally used to analyze the images. By using DL, it is expected that given enough data, the

923 network will learn image descriptors useful for analysis. However, it is possible to combine the

924 outputs of DL methods with the knowledge the field of medical imaging analysis has

925 accumulated with computer-extracted, hand-crafted features.[166] Several works, including

926 Antropova et al.,[197] Li et al.,[192] Huynh et al.,[196] and Ben-Cohen et al.,[307] combined features

927 extracted from the fully-connected layers of a DL architecture, with traditional hand-crafted

928 features (morphology, intensity, texture). Feature selection was performed to reduce the number

929 of features, then a machine learning classifier, such as SVM or RF, were used to generate a

930   model using the extracted features. These studies suggest that supplementing DL with

931   information already known to be useful, may improve the performance of these DL models.

932   **3.D. Supervised / Weakly supervised / Unsupervised learning**

933   The majority of the DL applications utilize supervised learning: there is ground truth or labels

934   that the system is trying to match. However, there are also unsupervised methods that attempt to

935   draw inferences from unlabeled data, i.e., without the help of a supervisor (or label) that provides

936   a degree of error for each observation, and weakly-supervised methods, that use noisy labels, or

937   images labeled as positive or negative, without localization information, to train for a specific

938   task.

939   Unsupervised learning in DL is generally performed by auto-encoders or independent subspace

940   analysis (ISA).[249, 250, 311] The outputs of these networks may be further processed in a supervised

941   manner, by extracting the features from the network and applying a machine learning classifier.

942   In weakly-supervised learning, the reference standard used to train does not contain the full

943   information.[311, 312] For example, Feng et al.[313] trained a system for lung nodule segmentation

944   with a binary label if a nodule was present for a given image slice. Yang et al.[167] used a weakly-

945   supervised network in a system that aimed to generate a cancer response map with each pixel

946   indicating the likelihood to be cancerous. Both methods refined the initial results with additional

947   deep learning networks. There are also methods that use a combination of weakly supervised and

948   supervised methods.[180, 314] Wang et al.[180] and Rajpurkar et al.[314] used supervised learning to label

949   chest x-rays with one or multiple specific lung diseases, and used weakly-supervised learning to

950   localize the region with the disease.

951   **4. EXPANDING DATA SETS FOR DEEP LEARNING**

952   As discussed above, DL performs significantly better than previous shallow learning methods

953   and hand-crafted image features. However, this comes at the cost of requiring greater amounts of

954   training data compared to previous methods. In the medical domain, publicly-available large-

955   scale image data sets that contain images from tens of thousands of patients are not available

956   (except the recently released ChestX-ray14 data set.[180]) Although vast amounts of clinical

957   images/annotations/reports are stored in many hospitals' digital warehouse, e.g., picture

958   archiving and communication systems (PACS) and oncology information system (OIS),

959   obtaining semantic labels on a large scale medical image database is another bottleneck to train

960   highly effective DL models for image analysis.

961   It is difficult to directly borrow conventional means of collecting image annotations that are used

962   for annotating natural scene images (e.g., Google image search uses terms from NEIL

963   knowledge[315] base followed by crowd-sourcing[28]) and apply them in medical images. Medical

964   annotations are difficult to obtain from clinically untrained annotators. On the other hand, using

965   well-trained radiologists is expensive. Moreover, the task of "assigning labels to images" is not

966   aligned with their regular clinical routine, which can cause drastic inter-observer variations or

967   inconsistency. There is a lot of definition ambiguity to assign image labels based on visible

968   anatomic structures, pathological findings or using both cues. In addition, a high quality or large

969   capacity medical image search engine is a prerequisite to locate relevant image studies. For

970   example, the radiological data stored in the PACS server are only indexed with dates, patient

971   names, and scan protocols, and it often takes extra effort to find all the cases with a disease

972   pattern of interest. Natural language processing based systems that text mine radiology reports

973   are just beginning to become available.[316]

974   A wide variety of techniques have been developed for tackling the data shortage problem for

975   both the general computer vision and medical image analysis domain. Data augmentation is the

976   most straightforward way to increase the size of a data set for training purposes. It has been

977   proved to be extremely effective for currently existing data sets,[160] which often contain a small

978   number (hundreds of cases) of hand-labeled data. Others believe that DL and humans-in-the-loop

979   inspection may have to be interleaved and integrated to construct labels for a large-scale image

980   database, rather than being employed as two independent labeling processes. It can involve

981   selectively labeling critical samples via active learning. A few recent works focus on transferring

982   the tremendous number of imaging studies accompanied by radiological reports (i.e., loosely

983   labeled samples) into machine trainable data format. Both image and textual features could be

984   utilized for this retrospective and cost-effective process. In addition to using hand-labeled

985   ground-truth, others[317, 318] utilize the algorithm-generated ground-truth of existing image data for

986   training the CNN models. They assume the model can learn from these less accurate examples

987   and produce refined results in an iterative training process. Furthermore, approaches based on

988   generative adversarial networks[38] (GAN) can create image samples for training, either from

989   random initialization or from more advanced clues for image generation. Recent results have

990   shown examples of its promising and useful outcomes. In the following sections, we will

991   summarize these techniques individually.

992   **4.A. Data augmentation**

993 Data augmentation creates new samples based on existing samples in a data set or according to a

994 generative model. These new samples can then be combined with the original samples to

995 increase the variability of data points in a data set. This class of techniques has become a

996 common practice in DL based applications since it has been shown to be extremely effective for

997   increasing the size of training sets, reducing the chance of overfitting and eliminating the

998   unbalance issue in multi-class data sets, which is critical for achieving generalizable models and

999   testing results.

1000   Common data augmentation techniques adopted in medical image analysis applications[84, 107, 319]

1001   include cropping, translation, rotation, flipping, and scaling of images. Instead of augmenting

1002   whole images, Gao et al.[206] randomly jittered and cropped sub-images as patches from each

1003   original CT slice to generate more samples for classifying interstitial lung diseases. Pezeshk et

1004   al.[320] introduced an image blending tool that can seamlessly embed a lesion patch into a CT scan

1005   or mammography. Furthermore, the lesion patches could be inserted with various types of

1006   transformations to the lesion shape and characteristics. Improved classification performances

1007   were presented even for small training data sets. Zhang et al.[321] intended to tackle the unbalanced

1008   data issue for common medical image classification tasks. They proposed a new data

1009   augmentation method called unified learning of feature representation and similarity matrix. A

1010   single DCNN was trained on the seed labeled data set to obtain image feature representations and

1011   a similarity matrix simultaneously, which could be used for searching more similar images to

1012   each class of colonoscopy and upper endoscopy images.

1013   Another type of data augmentation involves synthesizing images or data using an object model

1014   and physics principles of image formation. Depending on the ultimate purpose of the DL

1015   algorithm, the degree of sophistication for the models and image formation approximations can

1016   vary.[322] Yang et al.[240] created a synthetic CT data set through the use of the Radon transform for

1017   a known object and modeled different exposure conditions through adding noise to the data, for

1018   the purpose of training a CNN to estimate high-dose projections from low-dose ones. Cui et al.[323]

1019   simulated dynamic PET emission data in order to train a stacked sparse autoencoder based

1020 reconstruction framework for dynamic PET imaging. Chen et al.[237] synthesized noisy projections

1021 based on patient images to generate training data for developing a DL algorithm for noise

1022 reduction in reconstructed CT images. Miao et al.[253] used synthetic data only to train a CNN for

1023 2D/3D image registration.

1024 **4.B. Data annotation via mining text reports**

1025 Over the decades, large amounts of radiological data (e.g., images, clinical annotations, and

1026 radiological reports) have accumulated in many hospitals' PACS. How to transform those

1027 retrospective radiological data into a machine-learnable format has become a big challenge in the

1028 DL era. A radiological report could contain many types of information. Generally speaking, it is

1029 a free-text summary of all the clinical findings and impressions determined during examination

1030 of a radiological image study. It can contain richer information than just the description of

1031 disease findings, but also may consist of negation and uncertainty statements. In the 'findings'

1032 section, a list of normal and abnormal observations is listed for each part of the body examined

1033 in the image. Attributes of the disease patterns, e.g., specific location and severity, are also noted.

1034 Furthermore, critical diagnosis information is often presented in the 'impression' section by

1035 considering all findings, patient history, and previous studies. Additional or follow-up imaging

1036 studies are recommended if suspicious findings are located. As such, reports consist of a

1037 challenging mixture of information. A key for machine learning is extracting the relevant parts

1038 for particular applications.[324]

1039 Schlegl et al.[325] relied on existing optical coherence tomography (OCT) volume data and

1040 corresponding diagnostic reports to correlate image content and geometry with semantic

1041 concepts described in the reports. Increasing classification accuracy for intraretinal cystoid fluid,

1042   subretinal fluid and normal retinal tissue was demonstrated while mining the voxel-level

1043   annotation of class labels.

1044   Following an initial work using MeSH (medical subject headings) manual annotations on chest

1045   radiographs,[326] Shin et al.[33] extracted sentences from the original radiology reports describing

1046   key images (images identified during clinical image interpretation as having important findings).

1047   The authors used natural language processing (NLP) to analyze about 780,000 patients'

1048   radiology reports and found 215,786 key images mentioned in the reports from scans of 61,845

1049   unique patients. The key images were then extracted from their institution's PACS.

1050   Corresponding image labels were then mined via unsupervised hierarchical Bayesian document

1051   clustering, i.e. generative latent Dirichlet allocation topic modeling, to form 80 classes at the first

1052   level of hierarchy. Zech et al.[316] applied a similar methodology to a set of 96,303 head computed

1053   tomography reports. While mining topic labels in a fully unsupervised manner,[33] they adopted

1054   latent Dirichlet allocation together with bag of words to compute the feature representation of

1055   corpuses. Then, a regression model was trained using a small subset (1,004) of annotated reports

1056   to initialize the clustering of those unlabeled text reports.

1057   The purely text-computed information offers some coarse level of radiology semantics but is

1058   often limited and disconnected from the associated image. First, the classes could be highly

1059   unbalanced, which means that one dominating category may contain many more images while

1060   other classes may contain few. Furthermore, the images in a class assigned purely by text

1061   analysis may not be visually coherent since the image appearance is not considered in the

1062   clustering process. Wang et al.[327] exploited a combination of image features and textual

1063   information extracted from reports to label groups of images to alleviate these limitations. Fig. 6

1064   shows the flowchart of the framework. A CNN based joint mining framework was developed to

1065   iteratively improve the extracted CNN image features and clustering labels. Consequently, NLP-

1066   mined disease keywords were assigned to each image cluster.

1067

1068   More advanced NLP techniques have demonstrated better performances in extracted disease

1069   keywords for image labeling task in recent studies. Wang et al.[180] introduced a two-stage

1070   pathology extraction approach by first detecting all disease keywords mentioned in the report

1071   using ontology-based tools and then building negation and uncertainty elimination rules on the

1072   dependency graph of sentences. Fig. 7 shows sample disease categories mined from the

1073   retrospective data. The authors publicly released their data set of 112,120 frontal-view chest x-

1074   ray images of 30,805 unique patients along with image annotations of 14 disease categories.

1075   Subsequent research led to a 6% average improvement in the area under the receiver operating

1076   characteristic curve through the use of a multi-level attention model in a DL pipeline that

1077   included both CNNs and recurrent neural networks.[328]

1078   Chen et al.[329] applied a CNN based textual classification framework to find the presence,

1079   chronicity, and location of pulmonary embolism in CT examination reports. A human-in-the-

1080   loop NLP annotation strategy was adopted to reduce the labeling cost for CNN training. The

1081   final CNN model was trained using a total of 2,512 radiologist-annotated CT reports.

1082

1083   Yan et al.[330, 331] mined radiology reports and images to extract lesion measurements. The lesion

1084   measurements were made in the course of routine clinical interpretation of CT scans. They were

1085   bidimensional measurements performed for RECIST (Response Evaluation Criteria in Solid

1086   Tumors) assessment, many as part of oncology clinical trials. Their data set, named

1087 "DeepLesion", consisted of 32,120 axial CT slices, each containing a measured lesion, from

1088 10,594 CT imaging studies of 4,459 unique patients. The data set consists of a large variety of

1089 lesion types, including those involving lung, liver, kidney, pancreas and lymph nodes. The

1090 authors' deep learning algorithm, which used a triple network and ImageNet pretrained weights,

1091 was able to retrieve images of specified type, location and size with an average accuracy of

1092 90.5%.

1093 Possibilities for text mining do not need to be limited to radiology reports but extend to other

1094 clinical reports. The presence of electronic health records (EHR) yields the potential to collect

1095 both imaging and clinical/pathology data in order to input to DL to predict diagnosis, outcome,

1096 and guide treatments within a clinical workflow.[332] Dai et al.[333] proposed a clinical report guided

1097 CNN which leverages a small amount of supervised information in clinical reports to identify the

1098 potential microaneurysms in fundus images. During training, both fundus images and clinical

1099 reports are presented to the network. In the testing stage, the input is a fundus image only, and

1100 the output is a probabilistic map of the lesion types in the image. Zhang et al.[334] proposed a

1101 multimodal network that jointly learns from medical images and their diagnostic reports, in

1102 which semantic information interacts with visual information to improve the image

1103 understanding ability by teaching the network to distill informative features. Applied to bladder

1104 cancer images and the corresponding diagnostic reports, the network demonstrated improved

1105 performance compared to baseline CNN that only use image information for training.

1106 **4.C. Data annotation via active learning**

1107 Another approach for assembling large data sets for DL is to try to increase the efficiency of

1108 collecting hand-labeled data to minimize the annotation cost. Active learning is one group of

1109 methods for increasing number of annotated data points by including human annotators in the

1110 loop of incremental learning and performance improvement. Two key aspects are usually

1111 considered for selecting candidate data for the expensive annotation process, uncertainty and

1112 representativeness of the candidate data.

1113 Different types of information could be utilized to measure the uncertainty and

1114 representativeness in order to select samples. Top et al.[335] computed the uncertainty values of

1115 radius bone regions in the image for segmentation by considering boundary, regional,

1116 smoothness and entropy energies of those image regions. Annotators were then required to label

1117 those regions in a CT plane with maximum uncertainty. Zhu et al.[336] leveraged the structured

1118 information (e.g., data from individual patients) when selecting batch of candidate unlabeled

1119 samples. The proposed learning framework enforced a set of specifically designed diversity

1120 constraints for the histopathological image annotation task. The visual saliency of objects[337]

1121 inside an image were considered as a measure for selecting samples. The similarities between

1122 labeled and unlabeled data were computed and encoded in a graph. Then, random walks were

1123 adopted for searching the most informative node (with largest classification uncertainty and

1124 minimum overlap with labeled data). Lee et al.[338] believe the most informative instances (hard

1125 examples) are those closest to the SVM hyperplane. Together with balanced sampling, their

1126 proposed learning framework was able to achieve a more than 40% classification performance

1127 increase on the testing set.

1128 A batch mode based active learning[339] method was proposed and applied to medical image

1129 classification applications. The Fisher information matrix was adopted to select informative

1130 unlabeled samples in a group-wise manner. The framework developed an efficient greedy

1131 searching algorithm to find a subset of the unlabeled data that can minimize the Fisher

1132 information of remaining unlabeled set. The experiments demonstrated the effectiveness of this

1133 batch-mode based active learning approach. Konyushkova et al.[340] trained a segmentation

1134 classifier to decide if a set of supervoxels were most in need to be annotated in 3D image

1135 volumes. Geometric priors were utilized in this process to compute geometric uncertainty for

1136 each voxel, indicating whether a clear boundary was present. For segmenting electron

1137 microscopy images, the model trained using 100 selected pixels with annotations (less than

1138 0.03% of the total training set) achieved even higher classification performance than the one

1139 trained with all available labeled training pixels.

1140 Recent approaches further utilized DCNN features to compute and representativeness criteria.

1141 Yang et al.[341] presented a deep fully convolutional network based active learning framework to

1142 reduce annotation effort in image that contain multiple instances, e.g., pathological images. The

1143 uncertainty and similarity information computed from network activations is utilized to select the

1144 most cost-effective annotation areas. Zhou et al.[342] measured the uncertainty and diversity of

1145 candidate image samples using the CNN classification prediction values computed for all the

1146 image patches extracted from the candidate image. In comparison to previous methods, this

1147 method has the advantage that no seed labeled sample is required. A newly-annotated sample

1148 will further improve the candidate selection process after CNN mode is fine-tuned again based

1149 on the new training set. They demonstrated that the CNN's classification performance could be

1150 incrementally enhanced by continuously fine-tuning the CNN in an iterative manner.

1151 There are other methods that do not require even a small number of initial hand-labeled data.

1152 Gaur et al.[343] started the selection process with a deep model trained on a similar domain. Then,

1153 they interpreted the active learning problem of increasing the size of limited labeled data set as

1154 an optimization problem by maximizing both the uncertainty and abundancy. Only a minimum

1155 number of data fulfilling both criterions were selected and annotated by a human expert.

1156   Mosinska et al.[344] tailored the uncertainty sampling based active learning approach for the

1157   delineation of complex linear structures problem, which significantly reduced the size (up to

1158   80%) of training data set while achieving equivalent performance. Multiple samples inside the

1159   same image were simultaneously presented to the annotator while the interactive annotation

1160   framework kept the selected samples informative, representative and diverse.

1161   **4.D. Expanding the training data set via domain adaptation**

1162   Instead of manually annotating selective number of data, another strategy for training data-

1163   hungry DL paradigms is to leverage labeled data from a different domain, e.g., ImageNet

1164   database of natural images, and then fine-tune based on the pre-trained CNN parameters in the

1165   target domain via transfer learning, as discussed in Sec. 3B. The assumption is that the essential

1166   pattern learned and recorded in CNN weights, especially in the earlier layers, to some extent are

1167   shared by different kinds of images from different domains. Under this assumption, transfer

1168   learning using a pre-trained model is rather straightforward, but the underlying differences of

1169   structures and features in data cross domains are overlooked. In contrast to this straightforward

1170   application of pre-training, domain adaptation attempts to alter a source domain to bring the

1171   distribution of the source closer to that of the target. In-depth analyses have been conducted to

1172   measure the distribution difference or nonlinear mapping of features between source and target

1173   domains for domain adaptation.

1174   Heimann et al.[345] employed a discriminative learning based approach to localize the

1175   transesophageal echocardiography transducer in X-ray images. Instance weighting was applied

1176   on unlabeled fluoroscopy image samples to estimate the differences in feature space density and

1177   correct covariate shift to align the data distribution cross domains. Wachinger et al.[346] employed

1178   a similar instance weighting strategy in a supervised domain adaptation problem with a small

1179   training set as supervision from the target domain. Conjeti et al.[347] computed tissue-specific

1180   back-scattering signal statistics for calcified, lipidic and fibrotic arterial plaques and used

1181   decision forest based method to align the distribution shift of signal statistics between in-vitro

1182   and in-vivo image domains.

1183   Schlegl et al.[205] trained a CNN in an unsupervised manner for learning more general low-level

1184   image features for images from multiple sites (as domains). Then, another CNN model was fine-

1185   tuned based on the previous CNN model (with domain information injected) to classify lung

1186   tissue in high-resolution CT data using a small set of annotated data from on site. Improved

1187   classification performance was demonstrated by adopting unsupervised pre-training with data

1188   cross domains.

1189   Different acquisition and staining processes can cause large variability of microscopic brain

1190   images even on the same part of brain.[348] Normalized Cross Correlation was introduced to locate

1191   image patches in the images from target domain, which shared the similar selected features with

1192   an image patch from the source domain. Those located image patches will also share the same

1193   label as their counterpart from the annotated source domain. Then, a multiple instance learning

1194   based classification framework was used to utilize those newly labeled (and also possibly noisy)

1195   patches for the image classification task. For the same problem, Becker et al.[349] proposed to learn

1196   a nonlinear mapping of the data features between two domains (acquisitions in this case),

1197   together with decision boundary for the regression based classification.

1198   Azizi et al.[16] applied an unsupervised domain adaptation method based on DL for the prostate

1199   cancer detection problem. A deep belief network was trained using both B-mode (target domain)

1200   and radiofrequency (source domain) ultrasound images to effectively align features from two

1201 domains in a common latent feature space. The alignment was achieved by minimizing the

1202 divergence between the source and target distributions through the training. Similar ideas were

1203 presented for multiple sclerosis lesion segmentation in MR images using fully convolutional

1204 networks.[350] A modified U-Net architecture was designed to take both labeled (source domain)

1205 and unlabeled (target domain) data and simultaneously minimize both the segmentation loss and

1206 the discrepancy between embedded features from two domains.

1207 **4.E. Data synthesis via generative adversarial networks**

1208 Generative adversarial networks have attracted tremendous attentions and have grown into a big

1209 family of methods in the past two years, from the original GAN framework[38] to recent

1210 CycleGAN.[37] The quality of synthesized images also evolved rapidly from 32*32 snapshots to

1211 high-resolution CT/MR images. There have been quite a few successful applications of GANs in

1212 the medical imaging domain. Compared to the conventional generative models based method,

1213 e.g., characteristic modeling,[351] random walk sampling,[352] and image decomposition,[353] GANs

1214 intend to produce better images from an image appearance perspective. However, these images

1215 are often less meaningful from a clinical point of view since the image intensity on each pixel in

1216 a real clinical image has semantic meanings, e.g., high values in PET image usually represent

1217 high take-up tumor regions. To overcome such limitations, a variety of constraints and additional

1218 information need to be included to help produce more clinically meaningful medical images.

1219 Calimeri et al.[354] cascaded the GAN models as a multi-scale pyramid based refinement

1220 framework with different size image inputs at each level so that a high-resolution MR image

1221 could be synthesized and then improved from coarse to fine. Frid-Adar et al.[215] started with

1222 standard data augmentation methods to create a larger data set that could be used to train a deep

1223 convolutional GAN. The synthetic data samples created for each lesion class, i.e. cysts,

1224 metastases and hemangiomas, by the GAN were then inputted to the training process of the final

1225 lesion classifier together with the enlarged training set from previous data augmentation. Lahiri

1226 et al.[355] extended the discriminator for classifying patches from multiple categories in addition to

1227 answering the fake or real binary question. This design has proven to be more data efficient for

1228 adversarial training. Zhang et al.[356] applied the same strategy on the semantic segmentation task,

1229 where the discriminator not only evaluated the segmentation results itself but also tried to

1230 differentiate the labeled and unlabeled data. The segmentation results from unlabeled data was

1231 weighted less (compared to the counterpart from labeled data) in the adversarial training

1232 procedure to produce more accurate results for the next iteration.

1233 Generating realistic images from scratch (initialized with noise vectors from the latent space) is

1234 extremely challenging, especially for medical images. However, more meaningful images could

1235 be synthesized if some prior knowledge was provided, e.g. an image similar to the target one but

1236 in different modality.[357] Costa et al.[358] proposed to generate retinal images by using

1237 corresponding vessel tree images. Different from the standard pair-wise GAN generative

1238 framework, an auto-encoder was first trained to learn the distribution of realistic retinal vessel

1239 trees and the retinal images were generated from the representations learned via the auto-

1240 encoder.

1241 Instead of using paired images for training, Chartsias et al.[359] adopted the CycleGAN framework

1242 in synthesizing cardiac MR images and masks from view-aligned CT ones in a loosely

1243 supervised manner. The pair-wise constraints (e.g. paired images with similar anatomical

1244 structure) were eliminated in this case. A 15% increase in segmentation accuracy was

1245 demonstrated by using both real and synthetic data compared to using real data alone. The

1246 application of CycleGAN in the unpaired MRI to CT image synthesis was also demonstrated.[360]

1247 Although it is still in its early stage, GAN based medical image generation has provided a

1248 promising alternative to other data augmentation approaches. Chuquicusma et al.[361] reported a

1249 visual Turing test that involved two radiologists (with different years of experience) to evaluate

1250 the quality of the synthesized nodules. A mixed set of (benign or malignant) nodule patches was

1251 shown to the radiologists individually for determining whether they were real or generated. The

1252 results showed that the majority (67% and 100%, respectively) of the generated nodules were

1253 recognized as real by the two radiologists.

1254

1255 **5. CHALLENGES, LESSONS LEARNED, AND THE FUTURE**

1256 As discussed in previous sections, recent advances in DL show that computers can extract more

1257 information from images, more reliably, and more accurately than ever before. However, further

1258 developing and optimizing DL techniques for the characteristics of medical images and medical

1259 data remains an important and relevant research challenge.

1260 **5.A. Evaluation and robustness**

1261 As discussed previously, data augmentation is often used to alleviate the problem of limited data

1262 set sizes. Data augmentation is powerful, but must be used correctly. One cannot train a network

1263 on a set of images pertaining to a given case and then test this trained network on a different set

1264 of images pertaining to that same case. Similarly, when dealing with 3D images, it might be

1265 tempting to treat every image slice as an independent entity. This would be incorrect, however,

1266 since slices of the same case are correlated and slices of a given case either need to be all in the

1267 training/validation set or all in the test set. If not done correctly, the performance will be

1268 substantially overestimated and not be generalizable. It is also important to keep in mind that

1269 performance needs to be evaluated 'by case', whether a 'case' is a lesion, patient, or whatever is

1270 relevant to the clinical task at hand. No matter how one slices and dices the data, if there are 100

1271 patients, there really are only 100 patients, and evaluation needs to be done accordingly.

1272

1273 When DL is used as a feature extractor, even in transfer learning when a completely trained deep

1274 net is applied to new images, the sheer number of extracted features poses a challenge. With the

1275 use of data augmentation, one would hope that the number of features will not exceed the

1276 number of data points so that dimension reduction or feature selection is possible in a meaningful

1277 way before further classification with a different classifier such as a shallow neural net or

1278 support vector machine. Feature selection, however, is likely to be a rather unstable undertaking

1279 with different features being selected depending on how the data set is partitioned. Additionally,

1280 it is common practice to use p-values to choose which of numerous features should be used, but

1281 p-values themselves are highly variable.[362, 363] P-values are data dependent statistics that vary

1282 from sample to sample even when underlying effects, population, and sampling are the same.[364]

1283 Hence, utmost care needs to be taken when using DL methods as feature extractors.

1284 Robustness and repeatability are concerns with any machine learning approach,[365] and even more

1285 so with DL. Since medical image data sets are so difficult to come by compared to those of

1286 natural images and generally are of limited size, researchers like to re-use the same data for

1287 different tasks. Hence, correction for multiple comparisons[366, 367] is crucial in the statistical

1288 evaluation of performance. The requirement that data sets need to be of sufficient size and

1289 quality is not unique to DL or medical imaging. It is, for example, reminiscent of issues observed

1290 in genomics where lack of reproducibility was observed when looking for predictive gene lists in

1291   small data sets (~100s of cases).[368, 369] There, thousands of samples are needed to generate a

1292   robust gene list to predict the outcome in cancer.[369] A 2012 study of 53 landmark papers in basic

1293   cancer research was able to replicate the original results of just 6 of these studies.[370] Moreover, a

1294   study reviewing radiomics using texture features, i.e., 'conventional' radiomics, for the

1295   prediction of survival, found that all of the results of 9 published studies failed to reach statistical

1296   significance after properly correcting p-values for multiple comparisons and the use of an

1297   optimal cut-off (if applicable) in Kaplan-Meier analysis.[371] Results of DL-based methods, if

1298   analysis is not performed correctly, may be even less likely to hold up to scrutiny.

1299   **5.B. Data sets and curation**

1300   Perhaps the most important challenge when it comes to medical imaging data sets is to obtain

1301   data of a sufficiently large number of properly annotated cases. The bottleneck is not necessarily

1302   obtaining the images, but obtaining annotations and reference standards. For segmentation tasks,

1303   for example, the reference standard or 'truth' would be the manual outline of one, or preferably

1304   more, expert radiologists. For cancer classification tasks, for example, the reference standard

1305   would be the pathological truth as determined by biopsy or surgery which needs to be extracted

1306   from pathology reports. The reference standard has to be of high quality, especially when used

1307   for training but also for performance evaluation. Obtaining high quality image data, annotations,

1308   and reference standards is expensive and time consuming. Patient privacy laws, while absolutely

1309   necessary, further complicate data collection because all protected health information needs to be

1310   removed from image data and corresponding radiology, pathology, and other reports. Moreover,

1311   relevant information needs to be extracted from the radiology, pathology, and other text reports

1312   which is time consuming and potentially error prone when done manually and not trivial when

1313   performed automatically (section 4.B). There is immense value in sharing annotated image data

1314 and anonymized publicly accessible databases such as provided by the Cancer Imaging Archive

1315 (www.cancerimagingarchive.net/).

1316 Another challenge for medical image data sets is that imaging devices are not measurement

1317 devices. Unlike a ruler or a Volt meter, which are calibrated and expected to give consistent and

1318 correct results within the calibration accuracy, imaging devices generate images through often

1319 proprietary image processing techniques. Images are usually not quantitative and primarily

1320 designed to be interpretable by humans, not by computers. Robustness of 'conventional' and DL-

1321 based methods with respect to image manufacturer or image pre-processing methods needs to be

1322 investigated. There has been effort investigating robustness of 'conventional' methods with

1323 respect to manufacturer for breast cancer diagnosis on ultrasound,[372, 373] the assessment of risk of

1324 future breast cancer on digital mammography,[374] and lung nodule features.[375] Work has also been

1325 done towards the harmonization of image data with respect to different CT scanners.[376] One of

1326 the advantages of DL-based methods, however, is that they may be less sensitive than

1327 'conventional' methods to differences in images due to the use of imaging equipment of different

1328 manufacturers. Having been designed for natural images in which, for example, a dog in the

1329 shade is still a dog, may make them better able to deal with differences in image appearance and

1330 quality.

1331 Class imbalance is another challenge related to many medical imaging data sets, not only to DL

1332 based methods but to 'conventional' methods as well. In screening mammography, for example,

1333 the cancer prevalence is so low that developing a method to detect cancer without causing undue

1334 false-positives is a formidable task. One approach to alleviate the problem of class imbalance in

1335 the training of DL methods is to use data augmentation of the under-represented class only in

1336 classifier training as explained in more detail in Section 4.

1337   **5.C. Interpretability**

1338 When a deep neural net is used as a feature extractor thousands of features are extracted. Unlike

1339 engineered hand-crafted features these features do not directly relate to something radiologists

1340 can easily interpret. Engineered features often describe something directly related to

1341 characteristics radiologists use in their clinical assessment, such as lesion size or shape. Such

1342 characteristics can be described by multiple mathematical descriptors, i.e., engineered features.

1343 For example, the 'simplest' feature of maximum linear dimension is both used by a radiologist

1344 and can be automatically calculated by a radiomics method. It is then easy for a radiologist to

1345 assess whether to trust the radiomics output. But even for 'traditional' approaches, this direct

1346 interpretability diminishes for more 'complicated' features such as for the many that describe

1347 texture. For features extracted from deep neural nets, this interpretability is almost completely

1348 lost. Radiologists may not care about all the DL parameters and how an application works,

1349 however, and it may be more a matter of human trust in the capabilities of the proverbial DL

1350 'black box'. The 'believability' of DL approaches – both as classifiers and as feature extractors –

1351 then, relies on past performance reported for large independent test sets. For example, in

1352 diagnosis of breast cancer, the believability of the probability of malignancy output by a DL

1353 method relies on knowledge of past performance on independent test data. Acceptance of DL in

1354 medical imaging may benefit from success of DL in other applications such as self-driving cars

1355 and robotics. On the other hand, there may be legal implications to using DL in medical imaging

1356 applications since it will be more difficult than for 'conventional' applications to pinpoint

1357 exactly what went wrong if the output is incorrect (potentially negatively impacting patient care).

1358 Recently, there has been increasing interest in making AI methods (including those involving

1359 DL) transparent, interpretable, and explainable.[377] This, in part, has been driven by the European

1360 general data protection regulation that will go into effect in May 2018 and will make 'black-box'

1361 approaches difficult to use in business. These new rules require it to be at least possible to trace

1362 results on demand.[377] Whereas traditional approaches tend to be at least interpretable in the sense

1363 that users can understand the underlying math of an algorithm, until recently, DL systems tended

1364 to be more opaque offering little or no insight into their inner workings. However, there has been

1365 increasing effort in making DL methods more transparent and methods have been proposed to

1366 assess the sensitivity of the prediction with respect to changes in the input or to decompose the

1367 decision in terms of the input variables.[378]

1368 It is possible to provide visual 'explanations', for example, to show heat maps visualizing the

1369 importance of each pixel for the prediction. These visualization techniques could help to further

1370 optimize a CNN training approach and ensure that the CNN is 'paying attention' to the correct

1371 regions of an image in analysis. For example, if a CNN were to be trained to detect

1372 pneumothorax on chest X-rays it would be important to know whether the CNN correctly

1373 'looked at' the pneumothorax region of images or instead focused on chest tubes that are often

1374 present in patients with pneumothorax. Most popular visualization techniques are either

1375 perturbation-based or backpropagation-based. Perturbation-based methods modify parts of the

1376 image and study the effect on the CNN output.[379, 380] Backpropagation-based methods propagate

1377 either the output probability score, or the gradient of the output with respect to the input in order

1378 to construct heatmaps. Some of the most popular backpropagation-based methods include the

1379 saliency map,[381] the class-activation map,[382] and the gradient-weighted class activation map.[383]

1380 Backpropagation-based methods are computationally cheaper because they use the fundamental

1381 property of propagating signals through convolutions, instead of propagating each modification

1382 through the network as in done in perturbation-based methods.

1383   **5.D. Competitive challenges**

1384   There have been a number of competitive challenges in the field of medical image analysis

1385   (https://grand-challenge.org/all_challenges/). The prevalence of DL based methods has clearly

1386   increased over the last couple of years and DL methods have become top performers in medical

1387   image analysis competitions. They often, but not always, perform as well as or better than

1388   'conventional' methods. In a literature review on DL, Litjens et al.[384] noted that the exact DL

1389   architecture does not seem to be the most important determinant in getting a good solution. For

1390   example, in the Kaggle Diabetic Retinopathy Challenge (https://www.kaggle.com/c/diabetic-

1391   retinopathy-detection), many researchers used the exact same architectures, the same type of

1392   networks, but obtained widely varying results. Data augmentation methods and preprocessing

1393   techniques seem to contribute substantially to good performance and robustness. It remains an

1394   open question how results from these competitive challenges can be leveraged to benefit the

1395   medical image analysis research community at large.

1396   **5.E. Lessons learned**

1397   Looking back into the history of medical image analysis, it appears that popularity of certain

1398   methods fluctuated in time. For example, ANNs gathered a lot of attention in the early 90's, were

1399   replaced by support vector machines in many applications in late 1990's and early 2000's, only

1400   to make a comeback in the form of DL in the 2010's. Likewise, the popularity of wavelet

1401   methods and feature extraction techniques such as SIFT evolved in time. The successes already

1402   achieved by DL methods, many of them discussed above, are undeniable and well-established.

1403   We believe that the application areas of DL will evolve in time like other methods, and will

1404   likely be supplemented and complemented by newer methods. However, one important lesson

1405   learned that will likely be maintained into the future is one about data quality, or the 'garbage-in

1406 garbage-out' principle. Quality of the image data and annotations is crucial and analysis needs to

1407 be carried out correctly. Another important lesson is the difference between statistical

1408 significance and clinical significance/relevance. Although establishing statistical significance is a

1409 very important step in research and publications, we should never lose sight of what the

1410 clinically relevant questions are, and just because there is a newer more complicated CNN, does

1411 not necessarily mean that it will better help (or replace) radiologists. Expert knowledge about the

1412 clinical task can provide advantages that go beyond adding more layers to a CNN, and

1413 incorporating expert medical knowledge to optimize methods, for example through novel data

1414 preprocessing or augmentation techniques, for a specific clinical task is often crucial in obtaining

1415 good performance.

1416 Plenty of challenges remain for 'conventional' medical image analysis and DL-based methods,

1417 including computational and statistical aspects. We need to investigate and improve image data

1418 harmonization, develop standards for reporting as well as experiments, and have better access to

1419 annotated image data such as publicly available data sets to serve as independent benchmarks.

1420 **5.F. Future of deep learning in imaging and therapy**

1421 Machine learning, including DL, is a fast-moving research field that has great promise for future

1422 applications in imaging and therapy. It is evident that DL has already pervaded almost every

1423 aspect of medical image analysis. 'Conventional' image analysis methods were never intended to

1424 replace radiologists but rather to serve as a second opinion. Likewise, DL-based methods are

1425 unlikely to replace human experts any time soon. The performance of DL has equaled or

1426 surpassed human performance for some non-medical tasks such as playing computer games[385]

1427 and, as illustrated by the many cited publications in this paper, DL has also been quite successful

1428 in a variety of medical imaging applications. However, most medical imaging tasks are far from

1429 solved[386] and the optimal deep learning method and architecture for each individual task and

1430 application area have not yet been established. Moreover, the integration of medical image

1431 analysis methods and other patient data - such as patient history, age, and demographics - also

1432 remains an area of active research that could further improve performance of clinical decision

1433 making aids.

1434 Three aspects that will drive the DL revolution are availability of big data, advances in DL

1435 algorithms, and processing power. As discussed above, there is abundant new research aimed at

1436 alleviating the limited data set size problem in medical imaging, and some of the custom DL

1437 architectures and algorithms specifically designed for medical imaging have shown great

1438 promise. There has been an explosion of research papers published on DL in medical imaging,

1439 most within the past couple of years, and this trend is expected continue. The emergence of

1440 conferences solely dedicated to DL in medical imaging (such as the 'Medical Imaging with Deep

1441 Learning Conference' to be held in July 2018, https://midl.amsterdam/) is very telling. The

1442 potential of DL in medical imaging has also not gone unnoticed by the healthcare industry.

1443 Companies, both big and small, are taking big steps in developing and commercializing new

1444 applications that are based on DL, and large medical imaging vendors have already made

1445 significant investments. Deep learning is here to stay, and its future in medical imaging and

1446 radiation therapy seems bright.

1447 **ACKNOWLEDGMENTS**

1451 appointment to the Research Participation Program at the Center for Devices and Radiological

1452 Health administered by the Oak Ridge Institute for Science and Education through an

1453 interagency agreement between the U.S. Department of Energy and the U.S. Food and Drug

1454 Administration. XW and RMS were supported by the Intramural Research Programs of the NIH

1455 Clinical Center. The mention of commercial products, their sources, or their use in connection

1456 with material reported herein is not to be construed as either an actual or implied endorsement of

1457 such products by the Department of Health and Human Services.

1458 **CONFLICTS OF INTEREST**

1459 MLG is a stockholder in R2/Hologic, scientific advisor, co-founder, and equity holder in

1460 Quantitative Insights, makers of QuantX, shareholder in Qview, and receives royalties from

1461 Hologic, GE Medical Systems, MEDIAN Technologies, Riverain Medical, Mitsubishi, and

1462 Toshiba. KD receives royalties from Hologic. RMS receives royalties from iCAD, Inc.,

1463 Koninklijke Philips NV, ScanMed, LLC, and receives research support from Ping An Insurance

1464 Company of China, Ltd., Carestream Health, Inc. and NVIDIA Corporation.

1465 **REFERENCES**

1466 1.   D. Amodei, S. Ananthanarayanan, R. Anubhai, J. Bai, E. Battenberg, C. Case, J. Casper,
1467   B. Catanzaro, Q. Cheng and G. Chen, "Deep speech 2: End-to-end speech recognition in
1468   english and mandarin," International Conference on Machine Learning, 173-182, 2016.
1469 2.   M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee and L. Zettlemoyer,
1470   "Deep contextualized word representations," arXiv preprint arXiv:1802.05365 (2018).
1471 3.   B. Zoph, V. Vasudevan, J. Shlens and Q. V. Le, "Learning transferable architectures for
1472   scalable image recognition," arXiv preprint arXiv:1707.07012 **2** (2017).
1473 4.   D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J.
1474   Schrittwieser, I. Antonoglou, V. Panneershelvam and M. Lanctot, "Mastering the game of
1475   Go with deep neural networks and tree search," Nature **529**, 484-489 (2016).
1476 5.   D. Gandhi, L. Pinto and A. Gupta, "Learning to fly by crashing," arXiv:1704.05588
1477   (2017).
1478 6.   "Worldwide Semiannual Cognitive and Artificial Intelligence Systems Spending Guide",
1479   (2016). At https://www.idc.com/getdoc.jsp?containerId=prUS41878616.

1480 7. "The Fourth Industrial Revolution: what it means, how to respond", (2016). At
1481 https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-
1482 means-and-how-to-respond/.
1483 8. E. Brynjolfsson and A. McAfee, The second machine age: Work, progress, and
1484 prosperity in a time of brilliant technologies. (WW Norton & Company, 2014).
1485 9. K. Schwab, The fourth industrial revolution. (Crown Business, 2017).
1486 10. "Harnessing automation for a future that works", (2017). At
1487 https://www.mckinsey.com/global-themes/digital-disruption/harnessing-automation-for-
1488 a-future-that-works.
1489 11. Y. LeCun, Y. Bengio and G. Hinton, "Deep learning," Nature **521**, 436-44 (2015).
1490 12. J. Schmidhuber, "Deep learning in neural networks: an overview," Neural Networks **61**,
1491 85-117 (2015).
1492 13. R. Caruana and A. Niculescu-Mizil, "An empirical comparison of supervised learning
1493 algorithms," in Proceedings of the 23rd international conference on Machine learning,
1494 (ACM, Pittsburgh, Pennsylvania, USA, 2006), pp. 161-168.
1495 14. S. J. Pan and Q. Yang, "A Survey on Transfer Learning," IEEE Transactions on
1496 Knowledge and Data Engineering **22**, 1345-1359 (2010).
1497 15. K. Weiss, T. M. Khoshgoftaar and D. Wang, "A survey of transfer learning," Journal of
1498 Big Data **3**, 9 (2016).
1499 16. S. Azizi, P. Mousavi, P. Yan, A. Tahmasebi, J. T. Kwak, S. Xu, B. Turkbey, P. Choyke,
1500 P. Pinto, B. Wood and P. Abolmaesumi, "Transfer learning from RF to B-mode temporal
1501 enhanced ultrasound features for prostate cancer detection," International Journal of
1502 Computer Assisted Radiology and Surgery **12**, 1111-1121 (2017).
1503 17. R. K. Samala, H. P. Chan, L. Hadjiiski, M. A. Helvie, J. Wei and K. Cha, "Mass
1504 detection in digital breast tomosynthesis: Deep convolutional neural network with
1505 transfer learning from mammography," Med Phys **43**, 6654 (2016).
1506 18. A. Kumar, J. Kim, D. Lyndon, M. Fulham and D. Feng, "An ensemble of fine-tuned
1507 convolutional neural networks for medical image classification," IEEE Journal of
1508 Biomedical and Health Informatics **21**, 31-40 (2017).
1509 19. Y. Bengio, "Learning Deep Architectures for AI," Foundations and Trends in Machine
1510 Learning **2**, 1-127 (2009).
1511 20. D. J. Felleman and D. C. Van Essen, "Distributed hierarchical processing in the primate
1512 cerebral cortex," Cereb Cortex **1**, 1-47 (1991).
1513 21. C. M. Wessinger, J. VanMeter, B. Tian, J. Van Lare, J. Pekar and J. P. Rauschecker,
1514 "Hierarchical organization of the human auditory cortex revealed by functional magnetic
1515 resonance imaging," Journal of Cognitive Neuroscience **13**, 1-7 (2001).
1516 22. R. Caruana, N. Karampatziakis and A. Yessenalina, "An empirical evaluation of
1517 supervised learning in high dimensions," in Proceedings of the 25th international
1518 conference on Machine learning, (ACM, Helsinki, Finland, 2008), pp. 96-103.
1519 23. G. E. Hinton, S. Osindero and Y. W. Teh, "A fast learning algorithm for deep belief
1520 nets," Neural Comput **18**, 1527-54 (2006).
1521 24. I. Sutskever, J. Martens, G. Dahl and G. Hinton, "On the importance of initialization and
1522 momentum in deep learning," in Proceedings of the 30th International Conference on
1523 Machine Learning, Vol. 28, edited by D. Sanjoy and M. David (PMLR, Proceedings of
1524 Machine Learning Research, 2013), pp. 1139--1147.

1525   25.   X. Glorot, A. Bordes and Y. Bengio, "Deep Sparse Rectifier Neural Networks," in
1526   Proceedings of the Fourteenth International Conference on Artificial Intelligence and
1527   Statistics, Vol. 15, edited by G. Geoffrey, D. David and D. Miroslav (PMLR,
1528   Proceedings of Machine Learning Research, 2011), pp. 315--323.
1529   26.   N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, "Dropout: a
1530   simple way to prevent neural networks from overfitting," J. Mach. Learn. Res. **15**, 1929-
1531   1958 (2014).
1532   27.   L. Wan, M. Zeiler, S. Zhang, Y. L. Cun and R. Fergus, "Regularization of Neural
1533   Networks using DropConnect," in Proceedings of the 30th International Conference on
1534   Machine Learning, Vol. 28, edited by D. Sanjoy and M. David (PMLR, Proceedings of
1535   Machine Learning Research, 2013), pp. 1058--1066.
1536   28.   J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei, "Imagenet: A large-scale
1537   hierarchical image database," Proceedings of the IEEE conference on computer vision
1538   and pattern recognition, 248-255, 2009.
1539   29.   A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet Classification with Deep
1540   Convolutional Neural Networks," in Advances in Neural Information Processing
1541   Systems, Vol. 25, edited by F. Pereira, C. J. C. Burges, L. Bottou and K. Q. Weinberger
1542   (Curran Associates, Inc., 2012), pp. 1097--1105.
1543   30.   Y. LeCun, K. Kavukcuoglu and C. Farabet, "Convolutional networks and applications in
1544   vision," IEEE International Symposium on Circuits and Systems, 253-256, 2010.
1545   31.   S. Hochreiter, #252 and r. Schmidhuber, "Long Short-Term Memory," Neural Comput. **9**,
1546   1735-1780 (1997).
1547   32.   O. Vinyals, A. Toshev, S. Bengio and D. Erhan, "Show and tell: A neural image caption
1548   generator," Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference
1549   on, 3156-3164, 2015.
1550   33.   H.-C. Shin, L. Lu, L. Kim, A. Seff, J. Yao and R. Summers, "Interleaved text/image deep
1551   mining on a large-scale radiology database for automated image interpretation," Journal
1552   of Machine Learning Research **17**, 2 (2016).
1553   34.   A. Graves, A.-r. Mohamed and G. Hinton, "Speech recognition with deep recurrent
1554   neural networks," Acoustics, speech and signal processing (icassp), 2013 ieee
1555   international conference on, 6645-6649, 2013.
1556   35.   I. Sutskever, O. Vinyals and Q. V. Le, "Sequence to sequence learning with neural
1557   networks," Advances in neural information processing systems, 3104-3112, 2014.
1558   36.   M. Arjovsky, S. Chintala and L. Bottou, "Wasserstein gan," arXiv preprint
1559   arXiv:1701.07875 (2017).
1560   37.   J.-Y. Zhu, T. Park, P. Isola and A. A. Efros, "Unpaired Image-to-Image Translation using
1561   Cycle-Consistent Adversarial Networks," IEEE International Conference on Computer
1562   Vision, 2017.
1563   38.   I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A.
1564   Courville and Y. Bengio, "Generative adversarial nets," Advances in neural information
1565   processing systems, 2672-2680, 2014.
1566   39.   M. Mardani, E. Gong, J. Y. Cheng, S. Vasanawala, G. Zaharchuk, M. Alley, N. Thakur,
1567   S. Han, W. Dally and J. M. Pauly, "Deep generative adversarial networks for compressed
1568   sensing automates MRI," arXiv:1706.00051 (2017).

1569   40.   P. H. Meyers, C. M. Nice, H. C. Becker, W. J. Nettleton, J. W. Sweeney and G. R.
1570   Meckstroth, "Automated computer analysis of radiographic images," Radiology **83**,
1571   1029-1034 (1964).
1572   41.   H. C. Becker, W. J. Nettleton, P. H. Meyers, J. W. Sweeney and C. M. Nice, "Digital
1573   Computer Determination of a Medical Diagnostic Index Directly from Chest X-Ray
1574   Images " IEEE Trans Biomed Eng **11**, 67-72 (1964).
1575   42.   G. S. Lodwick, T. E. Keats and J. P. Dorst, "The coding of roentgen images for computer
1576   analysis as applied to lung cancer," Radiology **81**, 185-200 (1963).
1577   43.   H.-P. Chan, K. Doi, S. Galhotra, C. J. Vyborny, H. MacMahon and P. M. Jokich, "Image
1578   feature analysis and computer-aided diagnosis in digital radiography. 1. Automated
1579   detection of microcalcifications in mammography," Med Phys **14**, 538-548 (1987).
1580   44.   M. L. Giger, K. Doi and H. MacMahon, "Image feature analysis and computer aided
1581   diagnosis in digital radiography. 3. Automated detection of nodules in peripheral lung
1582   fields.," Med Phys **15**, 158-66 (1988).
1583   45.   K. Kanazawa, Y. Kawata, N. Niki, H. Satoh, H. Ohmatsu, R. Kakinuma, M. Kaneko, N.
1584   Moriyama and K. Eguchi, "Computer-aided diagnosis for pulmonary nodules based on
1585   helical CT images," Computerized Medical Imaging and Graphics **22**, 157-167 (1998).
1586   46.   C. Abe, C. E. Kahn, K. Doi and S. Katsuragawa, "Computer-Aided Detection Of Diffuse
1587   Liver-Disease In Ultrasound Images," Investigative Radiology **27**, 71-77 (1992).
1588   47.   K. Fukushima, S. Miyake and T. Ito, "Neocognitron: A neural network model for a
1589   mechanism of visual pattern recognition," IEEE Transactions on Systems, Man, and
1590   Cybernetics **SMC-13**, 826-834 (1983).
1591   48.   J. S. Lin, P. A. Ligomenides, M. T. Freedman and S. K. Mun, "Application of artificial
1592   neural networks for reduction of false-positive detections in digital chest radiographs,"
1593   Proc Annu Symp Comput Appl Med Care, 434-8 (1993).
1594   49.   S. C. Lo, S. L. Lou, J. S. Lin, M. T. Freedman and S. K. Mun, "Artificial convolution
1595   neural network techniques and applications to lung nodule detection," IEEE Transactions
1596   on Medical Imaging **14**, 711-718 (1995).
1597   50.   H.-P. Chan, S. C. B. Lo, B. Sahiner, K. L. Lam and M. A. Helvie, "Computer-aided
1598   detection of mammographic microcalcifications: Pattern recognition with an artificial
1599   neural network," Med Phys **22**, 1555-1567 (1995).
1600   51.   B. Sahiner, C. Heang-Ping, N. Petrick, W. Datong, M. A. Helvie, D. D. Adler and M. M.
1601   Goodsitt, "Classification of mass and normal breast tissue: a convolution neural network
1602   classifier with spatial domain and texture images," IEEE Transactions on Medical
1603   Imaging **15**, 598-610 (1996).
1604   52.   W. Zhang, K. Doi, M. L. Giger, Y. Wu, R. M. Nishikawa and R. A. Schmidt,
1605   "Computerized detection of clustered microcalcifications in digital mammograms using a
1606   shift-invariant artificial neural network," Med Phys **21**, 517-524 (1994).
1607   53.   W. Zhang, K. Doi, M. L. Giger, R. M. Nishikawa and R. A. Schmidt, "An improved
1608   shift-invariant artificial neural network for computerized detection of clustered
1609   microcalcifications in digital mammograms," Med Phys **23**, 595-601 (1996).
1610   54.   K. Suzuki, S. G. Armato, F. Li, S. Sone and K. Doi, "Massive training artificial neural
1611   network (MTANN) for reduction of false positives in computerized detection of lung
1612   nodules in low-dose computed tomography," Med. Phys **30**, 1602-1617 (2003).
1613   55.   K. Suzuki, F. Li, S. Sone and K. Doi, "Computer-Aided Diagnostic Scheme for
1614   Distinction Between Benign and Malignant Nodules in Thoracic Low-Dose CT by Use of

1615   Massive Training Artificial Neural Network," IEEE Transactions on Medical Imaging **24**,
1616   1138-1150 (2005).
1617   56.   K. M. He, X. Y. Zhang, S. Q. Ren and J. Sun, "Deep Residual Learning for Image
1618   Recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition,
1619   (2016), pp. 770-778.
1620   57.   M. Drozdzal, E. Vorontsov, G. Chartrand, S. Kadoury and C. Pal, "The Importance of
1621   Skip Connections in Biomedical Image Segmentation," in Deep Learning and Data
1622   Labeling for Medical Applications, Vol. 10008, edited by G. Carneiro, D. Mateus, L.
1623   Peter, A. Bradley, J. Tavares, V. Belagiannis, J. P. Papa, J. C. Nascimento, M. Loog, Z.
1624   Lu, J. S. Cardoso and J. Cornebise (2016), pp. 179-187.
1625   58.   H. Greenspan, B. van Ginneken and R. M. Summers, "Guest editorial deep learning in
1626   medical imaging: Overview and future promise of an exciting new technique," IEEE
1627   Transactions on Medical Imaging **35**, 1153-1159 (2016).
1628   59.   K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale
1629   image recognition," arXiv:1409.1556 (2014).
1630   60.   G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger, "Densely Connected
1631   Convolutional Networks," CVPR, 3, 2017.
1632   61.   F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," 2017
1633   IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1800-1807,
1634   2017.
1635   62.   C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the inception
1636   architecture for computer vision," Proceedings of the IEEE Conference on Computer
1637   Vision and Pattern Recognition, 2818-2826, 2016.
1638   63.   V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S.
1639   Venugopalan, K. Widner, T. Madams, J. Cuadros, R. Kim, R. Raman, P. C. Nelson, J. L.
1640   Mega and R. Webster, "Development and Validation of a Deep Learning Algorithm for
1641   Detection of Diabetic Retinopathy in Retinal Fundus Photographs," Jama-Journal of the
1642   American Medical Association **316**, 2402-2410 (2016).
1643   64.   T. Kooi, G. Litjens, B. van Ginneken, A. Gubern-Merida, C. I. Sancheza, R. Mann, A.
1644   den Heeten and N. Karssemeijer, "Large scale deep learning for computer aided detection
1645   of mammographic lesions," Medical Image Analysis **35**, 303-312 (2017).
1646   65.   Y. Liu, K. Gadepalli, M. Norouzi, G. E. Dahl, T. Kohlberger, A. Boyko, S. Venugopalan,
1647   A. Timofeev, P. Q. Nelson and G. S. Corrado, "Detecting cancer metastases on gigapixel
1648   pathology images," arXiv:1703.02442 (2017).
1649   66.   D. Ribli, A. Horváth, Z. Unger, P. Pollner and I. Csabai, "Detecting and classifying
1650   lesions in mammograms with Deep Learning," Sci Rep **8**, 4165 (2018).
1651   67.   N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway and J.
1652   M. Liang, "Convolutional Neural Networks for Medical Image Analysis: Full Training or
1653   Fine Tuning?," IEEE Transactions on Medical Imaging **35**, 1299-1312 (2016).
1654   68.   A. Coates, B. Huval, T. Wang, D. Wu, B. Catanzaro and N. Andrew, "Deep learning with
1655   COTS HPC systems," International Conference on Machine Learning, 1337-1345, 2013.
1656   69.   "Amazon Web Services". At https://aws.amazon.com/.
1657   70.   "NVidia GPU Cloud". At https://www.nvidia.com/en-us/gpu-cloud/.
1658   71.   "Google Cloud TPU". At https://cloud.google.com/tpu/.

1659 72. M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G.
1660   Irving and M. Isard, "TensorFlow: A System for Large-Scale Machine Learning," OSDI,
1661   265-283, 2016.
1662 73. Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama and
1663   T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," Proceedings
1664   of the 22nd ACM international conference on Multimedia, 675-678, 2014.
1665 74. R. Collobert, K. Kavukcuoglu and C. Farabet, "Torch7: A matlab-like environment for
1666   machine learning," BigLearn, NIPS workshop, 2011.
1667 75. F. Bastien, P. Lamblin, R. Pascanu, J. Bergstra, I. Goodfellow, A. Bergeron, N.
1668   Bouchard, D. Warde-Farley and Y. Bengio, "Theano: new features and speed
1669   improvements," arXiv:1211.5590 (2012).
1670 76. K. H. Cha, L. Hadjiiski, R. K. Samala, H. P. Chan, E. M. Caoili and R. H. Cohan,
1671   "Urinary bladder segmentation in CT urography using deep-learning convolutional neural
1672   network and level sets," Med Phys **43**, 1882 (2016).
1673 77. K. H. Cha, L. M. Hadjiiski, R. K. Samala, H. P. Chan, R. H. Cohan, E. M. Caoili, C.
1674   Paramagul, A. Alva and A. Z. Weizer, "Bladder Cancer Segmentation in CT for
1675   Treatment Response Assessment: Application of Deep-Learning Convolution Neural
1676   Network-A Pilot Study," Tomography **2**, 421-429 (2016).
1677 78. M. R. Avendi, A. Kheradvar and H. Jafarkhani, "Automatic segmentation of the right
1678   ventricle from cardiac MRI using a learning-based approach," Magn Reson Med **78**,
1679   2439-2448 (2017).
1680 79. T. A. Ngo, Z. Lu and G. Carneiro, "Combining deep learning and level set for the
1681   automated segmentation of the left ventricle of the heart from cardiac cine magnetic
1682   resonance," Med Image Anal **35**, 159-171 (2017).
1683 80. F. Lu, F. Wu, P. Hu, Z. Peng and D. Kong, "Automatic 3D liver location and
1684   segmentation via convolutional neural network and graph cut," International Journal of
1685   Computer Assisted Radiology and Surgery **12**, 171-182 (2017).
1686 81. F. Liu, Z. Zhou, H. Jang, A. Samsonov, G. Zhao and R. Kijowski, "Deep convolutional
1687   neural network and 3D deformable approach for tissue segmentation in musculoskeletal
1688   magnetic resonance imaging," Magn Reson Med **79**, 2379-2391 (2018).
1689 82. M. R. Avendi, A. Kheradvar and H. Jafarkhani, "A combined deep-learning and
1690   deformable-model approach to fully automatic segmentation of the left ventricle in
1691   cardiac MRI," Med Image Anal **30**, 108-119 (2016).
1692 83. B. Ibragimov and L. Xing, "Segmentation of organs-at-risks in head and neck CT images
1693   using convolutional neural networks," Med Phys **44**, 547-557 (2017).
1694 84. O. Ronneberger, P. Fischer and T. Brox, "U-Net: Convolutional Networks for
1695   Biomedical Image Segmentation," International Conference on Medical Image
1696   Computing and Computer-Assisted Intervention (MICCAI), 234-241, Cham, 2015.
1697 85. M. U. Dalmis, G. Litjens, K. Holland, A. Setio, R. Mann, N. Karssemeijer and A.
1698   Gubern-Merida, "Using deep learning to segment breast and fibroglandular tissue in MRI
1699   volumes," Med Phys **44**, 533-546 (2017).
1700 86. D. Nie, L. Wang, R. Trullo, J. Li, P. Yuan, J. Xia and D. Shen, "Segmentation of
1701   Craniomaxillofacial Bony Structures from MRI with a 3D Deep-Learning Based Cascade
1702   Framework," Mach Learn Med Imaging **10541**, 266-273 (2017).
1703 87. R. Cheng, H. R. Roth, N. Lay, L. Lu, B. Turkbey, W. Gandler, E. S. McCreedy, T.
1704   Pohida, P. A. Pinto, P. Choyke, M. J. McAuliffe and R. M. Summers, "Automatic

1705   magnetic resonance prostate segmentation by deep learning with holistically nested
1706   networks," J Med Imaging (Bellingham) **4**, 041302 (2017).

1707 88.   Y. Zhuge, A. V. Krauze, H. Ning, J. Y. Cheng, B. C. Arora, K. Camphausen and R. W.
1708   Miller, "Brain tumor segmentation using holistically nested neural networks in MRI
1709   images," Med Phys **44**, 5234-5243 (2017).

1710 89.   H. Lee, F. M. Troschel, S. Tajmir, G. Fuchs, J. Mario, F. J. Fintelmann and S. Do, "Pixel-
1711   Level Deep Segmentation: Artificial Intelligence Quantifies Muscle on Computed
1712   Tomography for Body Morphometric Analysis," J Digit Imaging **30**, 487-498 (2017).

1713 90.   Y. Wang, Y. Qiu, T. Thai, K. Moore, H. Liu and B. Zheng, "A two-step convolutional
1714   neural network based computer-aided detection scheme for automatically segmenting
1715   adipose tissue volume depicting on CT images," Comput Methods Programs Biomed
1716   **144**, 97-104 (2017).

1717 91.   P. Hu, F. Wu, J. Peng, Y. Bao, F. Chen and D. Kong, "Automatic abdominal multi-organ
1718   segmentation using deep convolutional neural network and time-implicit level sets,"
1719   International Journal of Computer Assisted Radiology and Surgery **12**, 399-411 (2017).

1720 92.   A. Mansoor, J. J. Cerrolaza, R. Idrees, E. Biggs, M. A. Alsharid, R. A. Avery and M. G.
1721   Linguraru, "Deep Learning Guided Partitioned Shape Model for Anterior Visual Pathway
1722   Segmentation," IEEE Transactions on Medical Imaging **35**, 1856-65 (2016).

1723 93.   H. Choi and K. H. Jin, "Fast and robust segmentation of the striatum using deep
1724   convolutional neural networks," J Neurosci Methods **274**, 146-153 (2016).

1725 94.   P. Moeskops, M. A. Viergever, A. M. Mendrik, L. S. de Vries, M. J. Benders and I.
1726   Isgum, "Automatic Segmentation of MR Brain Images With a Convolutional Neural
1727   Network," IEEE Transactions on Medical Imaging **35**, 1252-1261 (2016).

1728 95.   C. Wachinger, M. Reuter and T. Klein, "DeepNAT: Deep convolutional neural network
1729   for segmenting neuroanatomy," Neuroimage **170**, 434-445 (2018).

1730 96.   H. Chen, Q. Dou, L. Yu, J. Qin and P. A. Heng, "VoxResNet: Deep voxelwise residual
1731   networks for brain segmentation from 3D MR images," Neuroimage **170**, 446-455
1732   (2018).

1733 97.   J. Dolz, C. Desrosiers and I. Ben Ayed, "3D fully convolutional networks for subcortical
1734   segmentation in MRI: A large-scale study," Neuroimage **170**, 456-470 (2018).

1735 98.   M. Kallenberg, K. Petersen, M. Nielsen, A. Y. Ng, P. Diao, C. Igel, C. M. Vachon, K.
1736   Holland, R. R. Winkel, N. Karssemeijer and M. Lillholm, "Unsupervised Deep Learning
1737   Applied to Breast Density Segmentation and Mammographic Risk Scoring," IEEE
1738   Transactions on Medical Imaging **35**, 1322-1331 (2016).

1739 99.   L. K. Tan, R. A. McLaughlin, E. Lim, Y. F. Abdul Aziz and Y. M. Liew, "Fully
1740   automated segmentation of the left ventricle in cine cardiac MRI using neural network
1741   regression," J Magn Reson Imaging (2018).

1742 100.   L. K. Tan, Y. M. Liew, E. Lim and R. A. McLaughlin, "Convolutional neural network
1743   regression for short-axis left ventricle segmentation in cardiac cine MR sequences," Med
1744   Image Anal **39**, 78-86 (2017).

1745 101.   L. Yu, Y. Guo, Y. Wang, J. Yu and P. Chen, "Segmentation of Fetal Left Ventricle in
1746   Echocardiographic Sequences Based on Dynamic Convolutional Neural Networks,"
1747   IEEE Trans Biomed Eng **64**, 1886-1895 (2017).

1748 102.   T. L. Kline, P. Korfiatis, M. E. Edwards, J. D. Blais, F. S. Czerwiec, P. C. Harris, B. F.
1749   King, V. E. Torres and B. J. Erickson, "Performance of an Artificial Multi-observer Deep

1750		Neural Network for Fully Automated Segmentation of Polycystic Kidneys," J Digit
1751		Imaging **30**, 442-448 (2017).
1752	103.	K. Sharma, C. Rupprecht, A. Caroli, M. C. Aparicio, A. Remuzzi, M. Baust and N.
1753		Navab, "Automatic Segmentation of Kidneys using Deep Learning for Total Kidney
1754		Volume Quantification in Autosomal Dominant Polycystic Kidney Disease," Sci Rep **7**,
1755		2049 (2017).
1756	104.	P. Hu, F. Wu, J. Peng, P. Liang and D. Kong, "Automatic 3D liver segmentation based on
1757		deep learning and globally optimized surface evolution," Phys Med Biol **61**, 8676-8698
1758		(2016).
1759	105.	W. Kovacs, N. Hsieh, H. Roth, C. Nnamdi-Emeratom, W. P. Bandettini, A. Arai, A.
1760		Mankodi, R. M. Summers and J. Yao, "Holistic segmentation of the lung in cine MRI," J
1761		Med Imaging (Bellingham) **4**, 041310 (2017).
1762	106.	A. Farag, L. Lu, H. R. Roth, J. Liu, E. Turkbey and R. M. Summers, "A Bottom-up
1763		Approach for Pancreas Segmentation using Cascaded Superpixels and (Deep) Image
1764		Patch Labeling," IEEE Transactions on Image Processing (2016).
1765	107.	H. R. Roth, L. Lu, A. Farag, H.-C. Shin, J. Liu, E. B. Turkbey and R. M. Summers,
1766		"Deeporgan: Multi-level deep convolutional networks for automated pancreas
1767		segmentation," International Conference on Medical Image Computing and Computer-
1768		Assisted Intervention (MICCAI), 556-564, 2015.
1769	108.	Y. Guo, Y. Gao and D. Shen, "Deformable MR Prostate Segmentation via Deep Feature
1770		Learning and Sparse Patch Matching," IEEE Transactions on Medical Imaging **35**, 1077-
1771		89 (2016).
1772	109.	S. Liao, Y. Gao, A. Oto and D. Shen, "Representation learning: a unified deep learning
1773		framework for automatic prostate MR segmentation," International Conference on
1774		Medical Image Computing and Computer-Assisted Intervention (MICCAI), 254-61, 2013.
1775	110.	Z. Tian, L. Liu, Z. Zhang and B. Fei, "PSNet: prostate segmentation on MRI based on a
1776		convolutional neural network," J Med Imaging (Bellingham) **5**, 021208 (2018).
1777	111.	K. Men, J. Dai and Y. Li, "Automatic segmentation of the clinical target volume and
1778		organs at risk in the planning CT for rectal cancer using deep dilated convolutional neural
1779		networks," Med Phys **44**, 6377 (2017).
1780	112.	X. Li, Q. Dou, H. Chen, C. W. Fu, X. Qi, D. L. Belavy, G. Armbrecht, D. Felsenberg, G.
1781		Zheng and P. A. Heng, "3D multi-scale FCN with random modality voxel dropout
1782		learning for Intervertebral Disc Localization and Segmentation from Multi-modality MR
1783		Images," Med Image Anal **45**, 41-54 (2018).
1784	113.	X. Zhou, R. Takayama, S. Wang, T. Hara and H. Fujita, "Deep learning of the sectional
1785		appearances of 3D CT images for anatomical structure segmentation based on an FCN
1786		voting method," Med Phys **44**, 5221-5233 (2017).
1787	114.	Q. Dou, L. Yu, H. Chen, Y. Jin, X. Yang, J. Qin and P. A. Heng, "3D deeply supervised
1788		network for automated segmentation of volumetric medical images," Med Image Anal
1789		**41**, 40-54 (2017).
1790	115.	V. Alex, K. Vaidhya, S. Thirunavukkarasu, C. Kesavadas and G. Krishnamurthi,
1791		"Semisupervised learning using denoising autoencoders for brain lesion detection and
1792		segmentation," J Med Imaging (Bellingham) **4**, 041311 (2017).
1793	116.	P. Korfiatis, T. L. Kline and B. J. Erickson, "Automated Segmentation of Hyperintense
1794		Regions in FLAIR MRI Using Deep Learning," Tomography **2**, 334-340 (2016).

1795 117. S. Iqbal, M. U. Ghani, T. Saba and A. Rehman, "Brain tumor segmentation in multi-
1796   spectral MRI using convolutional neural networks (CNN)," Microsc Res Tech **81**, 419-
1797   427 (2018).
1798 118. M. A. Al-antari, M. A. Al-masni, M.-T. Choi, S.-M. Han and T.-S. Kim, "A fully
1799   integrated computer-aided diagnosis system for digital X-ray mammograms via deep
1800   learning detection, segmentation, and classification," International Journal of Medical
1801   Informatics **117**, 44-54 (2018).
1802 119. R. Zhang, L. Huang, W. Xia, B. Zhang, B. Qiu and X. Gao, "Multiple supervised residual
1803   network for osteosarcoma segmentation in CT images," Comput Med Imaging Graph **63**,
1804   1-8 (2018).
1805 120. L. Huang, W. Xia, B. Zhang, B. Qiu and X. Gao, "MSFCN-multiple supervised fully
1806   convolutional networks for the osteosarcoma segmentation of CT images," Comput
1807   Methods Programs Biomed **143**, 67-74 (2017).
1808 121. K. Kamnitsas, C. Ledig, V. F. J. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D.
1809   Rueckert and B. Glocker, "Efficient multi-scale 3D CNN with fully connected CRF for
1810   accurate brain lesion segmentation," Med Image Anal **36**, 61-78 (2017).
1811 122. Y. Liu, S. Stojadinovic, B. Hrycushko, Z. Wardak, S. Lau, W. G. Lu, Y. L. Yan, S. B.
1812   Jiang, X. Zhen, R. Timmerman, L. Nedzi and X. J. Gu, "A deep convolutional neural
1813   network-based automatic delineation strategy for multiple brain metastases stereotactic
1814   radiosurgery," Plos One **12** (2017).
1815 123. M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P. M.
1816   Jodoin and H. Larochelle, "Brain tumor segmentation with Deep Neural Networks," Med
1817   Image Anal **35**, 18-31 (2017).
1818 124. X. Zhao, Y. Wu, G. Song, Z. Li, Y. Zhang and Y. Fan, "A deep learning model
1819   integrating FCNNs and CRFs for brain tumor segmentation," Med Image Anal **43**, 98-
1820   111 (2018).
1821 125. L. Chen, P. Bentley and D. Rueckert, "Fully automatic acute ischemic lesion
1822   segmentation in DWI using convolutional neural networks," Neuroimage Clin **15**, 633-
1823   643 (2017).
1824 126. T. Brosch, L. Y. Tang, Y. Youngjin, D. K. Li, A. Traboulsee and R. Tam, "Deep 3D
1825   Convolutional Encoder Networks With Shortcuts for Multiscale Feature Integration
1826   Applied to Multiple Sclerosis Lesion Segmentation," IEEE Transactions on Medical
1827   Imaging **35**, 1229-1239 (2016).
1828 127. M. Ghafoorian, N. Karssemeijer, T. Heskes, I. W. M. van Uden, C. I. Sanchez, G.
1829   Litjens, F. E. de Leeuw, B. van Ginneken, E. Marchiori and B. Platel, "Location Sensitive
1830   Deep Convolutional Neural Networks for Segmentation of White Matter
1831   Hyperintensities," Sci Rep **7**, 5110 (2017).
1832 128. K. Men, X. Chen, Y. Zhang, T. Zhang, J. Dai, J. Yi and Y. Li, "Deep Deconvolutional
1833   Neural Network for Target Segmentation of Nasopharyngeal Cancer in Planning
1834   Computed Tomography Images," Front Oncol **7**, 315 (2017).
1835 129. J. Ma, F. Wu, T. Jiang, Q. Zhao and D. Kong, "Ultrasound image-based thyroid nodule
1836   automatic segmentation using convolutional neural networks," International Journal of
1837   Computer Assisted Radiology and Surgery **12**, 1895-1910 (2017).
1838 130. W. Li, F. Jia and Q. Hu, "Automatic segmentation of liver tumor in CT images with deep
1839   convolutional neural networks," Journal of Computer and Communications **3**, 146
1840   (2015).

131. S. Wang, M. Zhou, Z. Liu, Z. Liu, D. Gu, Y. Zang, D. Dong, O. Gevaert and J. Tian, "Central focused convolutional neural networks: Developing a data-driven model for lung nodule segmentation," Med Image Anal **40**, 172-183 (2017).

132. I. Nogues, L. Lu, X. Wang, H. Roth, G. Bertasius, N. Lay, J. Shi, Y. Tsehay and R. M. Summers, "Automatic lymph node cluster segmentation using holistically-nested neural networks and structured optimization in CT images," International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 388-397, 2016.

133. S. Trebeschi, J. J. M. van Griethuysen, D. M. J. Lambregts, M. J. Lahaye, C. Parmer, F. C. H. Bakers, N. Peters, R. G. H. Beets-Tan and H. Aerts, "Deep Learning for Fully-Automated Localization and Segmentation of Rectal Cancer on Multiparametric MR," Sci Rep **7**, 9 (2017).

134. M. H. Jafari, E. Nasr-Esfahani, N. Karimi, S. M. R. Soroushmehr, S. Samavi and K. Najarian, "Extraction of skin lesions from non-dermoscopic images for surgical excision of melanoma," International Journal of Computer Assisted Radiology and Surgery **12**, 1021-1030 (2017).

135. D. Yang, S. Zhang, Z. Yan, C. Tan, K. Li and D. Metaxas, "Automated anatomical landmark detection on distal femur surface using convolutional neural network," 2015 IEEE 12th International Symposium on Biomedical Imaging, 17-21, 2015.

136. H. Chen, D. Ni, J. Qin, S. L. Li, X. Yang, T. F. Wang and P. A. Heng, "Standard Plane Localization in Fetal Ultrasound via Domain Transferred Deep Neural Networks," IEEE Journal of Biomedical and Health Informatics **19**, 1627-1636 (2015).

137. C. F. Baumgartner, K. Kamnitsas, J. Matthew, S. Smith, B. Kainz and D. Rueckert, "Real-Time Standard Scan Plane Detection and Localisation in Fetal Ultrasound Using Fully Convolutional Neural Networks," International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 203-211, 2016.

138. A. Kumar, P. Sridar, A. Quinton, R. K. Kumar, D. Feng, R. Nanan and J. Kim, "Plane identification in fetal ultrasound images using saliency maps and convolutional neural networks," 2016 IEEE 13th International Symposium on Biomedical Imaging, 791-794, 2016.

139. F. C. Ghesu, E. Krubasik, B. Georgescu, V. Singh, Z. Yefeng, J. Hornegger and D. Comaniciu, "Marginal Space Deep Learning: Efficient Architecture for Volumetric Image Parsing," IEEE Transactions on Medical Imaging **35**, 1217-1228 (2016).

140. H. Wu, C. Bailey, P. Rasoulinejad and S. Li, "Automatic Landmark Estimation for Adolescent Idiopathic Scoliosis Assessment Using BoostNet," International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 127-135, 2017.

141. K. Yan, L. Lu and R. M. Summers, "Unsupervised body part regression using convolutional neural network with self-organization," arXiv:1707.03891 (2017).

142. Z. Yan, Y. Zhan, Z. Peng, S. Liao, Y. Shinagawa, S. Zhang, D. N. Metaxas and X. S. Zhou, "Multi-Instance Deep Learning: Discover Discriminative Local Anatomies for Bodypart Recognition," IEEE Transactions on Medical Imaging **35**, 1332-1343 (2016).

143. F. C. Ghesu, B. Georgescu, T. Mansi, D. Neumann, J. Hornegger and D. Comaniciu, "An Artificial Agent for Anatomical Landmark Detection in Medical Images," International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 229-237, 2016.

1887 144. F. C. Ghesu, B. Georgescu, Y. Zheng, S. Grbic, A. Maier, J. Hornegger and D.
1888   Comaniciu, "Multi-Scale Deep Reinforcement Learning for Real-Time 3D-Landmark
1889   Detection in CT Scans," IEEE Transactions on Pattern Analysis and Machine Intelligence
1890   (in print) (2017).
1891 145. F. C. Ghesu, B. Georgescu, S. Grbic, A. K. Maier, J. Hornegger and D. Comaniciu,
1892   "Robust Multi-scale Anatomical Landmark Detection in Incomplete 3D-CT Data,"
1893   International Conference on Medical Image Computing and Computer-Assisted
1894   Intervention (MICCAI), 194-202, 2017.
1895 146. Z. Xu, Q. Huang, J. Park, M. Chen, D. Xu, D. Yang, D. Liu and S. K. Zhou, "Supervised
1896   Action Classifier: Approaching Landmark Detection as Image Partitioning,"
1897   International Conference on Medical Image Computing and Computer-Assisted
1898   Intervention (MICCAI), 338-346, 2017.
1899 147. C. Payer, D. Štern, H. Bischof and M. Urschler, "Regressing Heatmaps for Multiple
1900   Landmark Localization Using CNNs," International Conference on Medical Image
1901   Computing and Computer-Assisted Intervention (MICCAI), 230-238, 2016.
1902 148. Y. Cai, M. Landis, D. T. Laidley, A. Kornecki, A. Lum and S. Li, "Multi-modal vertebrae
1903   recognition using Transformed Deep Convolution Network," Computerized Medical
1904   Imaging and Graphics **51**, 11-19 (2016).
1905 149. N. Baka, S. Leenstra and T. v. Walsum, "Ultrasound Aided Vertebral Level Localization
1906   for Lumbar Surgery," IEEE Transactions on Medical Imaging **36**, 2138-2147 (2017).
1907 150. Y. Zheng, D. Liu, B. Georgescu, H. Nguyen and D. Comaniciu, "3D Deep Learning for
1908   Efficient and Robust Landmark Detection in Volumetric Data," International Conference
1909   on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 565-572,
1910   2015.
1911 151. H. Chen, Q. Dou, D. Ni, J.-Z. Cheng, J. Qin, S. Li and P.-A. Heng, "Automatic Fetal
1912   Ultrasound Standard Plane Detection Using Knowledge Transferred Recurrent Neural
1913   Networks," International Conference on Medical Image Computing and Computer-
1914   Assisted Intervention (MICCAI), 507-514, 2015.
1915 152. X. Lu, D. Xu and D. Liu, "Robust 3D Organ Localization with Dual Learning
1916   Architectures and Fusion," Deep Learning and Data Labeling for Medical Applications,
1917   12-20, 2016.
1918 153. H. R. Roth, C. T. Lee, H.-C. Shin, A. Seff, L. Kim, J. Yao, L. Lu and R. M. Summers,
1919   "Anatomy-specific classification of medical images using deep convolutional nets,"
1920   Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on, 101-104, 2015.
1921 154. B. D. de Vos, J. M. Wolterink, P. A. de Jong, T. Leiner, M. A. Viergever and I. Išgum,
1922   "ConvNet-Based Localization of Anatomical Structures in 3-D Medical Images," IEEE
1923   transactions on medical imaging **36**, 1470-1481 (2017).
1924 155. J. Zhang, M. Liu and D. Shen, "Detecting Anatomical Landmarks From Limited Medical
1925   Imaging Data Using Two-Stage Task-Oriented Deep Neural Networks," IEEE
1926   Transactions on Image Processing **26**, 4753-4764 (2017).
1927 156. A. P. Harrison, Z. Xu, K. George, L. Lu, R. M. Summers and D. J. Mollura, "Progressive
1928   and multi-path holistically nested neural networks for pathological lung segmentation
1929   from CT images," International Conference on Medical Image Computing and
1930   Computer-Assisted Intervention, 621-629, 2017.

1931 157. J. Yao, W. Kovacs, N. Hsieh, C.-Y. Liu and R. M. Summers, "Holistic segmentation of
1932   intermuscular adipose tissues on thigh MRI," International Conference on Medical Image
1933   Computing and Computer-Assisted Intervention, 737-745, 2017.
1934 158. M. L. Giger, H. P. Chan and J. Boone, "Anniversary Paper: History and status of CAD
1935   and quantitative image analysis: The role of Medical Physics and AAPM," Med Phys **35**,
1936   5799-5820 (2008).
1937 159. M. L. Giger, N. Karssemeijer and J. A. Schnabel, "Breast Image Analysis for Risk
1938   Assessment, Detection, Diagnosis, and Treatment of Cancer," in Annual Review of
1939   Biomedical Engineering, Vol 15, Vol. 15, edited by M. L. Yarmush (2013), pp. 327-357.
1940 160. H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura and R. M.
1941   Summers, "Deep convolutional neural networks for computer-aided detection: CNN
1942   architectures, dataset characteristics and transfer learning," IEEE transactions on medical
1943   imaging **35**, 1285-1298 (2016).
1944 161. M. H. Yap, G. Pons, J. Martí, S. Ganau, M. Sentís, R. Zwiggelaar, A. K. Davison and R.
1945   Martí, "Automated breast ultrasound lesions detection using convolutional neural
1946   networks," IEEE Journal of biomedical and health informatics (2017).
1947 162. P. M. Cheng, T. K. Tejura, K. N. Tran and G. Whang, "Detection of high-grade small
1948   bowel obstruction on conventional radiography with convolutional neural networks,"
1949   Abdominal Radiology, 1-8 (2017).
1950 163. S. Belharbi, C. Chatelain, R. Hérault, S. Adam, S. Thureau, M. Chastan and R.
1951   Modzelewski, "Spotting L3 slice in CT scans using deep convolutional network and
1952   transfer learning," Computers in biology and medicine **87**, 95-103 (2017).
1953 164. S. Albarqouni, C. Baur, F. Achilles, V. Belagiannis, S. Demirci and N. Navab, "Aggnet:
1954   deep learning from crowds for mitosis detection in breast cancer histology images," IEEE
1955   Transactions on Medical Imaging **35**, 1313-1321 (2016).
1956 165. J. I. Orlando, E. Prokofyeva, M. del Fresno and M. B. Blaschko, "An ensemble deep
1957   learning based approach for red lesion detection in fundus images," Computer methods
1958   and programs in biomedicine **153**, 115-127 (2018).
1959 166. H. R. Roth, L. Lu, J. Liu, J. Yao, A. Seff, K. Cherry, L. Kim and R. M. Summers,
1960   "Improving computer-aided detection using convolutional neural networks and random
1961   view aggregation," IEEE transactions on medical imaging **35**, 1170-1181 (2016).
1962 167. X. Yang, C. Liu, Z. Wang, J. Yang, H. Le Min, L. Wang and K.-T. T. Cheng, "Co-trained
1963   convolutional neural networks for automated detection of prostate cancer in multi-
1964   parametric MRI," Medical image analysis **42**, 212-227 (2017).
1965 168. A. A. Setio, F. Ciompi, G. Litjens, P. Gerke, C. Jacobs, S. J. van Riel, M. M. Wille, M.
1966   Naqibullah, C. I. Sanchez and B. van Ginneken, "Pulmonary Nodule Detection in CT
1967   Images: False Positive Reduction Using Multi-View Convolutional Networks," IEEE
1968   Transactions on Medical Imaging **35**, 1160-1169 (2016).
1969 169. Q. Dou, H. Chen, L. Yu, J. Qin and P. A. Heng, "Multilevel Contextual 3-D CNNs for
1970   False Positive Reduction in Pulmonary Nodule Detection," IEEE Trans Biomed Eng **64**,
1971   1558-1567 (2017).
1972 170. M. Ghafoorian, N. Karssemeijer, T. Heskes, M. Bergkamp, J. Wissink, J. Obels, K.
1973   Keizer, F. E. de Leeuw, B. V. Ginneken, E. Marchiori and B. Platel, "Deep multi-scale
1974   location-aware 3D convolutional neural networks for automated detection of lacunes of
1975   presumed vascular origin," Neuroimage Clin **14**, 391-399 (2017).

171. J. Long, E. Shelhamer and T. Darrell, "Fully convolutional networks for semantic segmentation," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 3431-3440, 2015.

172. D. Qi, C. Hao, Y. Lequan, Z. Lei, Q. Jing, W. Defeng, V. C. Mok, S. Lin and H. Pheng-Ann, "Automatic Detection of Cerebral Microbleeds From MR Images via 3D Convolutional Neural Networks," IEEE Transactions on Medical Imaging **35**, 1182-1195 (2016).

173. 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," Medical image analysis **26**, 195-202 (2015).

174. S. Chen, J. Qin, X. Ji, B. Lei, T. Wang, D. Ni and J.-Z. Cheng, "Automatic scoring of multiple semantic attributes with multi-task feature leverage: A study on pulmonary nodules in CT images," IEEE transactions on medical imaging **36**, 802-814 (2017).

175. A. Teramoto, H. Fujita, O. Yamamuro and T. Tamaki, "Automated detection of pulmonary nodules in PET/CT images: Ensemble false‐positive reduction using a convolutional neural network technique," Med Phys **43**, 2821-2827 (2016).

176. H. Jiang, H. Ma, W. Qian, M. Gao and Y. Li, "An automatic detection system of lung nodule based on multi-group patch-based deep learning network," IEEE Journal of Biomedical and Health Informatics **In print** (2017).

177. Y. Bar, I. Diamant, L. Wolf, S. Lieberman, E. Konen and H. Greenspan, "Chest pathology detection using deep learning with non-medical training," Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on, 294-297, 2015.

178. M. Cicero, A. Bilbily, E. Colak, T. Dowdell, B. Gray, K. Perampaladas and J. Barfett, "Training and validating a deep convolutional neural network for computer-aided detection and classification of abnormalities on frontal chest radiographs," Investigative radiology **52**, 281-287 (2017).

179. P. Lakhani and B. Sundaram, "Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks," Radiology **284**, 574-582 (2017).

180. 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," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 3462-3471, 2017.

181. Y. Bar, I. Diamant, L. Wolf and H. Greenspan, "Deep learning with non-medical training used for chest pathology identification," in Proc. SPIE Medical Imaging, Vol. 9414, edited by L. M. Hadjiiski and G. D. Tourassi (2015), pp. 94140V.

182. T. Nakao, S. Hanaoka, Y. Nomura, I. Sato, M. Nemoto, S. Miki, E. Maeda, T. Yoshikawa, N. Hayashi and O. Abe, "Deep neural network‐based computer‐assisted detection of cerebral aneurysms in MR angiography," Journal of Magnetic Resonance Imaging **47**, 948-953 (2018).

183. M. U. Dalmış, S. Vreemann, T. Kooi, R. M. Mann, N. Karssemeijer and A. Gubern-Mérida, "Fully automated detection of breast cancer in screening MRI using convolutional neural networks," Journal of Medical Imaging **5**, 014502 (2018).

2021 184. J. Liu, D. Wang, L. Lu, Z. Wei, L. Kim, E. B. Turkbey, B. Sahiner, N. Petrick and R. M.
2022 Summers, "Detection and diagnosis of colitis on computed tomography using deep
2023 convolutional neural networks," Med Phys **44**, 4630-4642 (2017).
2024 185. J. J. Nappi, P. Pickhardt, D. H. Kim, T. Hironaka and H. Yoshida, "Deep learning of
2025 contrast-coated serrated polyps for computer-aided detection in CT colonography," in
2026 Medical Imaging 2017: Computer-Aided Diagnosis, Vol. 10134, edited by S. G. Armato
2027 and N. A. Petrick (2017).
2028 186. H. R. Roth, L. Lu, A. Seff, K. M. Cherry, J. Hoffman, S. J. Wang, J. M. Liu, E. Turkbey
2029 and R. M. Summers, "A New 2.5D Representation for Lymph Node Detection Using
2030 Random Sets of Deep Convolutional Neural Network Observations," International
2031 Conference on Medical Image Computing and Computer-Assisted Intervention
2032 (MICCAI), 520-527, 2014.
2033 187. R. Vivanti, A. Szeskin, N. Lev-Cohain, J. Sosna and L. Joskowicz, "Automatic detection
2034 of new tumors and tumor burden evaluation in longitudinal liver CT scan studies,"
2035 International journal of computer assisted radiology and surgery **12**, 1945-1957 (2017).
2036 188. J. Ma, F. Wu, T. a. Jiang, J. Zhu and D. Kong, "Cascade convolutional neural networks
2037 for automatic detection of thyroid nodules in ultrasound images," Med Phys **44**, 1678-
2038 1691 (2017).
2039 189. Y. Tsehay, N. Lay, X. Wang, J. T. Kwak, B. Turkbey, P. Choyke, P. Pinto, B. Wood and
2040 R. M. Summers, "Biopsy-guided learning with deep convolutional neural networks for
2041 Prostate Cancer detection on multiparametric MRI," Biomedical Imaging (ISBI 2017),
2042 2017 IEEE 14th International Symposium on, 642-645, 2017.
2043 190. J. Liu, K. Chellamuthu, L. Lu, M. Bagheri and R. M. Summers, "A coarse-to-fine
2044 approach for pericardial effusion localization and segmentation in chest CT scans," in
2045 Proc. SPIE Medical Imaging, Vol. 10575, edited by N. Petrick and K. Mori (2018), pp.
2046 105753B.
2047 191. K. Chellamuthu, J. Liu, J. Yao, M. Bagheri, L. Lu, V. Sandfort and R. M. Summers,
2048 "Atherosclerotic vascular calcification detection and segmentation on low dose computed
2049 tomography scans using convolutional neural networks," Biomedical Imaging (ISBI
2050 2017), 2017 IEEE 14th International Symposium on, 388-391, 2017.
2051 192. H. Li, M. L. Giger, B. Q. Huynh and N. O. Antropova, "Deep learning in breast cancer
2052 risk assessment: evaluation of convolutional neural networks on a clinical dataset of full-
2053 field digital mammograms," J Med Imaging (Bellingham) **4**, 041304 (2017).
2054 193. A. A. Mohamed, W. A. Berg, H. Peng, Y. Luo, R. C. Jankowitz and S. Wu, "A deep
2055 learning method for classifying mammographic breast density categories," Med Phys **45**,
2056 314-321 (2018).
2057 194. S. F. Li, J. Wei, H. P. Chan, M. A. Helvie, M. A. Roubidoux, Y. Lu, C. Zhou, L. M.
2058 Hadjiiski and R. K. Samala, "Computer-aided assessment of breast density: comparison
2059 of supervised deep learning and feature-based statistical learning," Phys Med Biol **63**
2060 (2018).
2061 195. J. Lee and R. M. Nishikawa, "Automated mammographic breast density estimation using
2062 a fully convolutional network," Med Phys **45**, 1178-1190 (2018).
2063 196. B. Q. Huynh, H. Li and M. L. Giger, "Digital mammographic tumor classification using
2064 transfer learning from deep convolutional neural networks," J Med Imaging (Bellingham)
2065 **3**, 034501 (2016).

2066 197. N. Antropova, B. Q. Huynh and M. L. Giger, "A deep feature fusion methodology for
2067 breast cancer diagnosis demonstrated on three imaging modality datasets," Med Phys **44**,
2068 5162-5171 (2017).
2069 198. R. K. Samala, H. P. Chan, L. M. Hadjiiski, M. A. Helvie, C. Richter and K. Cha,
2070 "Evolutionary pruning of transfer learned deep convolutional neural network for breast
2071 cancer diagnosis in digital breast tomosynthesis," Phys Med Biol (2018).
2072 199. R. K. Samala, H. P. Chan, L. M. Hadjiiski, M. A. Helvie, K. H. Cha and C. D. Richter,
2073 "Multi-task transfer learning deep convolutional neural network: application to computer-
2074 aided diagnosis of breast cancer on mammograms," Phys Med Biol **62**, 8894-8908
2075 (2017).
2076 200. N. Antropova, H. Abe and M. L. Giger, "Use of clinical MRI maximum intensity
2077 projections for improved breast lesion classification with deep convolutional neural
2078 networks," J Med Imaging (Bellingham) **5**, 014503 (2018).
2079 201. N. Antropova, B. Huynh and M. L. Giger, "Long short-term memory networks for
2080 efficient breast DCE-MRI classification," in NIPS: Neural Information Processing
2081 Systems, Medical Imaging Meets NIPS, (2017).
2082 202. T. Kooi, B. van Ginneken, N. Karssemeijer and A. den Heeten, "Discriminating solitary
2083 cysts from soft tissue lesions in mammography using a pretrained deep convolutional
2084 neural network," Med Phys **44**, 1017-1027 (2017).
2085 203. B. B. Shi, L. J. Grimm, M. A. Mazurowski, J. A. Baker, J. R. Marks, L. M. King, C. C.
2086 Maley, E. S. Hwang and J. Y. Lo, "Prediction of Occult Invasive Disease in Ductal
2087 Carcinoma in Situ Using Deep Learning Features," Journal of the American College of
2088 Radiology **15**, 527-534 (2018).
2089 204. A. Nibali, Z. He and D. Wollersheim, "Pulmonary nodule classification with deep
2090 residual networks," International Journal of Computer Assisted Radiology and Surgery
2091 **12**, 1799-1808 (2017).
2092 205. T. Schlegl, J. Ofner and G. Langs, "Unsupervised Pre-training Across Image Domains
2093 Improves Lung Tissue Classification," Medical Computer Vision: Algorithms for Big
2094 Data, 82-93, 2014.
2095 206. M. Gao, U. Bagci, L. Lu, A. Wu, M. Buty, H.-C. Shin, H. Roth, G. Z. Papadakis, A.
2096 Depeursinge and R. M. Summers, "Holistic classification of CT attenuation patterns for
2097 interstitial lung diseases via deep convolutional neural networks," Computer Methods in
2098 Biomechanics and Biomedical Engineering: Imaging & Visualization **6**, 1-6 (2018).
2099 207. M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe and S. Mougiakakou, "Lung
2100 Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural
2101 Network," IEEE Transactions on Medical Imaging **35**, 1207-1216 (2016).
2102 208. G. B. Kim, K. H. Jung, Y. Lee, H. J. Kim, N. Kim, S. Jun, J. B. Seo and D. A. Lynch,
2103 "Comparison of Shallow and Deep Learning Methods on Classifying the Regional
2104 Pattern of Diffuse Lung Disease," J Digit Imaging (2017).
2105 209. S. Christodoulidis, M. Anthimopoulos, L. Ebner, A. Christe and S. Mougiakakou,
2106 "Multisource Transfer Learning With Convolutional Neural Networks for Lung Pattern
2107 Analysis," IEEE J Biomed Health Inform **21**, 76-84 (2017).
2108 210. A. Masood, B. Sheng, P. Li, X. H. Hou, X. E. Wei, J. Qin and D. G. Feng, "Computer-
2109 Assisted Decision Support System in Pulmonary Cancer detection and stage classification
2110 on CT images," Journal of Biomedical Informatics **79**, 117-128 (2018).

211. G. Gonzalez, S. Y. Ash, G. Vegas-Sanchez-Ferrero, J. O. Onieva, F. N. Rahaghi, J. C. Ross, A. Diaz, R. S. J. Estepar, G. R. Washko, Copdgene and E. Investigator, "Disease Staging and Prognosis in Smokers Using Deep Learning in Chest Computed Tomography," American Journal of Respiratory and Critical Care Medicine **197**, 193-203 (2018).

212. N. Lessmann, B. van Ginneken, M. Zreik, P. A. de Jong, B. D. de Vos, M. A. Viergever and I. Isgum, "Automatic Calcium Scoring in Low-Dose Chest CT Using Deep Neural Networks With Dilated Convolutions," IEEE Transactions on Medical Imaging **37**, 615-625 (2018).

213. W. F. Xue, G. Brahm, S. Pandey, S. Leung and S. Li, "Full left ventricle quantification via deep multitask relationships learning," Medical Image Analysis **43**, 54-65 (2018).

214. P. M. Cheng and H. S. Malhi, "Transfer Learning with Convolutional Neural Networks for Classification of Abdominal Ultrasound Images," Journal of Digital Imaging **30**, 234-243 (2017).

215. M. Frid-Adar, E. Klang, M. Amitai, J. Goldberger and H. Greenspan, "Synthetic Data Augmentation using GAN for Improved Liver Lesion Classification," IEEE International Symposium on Biomedical Imaging, 2018.

216. K. Yasaka, H. Akai, O. Abe and S. Kiryu, "Deep Learning with Convolutional Neural Network for Differentiation of Liver Masses at Dynamic Contrast-enhanced CT: A Preliminary Study," Radiology **286**, 899-908 (2018).

217. R. Bharath and P. Rajalakshmi, "Deep Scattering Convolution Network based Features for Ultrasonic Fatty Liver Tissue Characterization," in 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, (2017), pp. 1982-1985.

218. J. Lao, Y. Chen, Z. C. Li, Q. Li, J. Zhang, J. Liu and G. Zhai, "A Deep Learning-Based Radiomics Model for Prediction of Survival in Glioblastoma Multiforme," Sci Rep **7**, 10353 (2017).

219. D. B. Larson, M. C. Chen, M. P. Lungren, S. S. Halabi, N. V. Stence and C. P. Langlotz, "Performance of a Deep-Learning Neural Network Model in Assessing Skeletal Maturity on Pediatric Hand Radiographs," Radiology **287**, 313-322 (2018).

220. H. Li, Y. Zhu, E. S. Burnside, E. Huang, K. Drukker, K. A. Hoadley, C. Fan, S. D. Conzen, M. Zuley, J. M. Net, E. Sutton, G. J. Whitman, E. Morris, C. M. Perou, Y. Ji and M. L. Giger, "Quantitative MRI radiomics in the prediction of molecular classifications of breast cancer subtypes in the TCGA/TCIA data set," NPJ Breast Cancer **2** (2016).

221. E. S. Burnside, K. Drukker, H. Li, E. Bonaccio, M. Zuley, M. Ganott, J. M. Net, E. J. Sutton, K. R. Brandt, G. J. Whitman, S. D. Conzen, L. Lan, Y. Ji, Y. T. Zhu, C. C. Jaffe, E. P. Huang, J. B. Freymann, J. S. Kirby, E. A. Morris and M. L. Giger, "Using computer-extracted image phenotypes from tumors on breast magnetic resonance imaging to predict breast cancer pathologic stage," Cancer **122**, 748-757 (2016).

222. Y. T. Zhu, H. Li, W. T. Guo, K. Drukker, L. Lan, M. L. Giger and Y. Ji, "Deciphering Genomic Underpinnings of Quantitative MRI-based Radiomic Phenotypes of Invasive Breast Carcinoma," Sci Rep **5** (2015).

223. H. Li, Y. T. Zhu, E. S. Burnside, K. Drukker, K. A. Hoadley, C. Fan, S. D. Conzen, G. J. Whitman, E. J. Sutton, J. M. Net, M. Ganott, E. Huang, E. A. Morris, C. M. Perou, Y. Ji and M. L. Giger, "MR Imaging Radiomics Signatures for Predicting the Risk of Breast

2156   Cancer Recurrence as Given by Research Versions of MammaPrint, Oncotype DX, and
2157   PAM50 Gene Assays," Radiology **281**, 382-391 (2016).

2158 224.   W. Guo, H. Li, Y. Zhu, L. Lan, S. Yang, K. Drukker, E. Morris, E. Burnside, G.
2159   Whitman, M. L. Giger and Y. Ji, "Prediction of clinical phenotypes in invasive breast
2160   carcinomas from the integration of radiomics and genomics data," J Med Imaging
2161   (Bellingham) **2**, 041007 (2015).

2162 225.   W. Q. Sun, T. L. Tseng, J. Y. Zhang and W. Qian, "Enhancing deep convolutional neural
2163   network scheme for breast cancer diagnosis with unlabeled data," Computerized Medical
2164   Imaging and Graphics **57**, 4-9 (2017).

2165 226.   A. R. Jamieson, M. L. Giger, K. Drukker and L. L. Pesce, "Enhancement of breast CADx
2166   with unlabeled data," Med Phys **37**, 4155-4172 (2010).

2167 227.   S. Katsuragawa, K. Doi, H. MacMahon, L. MonnierCholley, T. Ishida and T. Kobayashi,
2168   "Classification of normal and abnormal lungs with interstitial diseases by rule-based
2169   method and artificial neural networks," Journal of Digital Imaging **10**, 108-114 (1997).

2170 228.   B. van Ginneken, S. Katsuragawa, B. M. T. Romeny, K. Doi and M. A. Viergever,
2171   "Automatic detection of abnormalities in chest radiographs using local texture analysis,"
2172   IEEE Transactions on Medical Imaging **21**, 139-149 (2002).

2173 229.   H. Li, M. L. Giger, L. Lan, J. Bancroft Brown, A. MacMahon, M. Mussman, O. I.
2174   Olopade and C. Sennett, "Computerized analysis of mammographic parenchymal patterns
2175   on a large clinical dataset of full-field digital mammograms: robustness study with two
2176   high-risk datasets," J Digit Imaging **25**, 591-8 (2012).

2177 230.   H. He, X. Yang, L. Wu, H. Yan, Z. Gao, Y. Feng and G. Townsend, "Dual Long Short-
2178   Term Memory Networks for Sub-Character Representation Learning," arXiv:1712.08841
2179   (2018).

2180 231.   S. S. Garapati, L. Hadjiiski, K. H. Cha, H. P. Chan, E. M. Caoili, R. H. Cohan, A.
2181   Weizer, A. Alva, C. Paramagul, J. Wei and C. A. Zhou, "Urinary bladder cancer staging
2182   in CT urography using machine learning," Med Phys **44**, 5814-5823 (2017).

2183 232.   M. Henglin, G. Stein, P. V. Hushcha, J. Snoek, A. B. Wiltschko and S. Cheng, "Machine
2184   Learning Approaches in Cardiovascular Imaging," Circulation-Cardiovascular Imaging
2185   **10** (2017).

2186 233.   L. Zhang, L. Lu, R. M. Summers, E. Kebebew and J. Yao, "Convolutional Invasion and
2187   Expansion Networks for Tumor Growth Prediction," IEEE transactions on medical
2188   imaging **37**, 638-648 (2018).

2189 234.   K. Suzuki, "A supervised 'lesion-enhancement' filter by use of a massive-training
2190   artificial neural network (MTANN) in computer-aided diagnosis (CAD)," Phys Med Biol
2191   **54**, S31-45 (2009).

2192 235.   W. Yang, Y. Y. Chen, Y. B. Liu, L. M. Zhong, G. G. Qin, Z. T. Lu, Q. J. Feng and W. F.
2193   Chen, "Cascade of multi-scale convolutional neural networks for bone suppression of
2194   chest radiographs in gradient domain," Medical Image Analysis **35**, 421-433 (2017).

2195 236.   S. Mori, "Deep architecture neural network-based real-time image processing for image-
2196   guided radiotherapy," Physica Medica-European Journal of Medical Physics **40**, 79-87
2197   (2017).

2198 237.   H. Chen, Y. Zhang, W. Zhang, P. Liao, K. Li, J. Zhou and G. Wang, "Low-dose CT via
2199   convolutional neural network," Biomed Opt Express **8**, 679-694 (2017).

238. H. Chen, Y. Zhang, M. K. Kalra, F. Lin, Y. Chen, P. Liao, J. Zhou and G. Wang, "Low-Dose CT With a Residual Encoder-Decoder Convolutional Neural Network," IEEE Transactions on Medical Imaging **36**, 2524-2535 (2017).

239. E. Kang, J. Min and J. C. Ye, "A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction," Med Phys **44**, e360-e375 (2017).

240. X. Yang, V. De Andrade, W. Scullin, E. L. Dyer, N. Kasthuri, F. De Carlo and D. Gursoy, "Low-dose x-ray tomography through a deep convolutional neural network," Sci Rep **8**, 2575 (2018).

241. L. Xiang, Y. Qiao, D. Nie, L. An, Q. Wang and D. Shen, "Deep Auto-context Convolutional Neural Networks for Standard-Dose PET Image Estimation from Low-Dose PET/MRI," Neurocomputing **267**, 406-416 (2017).

242. K. H. Jin, M. T. McCann, E. Froustey and M. Unser, "Deep Convolutional Neural Network for Inverse Problems in Imaging," IEEE Transactions on Image Processing **26**, 4509-4522 (2017).

243. Y. B. Zhang, Y. Chu and H. Y. Yu, "Reduction of metal artifacts in x-ray CT images using a convolutional neural network," in Developments in X-Ray Tomography Xi, Vol. 10391, edited by B. Muller and G. Wang (2017).

244. Y. Han, J. Yoo, H. H. Kim, H. J. Shin, K. Sung and J. C. Ye, "Deep learning with domain adaptation for accelerated projection-reconstruction MR," Magn Reson Med (2018).

245. V. Golkov, A. Dosovitskiy, J. I. Sper, M. I. Menzel, M. Czisch, P. Samann, T. Brox and D. Cremers, "q-Space Deep Learning: Twelve-Fold Shorter and Model-Free Diffusion MRI Scans," IEEE Transactions on Medical Imaging **35**, 1344-1351 (2016).

246. K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock and F. Knoll, "Learning a variational network for reconstruction of accelerated MRI data," Magn Reson Med **79**, 3055-3071 (2018).

247. J. Schlemper, J. Caballero, J. V. Hajnal, A. N. Price and D. Rueckert, "A Deep Cascade of Convolutional Neural Networks for Dynamic MR Image Reconstruction," IEEE Transactions on Medical Imaging **37**, 491-503 (2018).

248. B. Zhu, J. Z. Liu, S. F. Cauley, B. R. Rosen and M. S. Rosen, "Image reconstruction by domain-transform manifold learning," Nature **555**, 487-492 (2018).

249. G. Wu, M. Kim, Q. Wang, Y. Gao, S. Liao and D. Shen, "Unsupervised deep feature learning for deformable registration of MR brain images," International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 649-56, 2013.

250. G. Wu, M. Kim, Q. Wang, B. C. Munsell and D. Shen, "Scalable High-Performance Image Registration Framework by Unsupervised Deep Feature Representations Learning," IEEE Trans Biomed Eng **63**, 1505-16 (2016).

251. X. Yang, R. Kwitt, M. Styner and M. Niethammer, "Quicksilver: Fast predictive image registration - A deep learning approach," Neuroimage **158**, 378-396 (2017).

252. J. Lv, M. Yang, J. Zhang and X. Y. Wang, "Respiratory motion correction for free-breathing 3D abdominal MRI using CNN-based image registration: a feasibility study," British Journal of Radiology **91**, 9 (2018).

253. S. Miao, Z. J. Wang and R. Liao, "A CNN Regression Approach for Real-Time 2D/3D Registration," IEEE Transactions on Medical Imaging **35**, 1352-1363 (2016).

254. J. Zheng, S. Miao, Z. Jane Wang and R. Liao, "Pairwise domain adaptation module for CNN-based 2-D/3-D registration," J Med Imaging (Bellingham) **5**, 021204 (2018).

2245 255. D. Nie, X. Cao, Y. Gao, L. Wang and D. Shen, "Estimating CT Image from MRI Data
2246   Using 3D Fully Convolutional Networks," Deep Learn Data Label Med Appl (2016)
2247   **2016**, 170-178 (2016).
2248 256. X. Han, "MR-based synthetic CT generation using a deep convolutional neural network
2249   method," Med Phys **44**, 1408-1419 (2017).
2250 257. A. P. Leynes, J. Yang, F. Wiesinger, S. S. Kaushik, D. D. Shanbhag, Y. Seo, T. A. Hope
2251   and P. E. Z. Larson, "Direct PseudoCT Generation for Pelvis PET/MRI Attenuation
2252   Correction using Deep Convolutional Neural Networks with Multi-parametric MRI: Zero
2253   Echo-time and Dixon Deep pseudoCT (ZeDD-CT)," J Nucl Med (2017).
2254 258. F. Liu, H. Jang, R. Kijowski, T. Bradshaw and A. B. McMillan, "Deep Learning MR
2255   Imaging-based Attenuation Correction for PET/MR Imaging," Radiology **286**, 676-684
2256   (2018).
2257 259. D. Nie, R. Trullo, J. Lian, C. Petitjean, S. Ruan, Q. Wang and D. Shen, "Medical image
2258   synthesis with context-aware generative adversarial networks," International Conference
2259   on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 417-425,
2260   2017.
2261 260. H. Choi and D. S. Lee, "Generation of structural MR images from amyloid PET:
2262   Application to MR-less quantification," J Nucl Med (2017).
2263 261. A. Ben-Cohen, E. Klang, S. P. Raskin, M. M. Amitai and H. Greenspan, "Virtual PET
2264   Images from CT Data Using Deep Convolutional Networks: Initial Results,"
2265   International Workshop on Simulation and Synthesis in Medical Imaging, 49-57, 2017.
2266 262. L. Wu, J. Z. Cheng, S. Li, B. Lei, T. Wang and D. Ni, "FUIQA: Fetal Ultrasound Image
2267   Quality Assessment With Deep Convolutional Networks," IEEE Transactions on
2268   Cybernetics **47**, 1336-1349 (2017).
2269 263. J. Neylon, Y. G. Min, D. A. Low and A. Santhanam, "A neural network approach for
2270   fast, automated quantification of DIR performance," Med Phys **44**, 4126-4138 (2017).
2271 264. J. H. Lee, B. R. Grant, J. H. Chung, I. Reiser and M. L. Giger, "Assessment of diagnostic
2272   image quality of computed tomography (CT) images of the lung using deep learning," in
2273   Proc. SPIE Medical Imaging, Vol. 10573, (2018), pp. 105731M
2274 265. S. J. Esses, X. Lu, T. Zhao, K. Shanbhogue, B. Dane, M. Bruno and H. Chandarana,
2275   "Automated image quality evaluation of T2 -weighted liver MRI utilizing deep learning
2276   architecture," J Magn Reson Imaging **47**, 723-728 (2018).
2277 266. C. McCollough, "Low Dose CT Grand Challenge, the Mayo Clinic, the American
2278   Association of Physicists in Medicine, and grants EB017095 and grants EB017185 from
2279   the National Institute of Biomedical Imaging and Bioengineering," (2016).
2280 267. G. Delso, F. Wiesinger, L. I. Sacolick, S. S. Kaushik, D. D. Shanbhag, M. Hullner and P.
2281   Veit-Haibach, "Clinical Evaluation of Zero-Echo-Time MR Imaging for the
2282   Segmentation of the Skull," Journal of Nuclear Medicine **56**, 417-422 (2015).
2283 268. K. H. Cha, L. Hadjiiski, H. P. Chan, A. Z. Weizer, A. Alva, R. H. Cohan, E. M. Caoili, C.
2284   Paramagul and R. K. Samala, "Bladder Cancer Treatment Response Assessment in CT
2285   using Radiomics with Deep-Learning," Sci Rep **7** (2017).
2286 269. A. Nielsen, M. B. Hansen, A. Tietze and K. Mouridsen, "Prediction of Tissue Outcome
2287   and Assessment of Treatment Effect in Acute Ischemic Stroke Using Deep Learning,"
2288   Stroke **49**, 1394-1401 (2018).
2289 270. B. Q. Huynh, N. Antropova and M. L. Giger, "Comparison of Breast DCE-MRI Contrast
2290   Time Points for Predicting Response to Neoadjuvant Chemotherapy Using Deep

2291   Convolutional Neural Network Features with Transfer Learning," in Proc. SPIE Medical
2292   Imaging, Vol. 10134, edited by S. G. Armato and N. A. Petrick (2017), pp. 101340U
2293 271.   K. Ravichandran, N. Braman, A. Janowczyk and A. Madabhushi, "A deep learning
2294   classifier for prediction of pathological complete response to neoadjuvant chemotherapy
2295   from baseline breast DCE-MRI," in Medical Imaging 2018: Computer-Aided Diagnosis,
2296   Vol. 10575, edited by N. Petrick and K. Mori (2018).
2297 272.   K. Men, P. Boimel, J. Janopaul-Naylor, H. Zhong, M. Huang, H. Geng, C. Cheng, Y.
2298   Fan, J. P. Plastaras, E. Ben-Josef and Y. Xiao, "Cascaded atrous convolution and spatial
2299   pyramid pooling for more accurate tumor target segmentation for rectal cancer
2300   radiotherapy," Phys Med Biol (2018).
2301 273.   M. Q. Ding, L. J. Chen, G. F. Cooper, J. D. Young and X. H. Lu, "Precision Oncology
2302   beyond Targeted Therapy: Combining Omics Data with Machine Learning Matches the
2303   Majority of Cancer Cells to Effective Therapeutics," Molecular Cancer Research **16**, 269-
2304   278 (2018).
2305 274.   N. Tong, S. Gou, S. Yang, D. Ruan and K. Sheng, "Fully Automatic Multi-Organ
2306   Segmentation for Head and Neck Cancer Radiotherapy Using Shape Representation
2307   Model Constrained Fully Convolutional Neural Networks," Med Phys (2018).
2308 275.   P. Jackson, N. Hardcastle, N. Dawe, T. Kron, M. S. Hofman and R. J. Hicks, "Deep
2309   Learning Renal Segmentation for Fully Automated Radiation Dose Estimation in
2310   Unsealed Source Therapy," Front Oncol **8**, 215 (2018).
2311 276.   M. Shehata, F. Khalifa, A. Soliman, M. Ghazal, F. Taher, M. Abou El-Ghar, A. Dwyer,
2312   G. Gimel'farb, R. Keynton and A. El-Baz, "Computer-Aided Diagnostic System for Early
2313   Detection of Acute Renal Transplant Rejection Using Diffusion-Weighted MRI," IEEE
2314   Trans Biomed Eng (2018).
2315 277.   B. Ibragimov, D. Toesca, D. Chang, Y. Yuan, A. Koong and L. Xing, "Development of
2316   deep neural network for individualized hepatobiliary toxicity prediction after liver
2317   SBRT," Med Phys (2018).
2318 278.   H. H. Tseng, Y. Luo, S. Cui, J. T. Chien, R. K. Ten Haken and I. El Naqa, "Deep
2319   reinforcement learning for automated radiation adaptation in lung cancer," Med Phys **44**,
2320   6690-6705 (2017).
2321 279.   M. D. Foote, B. Zimmerman, A. Sawant and S. Joshi, "Real-Time Patient-Specific Lung
2322   Radiotherapy Targeting using Deep Learning," in International Conference on Medical
2323   Imaging with Deep Learning (MIDL), (2018).
2324 280.   D. Nguyen, T. Long, X. Jia, W. Lu, X. Gu, Z. Iqbal and S. Jiang, "Dose prediction with
2325   U-Net: a feasibility study for predicting dose distributions from contours using deep
2326   learning on prostate IMRT patients," arXiv:1709.09233 (2017).
2327 281.   T. Kajikawa, N. Kadoya, K. Ito, Y. Takayama, T. Chiba, S. Tomori, K. Takeda and K.
2328   Jingu, "Automated prediction of dosimetric eligibility of patients with prostate cancer
2329   undergoing intensity-modulated radiation therapy using a convolutional neural network,"
2330   Radiol Phys Technol **11**, 320-327 (2018).
2331 282.   M. Maspero, M. H. G. Savenije, A. M. Dinkla, P. R. Seevinck, M. P. W. Intven, I. M.
2332   Jurgenliemk-Schulz, L. G. W. Kerkmeijer and C. A. T. van den Berg, "Fast synthetic CT
2333   generation with deep learning for general pelvis MR-only Radiotherapy,"
2334   arXiv:1802.06468 (2018).
2335 283.   X. Zhen, J. W. Chen, Z. C. Zhong, B. Hrycushko, L. H. Zhou, S. Jiang, K. Albuquerque
2336   and X. J. Gu, "Deep convolutional neural network with transfer learning for rectum

2337   toxicity prediction in cervical cancer radiotherapy: a feasibility study," Phys Med Biol
2338   **62**, 8246-8263 (2017).

2339 284.   J. E. Bibault, P. Giraud, C. Durdux, J. Taieb, A. Berger, R. Coriat, S. Chaussade, B.
2340   Dousset, B. Nordlinger and A. Burgun, "Deep Learning and Radiomics predict complete
2341   response after neo-adjuvant chemoradiation for locally advanced rectal cancer," Sci Rep
2342   **8**, 12611 (2018).

2343 285.   M. L. Giger, "Machine Learning in Medical Imaging," Journal of the American College
2344   of Radiology **15**, 512-520 (2018).

2345 286.   K. Men, T. Zhang, X. Chen, B. Chen, Y. Tang, S. Wang, Y. Li and J. Dai, "Fully
2346   automatic and robust segmentation of the clinical target volume for radiotherapy of breast
2347   cancer using big data and deep learning," Phys Med **50**, 13-19 (2018).

2348 287.   M. Everingham, S. M. A. Eslami, L. Van Gool, C. Williams, J. Winn and A. Zisserman,
2349   "The Pascal Visual Object Classes Challenge: A Retrospective," International journal of
2350   computer vision **111**, 98-136 (2015).

2351 288.   T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár and C. L.
2352   Zitnick, "Microsoft coco: Common objects in context," European conference on
2353   computer vision, 740-755, 2014.

2354 289.   A. Karpathy and L. Fei-Fei, "Deep Visual-Semantic Alignments for Generating Image
2355   Descriptions," IEEE transactions on pattern analysis and machine intelligence **39**, 664-
2356   676 (2017).

2357 290.   H. Nam, J.-W. Ha and J. Kim, "Dual attention networks for multimodal reasoning and
2358   matching," The IEEE Conference on Computer Vision and Pattern Recognition (CVPR),
2359   299-307, 2017.

2360 291.   B. Dai, Y. Zhang and D. Lin, "Detecting Visual Relationships with Deep Relational
2361   Networks," The IEEE Conference on Computer Vision and Pattern Recognition (CVPR),
2362   3076-3086, 2017.

2363 292.   F. Dyson, "A meeting with Enrico Fermi," Nature **427**, 297 (2004).

2364 293.   H. Mhaskar, Q. Liao and T. A. Poggio, "When and why are deep networks better than
2365   shallow ones?," Proceedings of the Thirty-First AAAI Conference on Artificial
2366   Intelligence, 2343--2349, San Francisco, CA, 2017.

2367 294.   R. Schwartz-Ziv and N. Tishby, "Opening the black box of deep neural networks via
2368   information," arXiv:1703.00810 (2017).

2369 295.   I. Goodfellow, Y. Bengio, A. Courville, G. I, I. Goodfellow, Y. Bengio and A. Courville,
2370   "Regularization for Deep Learning," Deep Learning, 221-265 (2016).

2371 296.   J. Prechelt, Early stopping - But when? (Springer Berlin Heidelberg, Berlin, Heidelberg,
2372   2012).

2373 297.   K. Fukunaga and R. R. Hayes, "Effects of sample size in classifier design," IEEE
2374   Transactions on Pattern Analysis and Machine Intelligence **11**, 873-885 (1989).

2375 298.   R. F. Wagner, H.-P. Chan, B. Sahiner, N. Petrick and J. T. Mossoba, "Finite-sample
2376   effects and resampling plans: applications to linear classifiers in computer-aided
2377   diagnosis," Medical Imaging 1997: Image Processing, 467-478, 1997.

2378 299.   H. P. Chan, B. Sahiner, R. F. Wagner and N. Petrick, "Classifier design for computer‐
2379   aided diagnosis: Effects of finite sample size on the mean performance of classical and
2380   neural network classifiers," Med Phys **26**, 2654-2668 (1999).

2381 300. B. Sahiner, H. P. Chan, N. Petrick, R. F. Wagner and L. Hadjiiski, "Feature selection and
2382 classifier performance in computer‐aided diagnosis: The effect of finite sample size,"
2383 Med Phys **27**, 1509-1522 (2000).
2384 301. M. A. Kupinski, D. C. Edwards, M. L. Giger and C. E. Metz, "Ideal observer
2385 approximation using Bayesian classification neural networks," IEEE transactions on
2386 medical imaging **20**, 886-899 (2001).
2387 302. J. Cho, K. Lee, E. Shin, G. Choy and S. Do, "How much data is needed to train a medical
2388 image deep learning system to achieve necessary high accuracy?," arXiv:1511.06348
2389 (2015).
2390 303. C. Sun, A. Shrivastava, S. Singh and A. Gupta, "Revisiting unreasonable effectiveness of
2391 data in deep learning era," 2017 IEEE International Conference on Computer Vision
2392 (ICCV), 843-852, 2017.
2393 304. R. Zhang, Y. Zheng, T. W. C. Mak, R. Yu, S. H. Wong, J. Y. Lau and C. C. Poon,
2394 "Automatic detection and classification of colorectal polyps by transferring low-level
2395 CNN features from nonmedical domain," IEEE Journal of Biomedical and Health
2396 Informatics **21**, 41-47 (2017).
2397 305. B. van Ginneken, A. A. A. Setio, C. Jacobs and F. Ciompi, "Off-the-shelf convolutional
2398 neural network features for pulmonary nodule detection in computed tomography scans,"
2399 in 2015 IEEE 12th International Symposium on Biomedical Imaging, (2015), pp. 286-
2400 289.
2401 306. U. K. Lopes and J. F. Valiati, "Pre-trained convolutional neural networks as feature
2402 extractors for tuberculosis detection," Computers in Biology and Medicine **89**, 135-143
2403 (2017).
2404 307. A. Ben-Cohen, E. Klang, I. Diamant, N. Rozendorn, S. P. Raskin, E. Konen, M. M.
2405 Amitai and H. Greenspan, "CT Image-based Decision Support System for Categorization
2406 of Liver Metastases Into Primary Cancer Sites: Initial Results," Academic Radiology **24**,
2407 1501-1509 (2017).
2408 308. A. Z. Abidin, B. T. Deng, A. M. Dsouza, M. B. Nagarajan, P. Coan and A. Wismuller,
2409 "Deep transfer learning for characterizing chondrocyte patterns in phase contrast X-Ray
2410 computed tomography images of the human patellar cartilage," Computers in Biology
2411 and Medicine **95**, 24-33 (2018).
2412 309. H. Lee, S. Tajmir, J. Lee, M. Zissen, B. A. Yeshiwas, T. K. Alkasab, G. Choy and S. Do,
2413 "Fully Automated Deep Learning System for Bone Age Assessment," J Digit Imaging **30**,
2414 427-441 (2017).
2415 310. R. K. Samala, H.-P. Chan, L. Hadjiiski, M. A. Helvie, C. Richter and K. Cha, "Cross-
2416 domain and multi-task transfer learning of deep convolutional neural network for breast
2417 cancer diagnosis in digital breast tomosynthesis," Medical Imaging 2018: Computer-
2418 Aided Diagnosis, 105750Q, 2018.
2419 311. H. C. Shin, M. R. Orton, D. J. Collins, S. J. Doran and M. O. Leach, "Stacked
2420 autoencoders for unsupervised feature learning and multiple organ detection in a pilot
2421 study using 4D patient data," IEEE Trans Pattern Anal Mach Intell **35**, 1930-43 (2013).
2422 312. S. Wang, Y. Cong, H. Fan, L. Liu, X. Li, Y. Yang, Y. Tang, H. Zhao and H. Yu,
2423 "Computer-Aided Endoscopic Diagnosis Without Human-Specific Labeling," IEEE
2424 Trans Biomed Eng **63**, 2347-2358 (2016).
2425 313. X. Feng, J. Yang, A. F. Laine and E. D. Angelini, "Discriminative Localization in CNNs
2426 for Weakly-Supervised Segmentation of Pulmonary Nodules," International Conference

2427   on Medical Image Computing and Computer-Assisted Intervention (MICCAI) **10435**,
2428   568-576 (2017).

2429 314. P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C.
2430   Langlotz and K. Shpanskaya, "CheXNet: Radiologist-Level Pneumonia Detection on
2431   Chest X-Rays with Deep Learning," arXiv:1711.05225 (2017).

2432 315. X. Chen, A. Shrivastava and A. Gupta, "NEIL: Extracting Visual Knowledge from Web
2433   Data," Proc. of ICCV, 2013.

2434 316. J. Zech, M. Pain, J. Titano, M. Badgeley, J. Schefflein, A. Su, A. Costa, J. Bederson, J.
2435   Lehar and E. K. Oermann, "Natural Language-based Machine Learning Models for the
2436   Annotation of Clinical Radiology Reports," Radiology **287**, 570-580 (2018).

2437 317. M. Ghafoorian, J. Teuwen, R. Manniesing, F.-E. de Leeuw, B. van Ginneken, N.
2438   Karssemeijer and B. Platel, "Student beats the teacher: deep neural networks for lateral
2439   ventricles segmentation in brain MR," in Proc. SPIE Medical Imaging, Vol. 10574,
2440   edited by E. D. Angelini and B. A. Landman (2018), pp. 105742U.

2441 318. L. Zhang, V. Gopalakrishnan, L. Lu, R. M. Summers, J. Moss and J. Yao, "Self-learning
2442   to detect and segment cysts in lung CT images without manual annotation," Biomedical
2443   Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on, 1100-1103, 2018.

2444 319. A. Asperti and C. Mastronardo, "The Effectiveness of Data Augmentation for Detection
2445   of Gastrointestinal Diseases from Endoscopical Images," the 5th International
2446   Conference on Bioimaging, 2017.

2447 320. A. Pezeshk, N. Petrick, W. Chen and B. Sahiner, "Seamless Lesion Insertion for Data
2448   Augmentation in CAD Training," IEEE Transactions on Medical Imaging **36**, 1005-1015
2449   (2017).

2450 321. C. Zhang, W. Tavanapong, J. Wong, P. C. de Groen and J. Oh, "Real Data Augmentation
2451   for Medical Image Classification," 67-76, 2017.

2452 322. A. Badano, A. Badal, S. Glick, C. G. Graff, F. Samuelson, D. Sharma and R. P. Zeng, "In
2453   silico imaging clinical trials for regulatory evaluation: initial considerations for VICTRE,
2454   a demonstration study," in Proc. SPIE Medical Imaging, Vol. 10132, edited by T. G.
2455   Flohr, J. Y. Lo and T. G. Schmidt (2017), pp. 1013220.

2456 323. J. Cui, X. Liu, Y. Wang and H. Liu, "Deep reconstruction model for dynamic PET
2457   images," PloS one **12**, e0184667 (2017).

2458 324. H.-C. Shin, L. Lu and R. M. Summers, "Natural Language Processing for Large-Scale
2459   Medical Image Analysis Using Deep Learning," in Deep Learning for Medical Image
2460   Analysis, (Elsevier, 2017), pp. 405-421.

2461 325. T. Schlegl, S. M. Waldstein, W.-D. Vogl, U. Schmidt-Erfurth and G. Langs, "Predicting
2462   Semantic Descriptions from Medical Images with Convolutional Neural Networks,"
2463   Information Processing in Medical Imaging, 437-448, 2015.

2464 326. H.-C. Shin, K. Roberts, L. Lu, D. Demner-Fushman, J. Yao and R. M. Summers,
2465   "Learning to read chest x-rays: Recurrent neural cascade model for automated image
2466   annotation," Proceedings of the IEEE Conference on Computer Vision and Pattern
2467   Recognition, 2497-2506, 2016.

2468 327. X. Wang, L. Lu, H.-C. Shin, L. Kim, M. Bagheri, I. Nogues, J. Yao and R. M. Summers,
2469   "Unsupervised joint mining of deep features and image labels for large-scale radiology
2470   image categorization and scene recognition," 2017 IEEE Winter Conference on
2471   Applications of Computer Vision (WACV), 998-1007, 2017.

2472 328. X. Wang, Y. Peng, L. Lu, Z. Lu and R. M. Summers, "TieNet: Text-Image Embedding
2473   Network for Common Thorax Disease Classification and Reporting in Chest X-rays,"
2474   International Conference of Computer Vision and Pattern Recognition, 2018.
2475 329. M. C. Chen, R. L. Ball, L. Yang, N. Moradzadeh, B. E. Chapman, D. B. Larson, C. P.
2476   Langlotz, T. J. Amrhein and M. P. Lungren, "Deep Learning to Classify Radiology Free-
2477   Text Reports," Radiology **286**, 845-852 (2018).
2478 330. K. Yan, X. Wang, L. Lu and R. M. Summers, "DeepLesion: Automated Deep Mining,
2479   Categorization and Detection of Significant Radiology Image Findings using Large-Scale
2480   Clinical Lesion Annotations," arXiv preprint arXiv:1710.01766 (2017).
2481 331. K. Yan, X. Wang, L. Lu, L. Zhang, A. Harrison, M. Bagheri and R. Summers, "Deep
2482   Lesion Graphs in the Wild: Relationship Learning and Organization of Significant
2483   Radiology Image Findings in a Diverse Large-scale Lesion Database," International
2484   Conference of Computer Vision and Pattern Recognition, 2018.
2485 332. J. E. Bibault, P. Giraud and A. Burgun, "Big Data and machine learning in radiation
2486   oncology: State of the art and future prospects," Cancer Letters **382**, 110-117 (2016).
2487 333. L. Dai, R. Fang, H. Li, X. Hou, B. Sheng, Q. Wu and W. Jia, "Clinical report guided
2488   retinal microaneurysm detection with multi-sieving deep learning," IEEE Transactions on
2489   Medical Imaging **37**, 1149-1161 (2018).
2490 334. Z. Zhang, P. Chen, M. Sapkota and L. Yang, "TandemNet: Distilling Knowledge from
2491   Medical Images Using Diagnostic Reports as Optional Semantic References,"
2492   International Conference on Medical Image Computing and Computer-Assisted
2493   Intervention (MICCAI), 320-328, 2017.
2494 335. A. Top, G. Hamarneh and R. Abugharbieh, "Active Learning for Interactive 3D Image
2495   Segmentation," International Conference on Medical Image Computing and Computer-
2496   Assisted Intervention (MICCAI), 603-610, Berlin, Heidelberg, 2011.
2497 336. Y. Zhu, S. Zhang, W. Liu and D. N. Metaxas, "Scalable histopathological image analysis
2498   via active learning," International Conference on Medical Image Computing and
2499   Computer-Assisted Intervention (MICCAI), 369-376, 2014.
2500 337. J. M. B. Dwarikanath Mahapatra, "Visual saliency-based active learning for prostate
2501   magnetic resonance imaging segmentation," Journal of Medical Imaging **3**, 3-3 (2016).
2502 338. J. Lee, Y. Wu and H. Kim, "Unbalanced data classification using support vector
2503   machines with active learning on scleroderma lung disease patterns," Journal of Applied
2504   Statistics **42**, 676-689 (2015).
2505 339. S. C. Hoi, R. Jin, J. Zhu and M. R. Lyu, "Batch mode active learning and its application
2506   to medical image classification," Proceedings of the 23rd international conference on
2507   Machine learning, 417-424, 2006.
2508 340. K. Konyushkova, R. Sznitman and P. Fua, "Introducing geometry in active learning for
2509   image segmentation," Proceedings of the IEEE International Conference on Computer
2510   Vision, 2974-2982, 2015.
2511 341. L. Yang, Y. Zhang, J. Chen, S. Zhang and D. Z. Chen, "Suggestive annotation: A deep
2512   active learning framework for biomedical image segmentation," International Conference
2513   on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 399-407,
2514   2017.
2515 342. Z. Zhou, J. Shin, L. Zhang, S. Gurudu, M. Gotway and J. Liang, "Fine-Tuning
2516   Convolutional Neural Networks for Biomedical Image Analysis: Actively and

2517   Incrementally," Computer Vision and Pattern Recognition (CVPR), 2017 IEEE
2518   Conference on, 4761-4772, 2017.

2519   343.   U. Gaur, M. Kourakis, E. Newman-Smith, W. Smith and B. S. Manjunath, "Membrane
2520   segmentation via active learning with deep networks," 2016 IEEE International
2521   Conference on Image Processing (ICIP), 1943-1947, 2016.

2522   344.   A. Mosinska-Domanska, R. Sznitman, P. Glowacki and P. Fua, "Active learning for
2523   delineation of curvilinear structures," Computer Vision and Pattern Recognition (CVPR),
2524   2016 IEEE Conference on, 5231-5239, 2016.

2525   345.   T. Heimann, P. Mountney, M. John and R. Ionasec, "Learning without labeling: Domain
2526   adaptation for ultrasound transducer localization," International Conference on Medical
2527   Image Computing and Computer-Assisted Intervention (MICCAI), 49-56, 2013.

2528   346.   C. Wachinger and M. Reuter, "Domain adaptation for Alzheimer's disease diagnostics,"
2529   NeuroImage **139**, 470-479 (2016).

2530   347.   S. Conjeti, A. Katouzian, A. G. Roy, L. Peter, D. Sheet, S. Carlier, A. Laine and N.
2531   Navab, "Supervised domain adaptation of decision forests: Transfer of models trained in
2532   vitro for in vivo intravascular ultrasound tissue characterization," Medical Image
2533   Analysis **32**, 1-17 (2016).

2534   348.   R. Bermúdez-Chacón, C. Becker, M. Salzmann and P. Fua, "Scalable unsupervised
2535   domain adaptation for electron microscopy," International Conference on Medical Image
2536   Computing and Computer-Assisted Intervention (MICCAI), 326-334, 2016.

2537   349.   C. Becker, C. M. Christoudias and P. Fua, "Domain adaptation for microscopy imaging,"
2538   IEEE transactions on medical imaging **34**, 1125-1139 (2015).

2539   350.   C. Baur, S. Albarqouni and N. Navab, "Semi-supervised Deep Learning for Fully
2540   Convolutional Networks," International Conference on Medical Image Computing and
2541   Computer-Assisted Intervention (MICCAI), 311-319, 2017.

2542   351.   R. Saunders, E. Samei, J. Baker and D. Delong, "Simulation of mammographic lesions,"
2543   Academic Radiology **13**, 860-870 (2006).

2544   352.   A. Rashidnasab, P. Elangovan, M. Yip, O. Diaz, D. R. Dance, K. C. Young and K. Wells,
2545   "Simulation and assessment of realistic breast lesions using fractal growth model," Phys
2546   Med Biol **58**, 5613-5627 (2013).

2547   353.   F. J. Martinez-Murcia, J. M. Górriz, J. Ramírez, I. A. Illán, F. Segovia, D. Castillo-
2548   Barnes and D. Salas-Gonzalez, "Functional Brain Imaging Synthesis Based on Image
2549   Decomposition and Kernel Modeling: Application to Neurodegenerative Diseases,"
2550   Frontiers in neuroinformatics **11**, 65 (2017).

2551   354.   F. Calimeri, A. Marzullo, C. Stamile and G. Terracina, "Biomedical Data Augmentation
2552   Using Generative Adversarial Neural Networks," Artificial Neural Networks and
2553   Machine Learning (ICANN) 626-634, 2017.

2554   355.   A. Lahiri, K. Ayush, P. K. Biswas and P. Mitra, "Generative adversarial learning for
2555   reducing manual annotation in semantic segmentation on large scale miscroscopy images:
2556   Automated vessel segmentation in retinal fundus image as test case," Proceedings of the
2557   IEEE Conference on Computer Vision and Pattern Recognition Workshops, 42-48, 2017.

2558   356.   L. Zhang, A. Gooya and A. F. Frangi, "Semi-supervised assessment of incomplete lv
2559   coverage in cardiac mri using generative adversarial nets," International Workshop on
2560   Simulation and Synthesis in Medical Imaging, 61-68, 2017.

2561   357.   N. Bayramoglu, M. Kaakinen, L. Eklund and J. Heikkila, "Towards virtual h&e staining
2562   of hyperspectral lung histology images using conditional generative adversarial

2563 networks," Proceedings of the IEEE Conference on Computer Vision and Pattern
2564 Recognition, 64-71, 2017.
2565 358. P. Costa, A. Galdran, M. I. Meyer, M. Niemeijer, M. Abràmoff, A. M. Mendonça and A.
2566 Campilho, "End-to-end adversarial retinal image synthesis," IEEE transactions on
2567 medical imaging (2017).
2568 359. A. Chartsias, T. Joyce, R. Dharmakumar and S. A. Tsaftaris, "Adversarial Image
2569 Synthesis for Unpaired Multi-modal Cardiac Data," International Workshop on
2570 Simulation and Synthesis in Medical Imaging, 3-13, 2017.
2571 360. J. M. Wolterink, A. M. Dinkla, M. H. Savenije, P. R. Seevinck, C. A. van den Berg and I.
2572 Išgum, "Deep MR to CT synthesis using unpaired data," International Workshop on
2573 Simulation and Synthesis in Medical Imaging, 14-23, 2017.
2574 361. M. J. Chuquicusma, S. Hussein, J. Burt and U. Bagci, "How to Fool Radiologists with
2575 Generative Adversarial Networks?: A Visual Turing Test for Lung Cancer Diagnosis,"
2576 IEEE International Symposium on Biomedical Imaging 2018.
2577 362. N. A. Obuchowski, H. X. Barnhart, A. J. Buckler, G. Pennello, X. F. Wang, J. Kalpathy-
2578 Cramer, H. J. Kim, A. P. Reeves and C. E. W. Grp, "Statistical issues in the comparison
2579 of quantitative imaging biomarker algorithms using pulmonary nodule volume as an
2580 example," Statistical Methods in Medical Research **24**, 107-140 (2015).
2581 363. N. A. Obuchowski, A. P. Reeves, E. P. Huang, X. F. Wang, A. J. Buckler, H. J. Kim, H.
2582 X. Barnhart, E. F. Jackson, M. L. Giger, G. Pennello, A. Y. Toledano, J. Kalpathy-
2583 Cramer, T. V. Apanasovich, P. E. Kinahan, K. J. Myers, D. B. Goldgof, D. P. Barboriak,
2584 R. J. Gillies, L. H. Schwartz, D. C. Sullivan and A. C. W. Grp, "Quantitative imaging
2585 biomarkers: A review of statistical methods for computer algorithm comparisons,"
2586 Statistical Methods in Medical Research **24**, 68-106 (2015).
2587 364. L. C. Lazzeroni, Y. Lu and I. Belitskaya-Levy, "P-values in genomics: Apparent
2588 precision masks high uncertainty," Molecular Psychiatry **19**, 1336-1340 (2014).
2589 365. K. Drukker, L. Pesce and M. Giger, "Repeatability in computer-aided diagnosis:
2590 Application to breast cancer diagnosis on sonography," Med Phys **37**, 2659-2669 (2010).
2591 366. S. Holm, "A Simple Sequentially Rejective Multiple Test Procedure," Scandinavian
2592 Journal of Statistics **6**, 65-70 (1979).
2593 367. D. Curran-Everett, "Multiple comparisons: philosophies and illustrations," American
2594 Journal of Physiology-Regulatory Integrative and Comparative Physiology **279**, R1-R8
2595 (2000).
2596 368. L. Ein-Dor, I. Kela, G. Getz, D. Givol and E. Domany, "Outcome signature genes in
2597 breast cancer: is there a unique set?," Bioinformatics **21**, 171-178 (2005).
2598 369. L. Ein-Dor, O. Zuk and E. Domany, "Thousands of samples are needed to generate a
2599 robust gene list for predicting outcome in cancer," Proceedings of the National Academy
2600 of Sciences of the United States of America **103**, 5923-5928 (2006).
2601 370. C. G. Begley and L. M. Ellis, "Raise standards for preclinical cancer research," Nature
2602 **483**, 531-533 (2012).
2603 371. A. Chalkidou, M. J. O'Doherty and P. K. Marsden, "False Discovery Rates in PET and
2604 CT Studies with Texture Features: A Systematic Review," Plos One **10** (2015).
2605 372. N. P. Gruszauskas, K. Drukker, M. L. Giger, C. A. Sennett and L. L. Pesce,
2606 "Performance of breast ultrasound computer-aided diagnosis: Dependence on image
2607 selection," Academic Radiology **15**, 1234-1245 (2008).

2608 373. N. P. Gruszauskas, K. Drukker, M. L. Giger, R. F. Chang, C. A. Sennett, W. K. Moon
2609 and L. L. Pesce, "Breast US Computer-aided Diagnosis System: Robustness across Urban
2610 Populations in South Korea and the United States," Radiology **253**, 661-671 (2009).

2611 374. K. R. Mendel, H. Li, L. Lan, C. M. Cahill, V. Rael, H. Abe and M. L. Giger,
2612 "Quantitative texture analysis: robustness of radiomics across two digital mammography
2613 manufacturers' systems," Journal of Medical Imaging **5** (2018).

2614 375. J. Kalpathy-Cramer, A. Mamomov, B. S. Zhao, L. Lu, D. Cherezov, S. Napel, S.
2615 Echegaray, D. Rubin, M. McNitt-Gray, P. Lo, J. C. Sieren, J. Uthoff, S. K. N. Dilger, B.
2616 Driscoll, I. Yeung, L. Hadjiiski, K. Cha, Y. Balagurunathan, R. Gillies and D. Goldgof,
2617 "Radiomics of Lung Nodules: A Multi-Institutional Study of Robustness and Agreement
2618 of Quantitative Imaging Features," Tomography **2**, 430-437 (2016).

2619 376. L. E. Court, "Harmonization & Robustness in Radiomics," Med Phys **43**, 3695-3696
2620 (2016).

2621 377. A. Holzinger, C. Biemann, C. S. Pattichis and D. B. Kell, "What do we need to build
2622 explainable AI systems for the medical domain?," arXiv:1712.09923 (2017).

2623 378. W. Samek, T. Wiegand and K.-R. Müller, "Explainable Artificial Intelligence:
2624 Understanding, Visualizing and Interpreting Deep Learning Models," arXiv:1708.08296
2625 (2017).

2626 379. M. D. Zeiler and R. Fergus, "Visualizing and Understanding Convolutional Networks,"
2627 arXiv:1311.2901 (2014).

2628 380. M. T. Ribeiro, S. Singh and C. Guestrin, ""Why Should I Trust You?": Explaining the
2629 Predictions of Any Classifier," arXiv:1602.04938 (2016).

2630 381. K. Simonyan, A. Vedaldi and A. Zisserman, "Deep inside convolutional networks:
2631 Visualising image classification models and saliency maps," arXiv:1312.6034 (2014).

2632 382. B. Zhou, A. Khosla, A. Lapedriza, A. Oliva and A. Torralba, "Learning Deep Features
2633 for Discriminative Localization," arXiv:1512.04150 (2015).

2634 383. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra, "Grad-
2635 CAM: Visual Explanations from Deep Networks via Gradient-based Localization,"
2636 arXiv:1610.02391 (2017).

2637 384. G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. van der
2638 Laak, B. van Ginneken and C. I. Sanchez, "A survey on deep learning in medical image
2639 analysis," Medical Image Analysis **42**, 60-88 (2017).

2640 385. K. Adil, F. Jiang, S. H. Liu, A. Grigorev, B. B. Gupta and S. Rho, "Training an Agent for
2641 FPS Doom Game using Visual Reinforcement Learning and VizDoom," International
2642 Journal of Advanced Computer Science and Applications **8**, 32-41 (2017).

2643 386. R. M. Summers, "Are we at a crossroads or a plateau? Radiomics and machine learning
2644 in abdominal oncology imaging," Abdominal Radiology, 1-5 (2018).

2645

2646

2647 **Figure Legends**

2648 Fig. 1: CNN with two convolution layers each followed by a pooling layer, and one fully connected layer.

2649 Fig. 2: Number of peer-reviewed publications in radiologic medical imaging that involved DL. Peer-

2650 reviewed publications were searched on PubMed using the criteria ("deep learning" OR "deep neural

2651 network" OR deep convolution OR deep convolutional OR convolution neural network OR "shift-

2652 invariant artificial neural network" OR MTANN) AND (radiography OR x-ray OR mammography OR

2653 CT OR MRI OR PET OR ultrasound OR therapy OR radiology OR MR OR mammogram OR SPECT).

2654 The search only covered the first three months of 2018 and the result was linearly extended to the rest of

2655 2018.

2656 Fig 3: Use of CNN as a feature extractor.[196] (a) Each ROI is sent through AlexNet and the outputs from

2657 each layer are preprocessed to be used as sets of features for an SVM. The filtered image outputs from

2658 some of the layers can be seen in the left column. The numbers in parentheses for the center column

2659 denote the dimensionality of the outputs from each layer. The numbers in parentheses for the right

2660 column denote the length of the feature vector per ROI used as an input for the SVM after zero-variance

2661 removal. (b) Performance in terms of area under the receiver operating characteristic curve for classifiers

2662 based on features from each layer of AlexNet in the task of distinguishing between malignant and benign

2663 breast tumors.

2664 Fig. 4: CNN-extracted and conventional features can be combined in a number of ways,

2665 including a traditional classifier such as an SVM.[196]

2666 Fig. 5: The use of training, validation, and test sets for the design and performance evaluation of a

2667 supervised machine learning algorithm.

2668 Fig. 6: A disease image categorization framework using both images and texts.[327]

2669   Fig. 7: Eight sample disease keywords and images mined from PACS.[180]

2670

2671   **Table Headings**

Table I: Organ and substructure segmentation summary and performance using DL.

Region	Segmentation Object	Network Input	Network Architecture Basis	Data Set (train/test)	Dice Coefficient on Test Set
Abdomen	Skeletal muscle[89]	Whole Image	FCN	250/150 patients	0.93
	Subcutaneous and visceral fat areas[90]	Image Patch	Custom	20/20 patients	0.92 - 0.98
	Liver, spleen, kidneys[91]	Whole Image	Custom	140 scans 5-fold CV	0.94 – 0.96
Bladder	Bladder[76]	Image Patch	CifarNet	81/93 patients	0.86
Brain	Anterior visual pathway[92]	Whole Image	AE	165 patients LOO CV	0.78
	Bones[86]	Whole image	U-net	16 patients LOO CV	0.94
	Striatum[93]	Whole Image	Custom	15/18 patients	0.83
	Substructures[94]	Image Patch	Custom	15/20 patients	0.86 – 0.95
	Substructures[95]	Image Patch	Custom	20/10 patients	0.92
	Substructures[96]	Image Patch	Deep Residual Network[92]	18 patients 6-fold CV	0.69 – 0.83
	Substructures[97]	Whole Image	FCN	150/947 patients	0.86 – 0.92
Breast	Dense tissue and fat[98]	Image Patch	Custom	493 images 5-fold CV	0.63 – 0.95

	Breast and fibroglandular tissue[85]	Whole Image	U-net	66 patients 3-fold CV	0.85 – 0.94
Head and Neck	Organs-at-risk[83]	Image Patch	Custom	50 patients 5-fold CV	0.37 – 0.90
Heart	Left ventricle[79]	Whole Image	AE	15/15 patients	0.93
	Left ventricle[82]	Whole Image	AE	15/15 patients	0.94
	Left ventricle[99]	Image Patch	Custom	100/100 patients	0.86
	Left ventricle[100]	Image Patch	Custom	100/100 patients	0.88
	Fetal left ventricle[101]	Image Patch	Custom	10/41 patients	0.95
	Right ventricle[78]	Whole Image	AE	16/16 patients	0.82
Kidney	Kidney[102]	Whole Image	Custom	2000/400 patients	0.97
	Kidney[103]	Whole Image	FCN	165/79 patients	0.86
Knee	Femur, femoral cartilage, tibia, tibial cartilage[81]	Whole Image	Custom	60/40 images	-
Liver	Liver[80]	Image Patch	Custom	78/40 patients	-
	Liver[104]	Image Patch	Custom	109/32 patients	0.97
	Portal vein[83]	Image Patch	Custom	72 scans 8-fold CV	0.70

Lung	Lung[105]	Whole Image	HNN	62 slices/31 patients	0.96 – 0.97
Pancreas	Pancreas[106]	Image Patch	Custom	80 patients 6-fold CV	0.71
	Pancreas[107]	Image Patch	Custom	82 patients 4-fold CV	0.72
Prostate	Prostate[108]	Image Patch	AE	66 patients 2-fold CV	0.87
	Prostate[109]	Image Patch	Custom	30 patients LOO CV	0.87
	Prostate[110]	Whole Image	FCN	41/99 patients	0.85
	Prostate[87]	Whole Image	HNN	250 patients 5-fold CV	0.90
Rectum	Organs-at-risk[111]	Whole Image	VGG-16	218/60 patients	0.88 – 0.93
Spine	Intervertebral disk[112]	Image Patch	Custom	18/6 scans	0.91
Whole body	Multiple organs[113]	Whole Image	FCN	228/12 scans	-
Multiple organs	Liver and heart (blood pool, myocardium)[114]	Whole Image	Custom	Liver: 20/10 patients Heart: 10/10 patients	0.74 – 0.93

Note: A "-" on the performance metrics means that the authors report different segmentation accuracy metrics. Abbreviations: AE: Auto-encoder. FCN: Fully Convolutional Network. HNN: Holistically-Nested Network. LOO: Leave-one-out. CV: Cross-validation.

Table II: Lesion segmentation summary and performance using DL.

Region	Segmentation Object	Network Input	Network Architecture Basis	Data Set (train/test)	Dice Coefficient on Test Set
Bladder	Bladder lesion[77]	Image Patch	CifarNet	62 patients LOO CV	0.51
Breast	Breast lesion[118]	Image Patch	Custom	107 patients 4-fold CV	0.93
Bone	Osteosarcoma[119]	Whole Image	ResNet-50	15/8 patients	0.89
	Osteosarcoma[120]	Whole Image	FCN	1900/405 images from 23 patients	0.90
Brain	Brain lesion[121]	Image Patch	Custom	61 patients 5-fold CV	0.65
	Brain metastases[122]	Image Patch	Custom	225 patients 5-fold CV	0.67
	Brain tumor[115]	Image Patch	AE	HGG: 150/69 patients, LGG: 20/23 patients	HGG: 0.86 LGG: 0.82
	Brain tumor[117]	Image Patch	Custom	HGG: 220, LGG: 54, 5-fold CV	HGG: 0.85 – 0.91 LGG: 0.83 – 0.86
	Brain tumor[123]	Whole Image	Custom	30/25 patients	0.88
	Brain tumor[124]	Whole Image	FCN	274/110 patients	0.82
	Brain tumor[88]	Whole Image	HNN	20/10 patients	0.83

	Ischemic lesions[125]	Whole Image	DeConvNet	380/381 patients	0.88
	Multiple sclerosis lesion[126]	Whole Image	Custom	250/77 patients	0.64
	White matter hyper-intensities[116]	Image Patch	AE	100/135 patients	0.88
	White matter hyper-intensities[127]	Image Patch	Custom	378/50 patients	0.79
Head and neck	Nasopharyngeal cancer[128]	Whole Image	VGG-16	184/46 patients	0.81 – 0.83
	Thyroid nodule[129]	Image Patch	HNN	250 patients 5-fold CV	0.92
Liver	Liver lesion[130]	Image Patch	Custom	26 patients LOO CV	0.80
Lung	Lung nodule[131]	Image Patch	Custom	350/493 nodules	0.82
Lymph nodes	Lymph nodes[132]	Whole Image	HNN	171 patients 4-fold CV	0.82
Rectum	Rectal cancer[133]	Image Patch	Custom	70/70 patients	0.68
Skin	Melanoma[134]	Image Patch	Custom	126 images 4-fold CV	-

Note: A "-" on the performance metrics means that the authoers report different segmentation accuracy metrics. Abbreviations: AE: Auto-encoder. FCN: Fully Convolutional Network. HNN: Holistically-Nested Network. LOO: Leave-one-out. CV: Cross-validation. HGG: High Grade Glioma. LGG: Low Grade Glioma.

Table III: Organ and Anatomical structure detection summary and performance.

Organ	Detection Object	Network Input	Network Architecture Basis	Data Set (train/test)	Error (Mean± STD)
Bone	37 hand landmarks[147]	X-ray images	Custom CNN	895 images 3-fold CV	1.19±1.14 mm
	Femur bone[135]	MR 2.5D image patches	Custom 3D CNN	40/10 volumes	4.53±2.31 mm
	vertebrae[148]	MR/CT image patches	Custom CNN	1150 patches/ 110 images	3.81±2.98 mm
	vertebrae[149]	US/X-ray images	U-Net	22/19 patients	F1:0.90
Vessel	carotid artery[150]	CT 3D image patches	Custom 3D CNN	455 patient four-fold CV	2.64±4.98 mm
	ascending aorta[139]	3D US	Custom CNN	719/150 patients	1.04±0.50 mm
Fetal anatomy	Abdominal standard scan plane[136, 151]	US image patches	Custom CNN	11942/8718 images	F1:0.71[136], 0.75[151]
	12 standard scan planes[137]	US images	Custom CNN	800/200 images	F1:0.42-0.93
	13 standard scan planes[138]	US images	AlexNet	5229/2339 images	Acc: 0.10-0.94
Body	Body parts[152]	CT images	AlexNet + FCN	450/49 patients	3.9±4.7 voxels
	Body parts[153]	CT images	AlexNet	3438/860 images	AUC: 0.998
	Multiple	3D CT images	Custom CNN	200/200	F1:0.97

	Organ[154]			scans	
	Body parts[141, 142]	CT images	LeNet	2413/4043 images	F1:0.92
Brain	Brain landmarks[155]	MR images	FCN	350/350 images	2.94±1.58 mm
Lung	Pathologic Lung[156]	CT images	FCN	929 scans 5-fold CV	0.76±0.53 mm
Extremities	Thigh muscle[157]	MR images	FCN	15/10 patients	1.4±0.8 mm
Heart	Ventricle landmarks[143-145]	MRI images	Custom CNN + RL	801/90 images	2.9±2.4 mm

Abbreviations: FCN: Fully Convolutional Network. RL: Reinforcement learning. F1: harmonic average of the precision (positive predictive value) and recall (sensitivity). AUC: Area under the receiver operating characteristic curve. CV: Cross-validation.

Table IV: Lesion detection using DL.

Detection Organ	Lesion Type	Data set (train/test)	Network Input	Network Architecture Basis
Lung and Thorax	Pulmonary Nodule	888 patients 5-fold CV[168]	Image Patch[168, 169, 173-177] Whole Image[178-180]	CNN[168, 169, 173, 175-180] SDAE/CNN[174]
		888 patients 10-fold CV[169]		
		303 patients 10-fold CV[173]		
		2400 images 10-fold CV[174]		
		104 patients 5-fold CV[175]		
		1006 patients 10-fold CV[176]		
	Multiple	35,038/2,443		

	Pathologies	radiographs[178]		
		76,000/22,000 chest x-rays[180]		
		ImageNet Pre-training, 433 patients LOO CV[181]		
	Tuberculosis	685/151 chest radiographs[179]		
Brain	Cerebral Aneurism	300/100 magnetic resonance angiography images [182]	Image Patch[182] Whole Image[170, 172]	CNN[182] FCN/CNN[170, 172]
	Cerebral microbleed	230/50 brain MR scans[172]		
	Lacune	868/111 brain MR scans[170]		
Breast	Solid Cancer	40,000/18,000 mammographic images[64]	Image Patch[17, 64, 183] Whole Image[66, 161]	CNN[17, 64, 66, 183] FCN/CNN[161]
		161/160 Breast MR images[183]		
	Mass	Pre-training on ~2,300 mammography images, 277/47 DBT cases[17]		
		ImageNet Pre-training, 306/163 breast ultrasounds images[161]		
	Malignant mass & Mirco-calcification	ImageNet Pre-training, 3476/115 FFDM images[66]		
Colon	Polyp	394/792 CT colonography cases[166]	Whole Image[184]	CNN[166, 184, 185]

		101 CT colonography cases;10-fold CV[185]	Image Patch[166, 185]	
	Colitis	ImageNet Pre-training,160 abdominal CT cases; 4-fold CV[184]		
Multiple	Lymph Node	ImageNet Pre-training, 176 CT cases; 3-fold CV[160]	Image Patch[160, 166, 186]	CNN[160, 166, 186]
		69/17 abdominal CT cases[166]		
		176 abdominal CT cases; 3-fold CV[186]		
Liver	Tumor	NA/37[187]	Image Patch[187]	CNN[187]
Thyroid	Nodule	21,523 Ultrasound images; 10-fold CV[188]	Image Patch[188]	CNN[188]
Prostate	Cancer	196 MR cases; 10-fold CV[189]	Whole Image[189]	FCN[189]
Pericardium	Effusion	20/5 CT cases[190]	Whole Image[190]	FCN[190]
Vascular	Calcification	ImageNet Pre-training; 84/28[191]	Image Patch[191]	FCN[191]

Abbreviations: SDAE: Stacked Denoising Auto-encoder. FCN: Fully Convolutional Network. LOO: Leave-one-out. CV: Cross-validation.

Table V: Characterization using DL.

Anatomic Site	Object or Task	Network Input	Network Architecture	Data Set (train/test)
Breast	Cancer risk	Mammograms	Pre-trained Alexnet	456 patients LOO

	assessment[192]		followed by SVM	CV
	Cancer risk assessment[193]	Mammograms	Modified AlexNet	14,000/1850 images randomly selected 20 times
	Cancer risk assessment[194]	Mammograms	Custom DCNN	478/183 mammograms
	Cancer risk assessment[195]	Mammograms	Fine-tuned a pre-trained VGG16Net	513/91 women
	Diagnosis[196]	Mammograms	Pre-trained AlexNet followed by SVM	607 cases 5-fold CV
	Diagnosis[197]	Mammograms, MRI, US	Pre-trained VGG19Net followed by SVM	690 MRI, 245 FFDM 1125 US, LOO CV
	Diagnosis[198]	Breast Tomosynthesis	Pre-trained Alexnet followed by evolutionary pruning	2682/89 masses
	Diagnosis[199]	Mammograms	Pre-trained AlexNet	1545/909 masses
	Diagnosis[200]	MRI MIP	Pre-trained VGG19Net followed by SVM	690 cases with 5-fold CV
	Diagnosis[201]	DCE-MRI	LSTM	562/141 cases
	Solitary cyst diagnosis[202]	Mammograms	Modified VGG Net	1,600 lesions 8-fold CV
	Prognosis[203]	Mammograms	VGG16Net followed by logistic regression classifier	79/20 cases randomly selected 100 times
Chest - Lung	Pulmonary Nodule Classification[204]	CT patches	ResNet	665/166 nodules

Tissue Classification[205]	CT patches	Restricted Boltzmann Machines	training 50/100/150/200; testing 20,000/1,000/20,000/20,000 image patches
Interstitial Disease[206]	CT patches	Modified AlexNet	100/20 patients
Interstitial Disease[207]	CT patches	Modified VGG	public: 71/23 scans local: 20/6 scans
Interstitial Disease[208]	CT patches	Custom	480/(120 and 240)
Interstitial Disease[209]	CT patches	Custom	36,106/1,050 patches
Pulmonary Nodule Staging[210]	CT	DFCNet	11/7 patients
Prognosis[211]	CT	Custom	7,983/ (1000 and 2164) subjects
Chest - cardiac	Calcium Scoring[212]	CT	Custom	1181/506 scans
Ventricle Quantification[213]	MR	Custom (CNN + RNN +Bayesian multitask)	145 cases, 5-fold CV
Abdomen	Tissue Classification[214]	Ultrasound	CaffeNet and VGGNet	136/49 Studies
Liver Tumor Classification[215]	Portal Phase 2D CT	GAN	182 cases, 3-fold CV
Liver Fibrosis[216]	DCE-CT	Custom CNN	460/100 scans

	Fatty Liver Disease[217]	US	Invariant Scattering Convolution Network	650 patients, 5- and 10-fold CV
Brain	Survival[218]	Multiparametric MR	Transfer learning as feature extractor, CNN-S	75/37 patients
Skeletal	Maturity[219]	Hand Radiographs	Deep Residual Network	14036/ (200 and 913) exams

Abbreviations: FCN: Fully Convolutional Network. LOO: Leave-one-out. CV: Cross-validation.

Table VI: Image processing and reconstruction with DL.

Task	Imaging Modality	Performance Measure	Network Output	Network Architecture Basis
Filtering	CT[234] Chest X-Ray[235] X-ray fluoro[236]	MSE[234], CAD Performance[234] PSNR[235, 236] SSIM[235, 236] Runtime[236]	Likelihood of Nodule[234] Bone Image[235] CLAHE filtering[236]	Custom CNN[234, 235] Residual CNN[236] Residual AE[236]
Noise reduction	CT[237-240] PET[241]	PSNR[237-241] RMSE[237, 238] SSIM[237, 238, 240] NRMSE[239] NMSE[241]	Noise-reduced image[237-241]	Custom CNN[237-239] Residual AE[237, 238] Concatenated CNNs[241] U-net[240]
Artifact reduction	CT[242, 243] MRI[244]	SNR[242, 243] NMSE[244] Qualitative[243] Runtime[244]	Sparse-view recon[242, 244] Metal artifact reduced image[243]	U-net[242, 244] Custom CNN[243]
Recons	MRI[245-248]	RMSE[245, 248] Runtime[245] MSE[246, 247] NRMSE[246] SSIM[246] SNR[248]	Image of scalar measures[245] MR reconstruction[246-248]	Custom CNN[245, 248] Custom NN[246] Cascade of CNNs[247]
Registration	MRI[249-252] X-ray to	DICE[249, 250] Runtime[250] Target overlap[251] SNR[252]	Deformable registration [249-252]	Custom CNN[249, 251-254] SAE[250]

	$3D^{253, 254}$	TRE^{254} Image & vessel sharpness[252] $mTREproj^{253}$	Rigid body 3D transformation[253, 254]	
Synthesis of one modality from another	CT from MRI[255-259] MRI from PET[260] PET from CT[261]	$MAE^{255, 256}$ $PSNR^{255, 259}$ ME^{256} MSE^{256} Pearson Correl[256] PET Image Quality[257, 258] $SSIM^{260}$ SUVR of MR-less methods[260] Tumor detection by radiologist[261]	Synthetic CT[255-258] Synthetic MRI[260] Synthetic PET[261]	Custom 3D FCN[255] GAN[259-261] U-net[256, 257] AE[258]
Image quality assessment	US[262] CT[263, 264] MRI[265]	$AUC^{262, 264}$ IOU^{262} Correlation between TRE estimation and ground trutth[263] Concordance with readers[265]	ROI localization & classification[262] TRE estimation[263] estimate of image diagnostic value[264, 265]	Custom CNN[262, 265] Custom NN[263] VGG19[264]

Abbreviations: MSE: Mean-squared error, RMSE: Root MSE, NSME: Normalized MSE, NRMSE: Normalized RMSE, SNR: signal-to-noise ratio, PSNR: Peak SNR, SSIM: Structural similarity, DICE: Segmentation overlap index, TRE: Target registration error, mTREproj: mean TRE in projection direction, MAE: Mean absolute error, ME: Mean error, SUVR: Standardized uptake value ratio, AUC: Area under the receiver operating characteristic curve, IOU: Intersection over union, CLAHE: Contrast-limited adaptive histogram equalization.

Table VII: Radiotherapy and assessment of response to treatment with DL.

Anatomic Site	Object or Task	Network Input	Network Architecture	Dataset (train/test)
Bladder	Treatment response assessment[268]	CT	CifarNet	82/41 patients
Brain	Glioblastoma multiforme treatment options and survival prediction[218]	MRI	Custom	75/37 patients
	Assessment of treatment effect in acute ischemic	MRI	CNN based on	158/29 patients

			SegNet	
Breast	Response to neoadjuvant chemotherapy[270]	MRI	Pre-trained VGGNet followed by LDA	561 exams from 64 subjects LOO CV
	Response to neoadjuvant chemotherapy[271]	MRI	Custom	133/33 patients
	Segmentation of clinical target volume[272]	CT	Deep Dilated Residual Network	800 patients 5-fold CV
Cancer cell lines	Prediction of drug effectiveness in cancer cell lines[273]	Multiple omics data from cancer cells (gene expression data, copy number variation data, mutation data, and cell line annotations)	Deep autoencoder	520/104 cell lines
Head and Neck	Organ segmentation[274]	CT	U-Net based with shape retention model	22/10 scans
Kidney	Renal segmentation[275]	CT	Custom	89/24 patients
	Early detection of acute renal transplant rejection[276]	DWI-MRI	Stacked autoencoders	100 patients 4-fold, 10-fold and LOO CV
Liver	Hepatobiliary toxicity prediction after liver SBRT[277]	CT and patient demographics, clinical information	Custom CNN trained on other organs, fine-tuned on liver	125 patients 20-fold CV

				SBRT	
Lung		Estimation of dose protocols in Radiotherapy[278]	FDG-PET/CT, clinical, genetic, imaging radiomics features, tumor and lung dosimetric variables, treatment plans	Deep Q-Network	114 real train / 4000 synthesized test cases
		Dynamic tracking during therapy[279]	DRRs from 4D CT	DenseNet	1/9 volumes
Prostate		Prediction of dose from patient image contours[280]	IMRT	U-Net	80/8 patients
		Prediction of dosimetric eligibility of prostate cancer patients undergoing IMRT[281]	CT	Fine-tuned AlexNet	60 patients 5-fold CV
Pelvis		Generating synthetic CTs from MR-only radiotherapy[282]	MRI	cGAN	123/59 patients
		Assessment of toxicity to normal organs and tissue[283]	Rectum surface dose maps	Fine-tuned VGG-16	42 patients 10-fold and LOO CV
Rectum		Segmentation of rectal tumors on T2-MRI and clinical target volume segmentation on CT[272]	T2-MRI or CT	Novel CNN involving cascaded atrous convolution and spatial pyramid pooling	70 T2-MR and 100 CT 5-fold CV
		Prediction of pathologic	CT	DNN Classifier	95 patients 5-

	complete response after chemoradiation[284]		Custom Estimator	fold CV

Abbreviations: IMRT: Intensity-modulated radiation therapy. SBRT: Stereotactic body radiotherapy. DWI: Diffusion-weighted MRI. DRR: Digitally reconstructed radiographs. LDA: Linear discriminant analysis. LOO: Leave-one-out. CV: Cross-validation.

mp_13264_f2.tif

mp_13264_f3a.tif

Created by Universal Document Converter

mp_13264_f3b.tif

mp_13264_f4.tif

mp_13264_f5.tif

mp_13264_f6.tif

mp_13264_f7.tif