nature portfolio

https://doi.org/10.1038/s41571-021-00560-7

Supplementary information

Predicting cancer outcomes with radiomics and artificial intelligence in radiology

In the format provided by the authors and unedited

Supplementary Box 1 | Other network types of neural networks in radiology AI

Fully convolutional neural networks

Fully convolutional neural networks (FCNs)[92] are a variety of convolutional neural networks (CNN) tailored to providing pixel-level rather than image-level predictions. Whereas the final layers of CNNs generally consist of fully connected layers that transform image representations into a vector of predictions, FCNs contain only convolutional layers and produce outputs in the image space. Accordingly, FCNs are most commonly utilized for segmentation in medical images. The most popular FCN in medical imaging is the U-net model[178], which first reduces then restores the size of an input volume, making use of information at multiple spatial scales. FCN segmentation models are a powerful option for obtaining automated segmentations to enable AI biomarkers that require precise manual annotation of ground truth, such as radiomics models[179].

Convolutional autoencoders, a form of FCN trained to compress image inputs to a low-dimensional representation and then reconstruct the original image, are a commonly utilized unsupervised feature-learning strategy. FCNs can typically be trained using much smaller datasets than CNNs for classification. The resulting feature representations learned by the FCN can then be used as input to train a secondary, simpler predictive model. This strategy enables the training of supervised models from deep feature representations when patient data is limited[180,181].

Neural networks for longitudinal data

Whereas typical deep learning (DL) approaches, such as CNNs, are limited to data collected from a single time point, specialized neural networks have been proposed for detecting temporal patterns in time-series data. Long short-term memory (LSTM)[182] and other recurrent neural networks include ‘memory’ states that can process temporal data sequentially. Pairing CNNs for feature extraction with recurrent networks can enable them to analyse temporal imaging data. This scheme can be applied for prediction from longitudinal imaging exams[183] and dynamic imaging modalities[184]. Recently, transformer networks[185] have vastly improved[186] the analysis of longitudinal information by allowing a network to incorporate information from all time points at once rather than sequentially. Transformer-based strategies do not require consistent time intervals between patients or for data types to be provided in a fixed order[138], and thus offer immense potential to leverage real-world longitudinal imaging and clinical data.

Supplementary Box 2 | Popular convolutional neural network architectures

A number of CNN architectures have been proposed for pattern recognition in image data, and these myriad approaches can prove daunting for clinical researchers to compare and choose from. In this section, we present an overview of some of the CNN architectures most commonly applied to radiology images, the design innovations that have made them popular, and some guidelines on their use. Although we focus on their original incarnations for natural images, the majority of these architectures have been adapted for 3D images with open-source implementations. We also recommend several high-quality review articles for readers interested in learning more about CNN varieties. Morid et al.[187] provide a thorough meta-review of CNN approaches in radiology transfer learning studies. They provide detailed statistics on what architectures are most popular in different imaging modalities and anatomical sites. Khan et al.[188] provide a more thorough account of popular CNNs in computer vision, discussing a larger range of architectures and their contributions in greater technical depth.

Although CNNs were first introduced in the 1980s, the modern era of DL-based computer vision began in the early 2010s as researchers utilizing CNNs began to dominate pattern recognition competitions[189–192]. Most significant of these was AlexNet[193], which achieved unprecedented performance in the 2012 ImageNet Large Scale Visual Recognition Challenge. AlexNet is a relatively simple architecture consisting of 8 layers containing convolution kernels with large receptive fields. The visual geometry group (VGG)[194] family of CNNs is based on the AlexNet architecture, but improved upon its performance by increasing network depth (the architecture has 16-layer and 19-layer configurations, denoted as VGG16 and VGG19, respectively), with each convolutional layer containing uniform, small receptive fields. Although AlexNet and VGG architectures are relatively large in terms of number of parameters, their shallow and straightforward configurations relative to later CNN architectures have nonetheless made them a popular choice for transfer learning in medical imaging[187].

A number of architectures were subsequently proposed that enabled larger and more powerful networks while using considerably fewer parameters. These approaches introduced more complex layer configurations for more efficient learning. The inception family of CNNs (including GoogLeNet[195], Inception-V3[196], and Inception-V4[197]) introduced the inception module, which performs convolutions at multiple receptive field sizes in parallel then concatenates the outputs of each. This design enforces sparsity in network connections and better captures information across multiple spatial scales, resulting in improved performance and greater computational efficiency. Inception is an especially popular choice for transfer learning applications for radiography and MRI. Residual networks (ResNet)[198] achieved similar performance boosts through the introduction of the residual unit, which creates a ‘skip connection’ between several convolutional layers in series that combines deep features output by a convolutional layer with deep features from a few layers earlier in the network by summing them. These skip connections help facilitate network training by maintaining the strength of the signal found in lower levels of the network, which in turn allows networks to be constructed with significantly more layers. Several ResNet variants exist with different numbers of layers, such as ResNet-34, ResNet-50, and ResNet-101. ResNet CNNs have been shown in multiple high-profile studies[88,199] to match human performance in the interpretation of mammograms. Densely connected convolutional neural networks (DenseNet)[200] builds on the principles of ResNet but adds more connections between layers and concatenates features rather than summing them, enabling strong performance with up to 250 layers. Inception-ResNet[197] combines the advantages of these two approaches to achieve an architecture with both high depth and width.

More recently, Google introduced the EfficientNet[201] family of CNNs, which seek to optimize the balance of network depth, width and input resolution, and achieve substantial reductions in model size, while maintaining strong performance. While EfficientNet is relatively new and has thus far seen only a handful applications in cancer imaging[202], its balance of performance and network size are likely to make it an important tool in training DL models from limited radiology data.

Supplementary Box 3 | Strategies for developing AI models with limited training data

Transfer learning

A significant barrier to training supervised deep learning (DL) models for medical applications is limited availability of training data. State-of-the-art convolutional neural networks (CNNs) are designed for high performance when trained with massive natural image datasets and often feature numerous layers of convolutional operations with tens of millions of learnable parameters. As a result, such models can fail when trained within medical imaging datasets many orders of magnitude smaller. Transfer learning is a form of supervised learning where a model is first trained on a larger dataset for a different pattern recognition task, then applied or adapted to a smaller dataset. There are two main transfer learning strategies applied in medical imaging: feature extraction and fine-tuning[187]. In feature extraction, a pretrained CNN can be applied directly to medical images without any additional training to extract a set of features that will be used as input to a simpler downstream machine learning (ML) model. Typically, the final layers of a CNN that output a prediction are removed in these approaches and instead image features from an intermediate layer are used to train the secondary model. In fine-tuning, the previously learned weights are used as a starting point for a trainable model that is incrementally trained for a new application on a smaller dataset. The underlying theory behind this approach is that initializing a model for one pattern recognition task will learn some information that can be applied to new problems. A network can be fine-tuned in entirety, or the learnable weights of earlier layers can be frozen to prevent them from changing. These lower level layers tend to encode the simplest image patterns in CNNs and have the highest likelihood of having utility in multiple pattern recognition tasks. When frozen, they function as fixed feature extraction layers while the remainder of the network is fine-tuned, and thus reduce the number of parameters in the model and the risk of overfitting on small datasets.

While transfer learning can significantly reduce the quantity of training data required, it also introduces its own set of challenges and design constraints. In particular, inputs to a model with pretrained weights must possess similarities to the original image data used to train them. In the context of radiology studies, this most often means processing 3D medical imaging volumes to resemble 3-channel natural RGB images. Previous studies have created artificial RGB images from radiology by inserting adjacent slices[203], multiple scan types[33] or different 2D views into the colour channels[90] in order to repurpose networks pretrained on natural images for outcome prediction.

Expanding datasets with artificial data

A common tactic for training DL models in spite of limited data is the generation of synthetic data to expand the training set. Augmentation is the process of applying perturbations to training images in order to increase the variation seen by a model during training. Popular augmentation transformations include random flipping, rotation, resizing, deformation, histogram shifting, and addition of synthetic scanner artifacts[204]. More recent approaches have sought to apply more sophisticated transformations or even generate entirely new, fully synthetic data using a separate DL model known as a generative adversarial network (GAN)[205]. GANs have been successfully employed to artificially change the scan protocol of training data[206] or generate synthetic lesions for model training[207], which improved performance and generalizability over traditional augmentation strategies.

Similar principles can be applied to radiomic model development under data limitations. For instance, synthetic minority over-sampling technique (SMOTE)[208] is a popularly utilized approach

to generate synthetic radiomic features of a less common class (for instance, responders to a treatment with a low rate of response) in order to synthetically balance training datasets.[209]

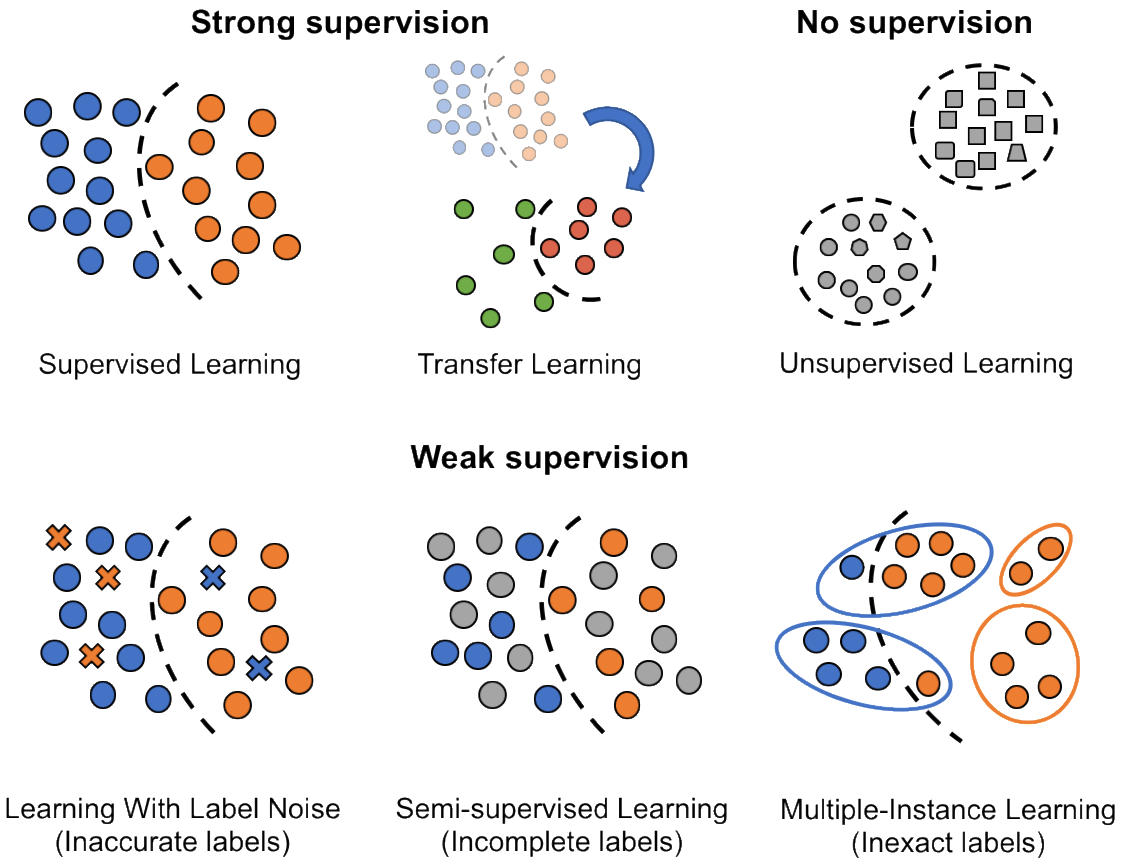
Unsupervised learning

In situations where data is more limited or there is a lack of well-validated clinical labels to rely on for model training, ML practitioners often employ unsupervised learning strategies. In unsupervised learning, no data labels are provided to train the learning model. Instead, an algorithm learns to organize samples and obtain a reduced dimensional representation of data that reflects some inherent hidden structure. A popular variety of unsupervised learning strategy in radiomics is clustering, wherein patients are stratified based on similarities across a high-dimensional set of quantitative image features. One avenue for radiomics-based clustering is the discovery of distinct disease phenotypes based on imaging appearance, which can then be assessed for association with overarching prognostic or genotypic patterns[141,210]. Unsupervised learning strategies can also be employed as a precursor to supervised model training in data-constrained scenarios, such as fully convolutional neural networks (FCNs)[180].

Weakly supervised learning

In many cases, image data might be readily available, but the corresponding clinical labels necessary to train a supervised model are missing or flawed. Weakly supervised learning[94] refers broadly to a set of learning approaches tailored to developing ML models from imperfectly labelled data. Some forms of weakly supervised learning are included in Supplementary Figure 1. In the case where only a portion of data has outcome labels, semi-supervised learning makes use of unlabelled data but making informed assumptions about its group membership. For instance, a model can be initially trained with only labelled data, then its high confidence predictions can be used to assign labels to new data for the training set[211]. Multi-task learning can also make use of partially trained data by jointly training a model for the task of interest (e.g. outcome prediction) and an easier task with greater label availability, such as diagnostic labels.

Supplementary Figure 1 | Types of weakly supervised learning. These types include: inaccurate supervision, utilizing labels that might include errors; incomplete supervision, utilizing a dataset that is only partially labelled; and inexact supervision, utilizing imprecise or course labels.



Supplementary Table 1 | Prognostic and predictive radiomic approaches currently studied across the most common cancer types

Indication	Type/stage	AI methods	Approaches	Novel applications
Lung cancer				
Prognosis	NSCLC: •Early stage •Locally advanced stage •Metastatic cancer SCLC and mesothelioma	*Radiomics:* •Shape, texture, intensity, edge sharpness, vessel tortuosity •Radiomics of the intratumoural area, the tumour microenvironment as well as lymph nodes *DL:* •Mostly CNN, occasionally recurrent networks for longitudinal analysis	•Survival outcomes in both early and advanced-stage NSCLC, which includes DFS, RFS as well as OS	• 3D CNN for prognosticating outcome in patients with NSCLC treated with radiotherapy and/or surgery[98] • FDG-PET radiomics to predict outcome in mesothelioma[97] • Tumour elongation features prognostic of 2-year OS, LRR and DM in SCLC chemotherapy recipients[212]
Response prediction	NSCLC: •Adjuvant chemotherapy in early stage disease •Combination chemotherapy in locally advanced cancers •Trimodality therapy •Targeted therapy •ICI therapy in metastatic cancer SCLC: standard of care chemotherapy		• RECIST response • Major pathological response in locally advanced cancers treated with chemotherapy • Survival differences to differentiate between responders and non-responders to therapy	• A radiomic nomogram[65] to predict benefit from adjuvant chemotherapy following surgery, associated with the spatial arrangement of TILs and biological pathways of cellular differentiation and angiogenesis • Clinical trial-derived CT radiomics signature[136] estimated tumour-infiltrating $CD8^+$ T cells, predicted response to ICIs in different solid tumours • Radiomic biomarker[64] on pre-treatment and immediate post-

				treatment CT scans was used to predict response to checkpoint therapy in terms of RECIST and OS • Radiomics of primary/metastatic lesions in NSCLC and melanoma[137] predicted ICI response, associated with molecular pathways of mitosis and proliferation • A DL transformer network[138] integrated clinical measurements over time, interventions, and CT radiomics to predict anti-PD-1 immunotherapy response • Radiomic methods[139,140] to differentiate atypical response patterns like hyperprogressors to ICI therapy
Breast cancer				
Prognosis	•Early stage node-negative cancers across all the hormone receptor subtypes •Metastatic node-positive cancers •Resistant triple-negative breast cancers •Hormone receptor-positive breast cancers	*Radiomics:* •Intensity, texture, morphology •Peritumoural radiomics and radiomics of axillary lymph nodes has shown to improve model performance *DL:*	•End points studied have included DFS, RFS, DM as well as OS	•Nomogram[99] of radiomics from DCE-MRI and clinic-pathological features predicted DFS in invasive breast cancer •AI analysis on perfusion sequences of breast images[100] to build a spatial interaction matrix to predict RFS •Radiomics of intratumoural and peritumoural regions and axillary lymph nodes to

		• CNNs, and occasionally recurrent networks for temporal analysis of dynamic contrast-enhanced		estimate DFS in invasive breast cancer[101] • A recurrent network[103] predicted RFS from multiple DCE-MRI images throughout neoadjuvant chemotherapy
Response prediction	• Neoadjuvant chemotherapy before surgery in invasive hormone-positive and triple-negative tumours • HER2-targeted therapy in HER2-positive tumours • CDK4/6 inhibitors in hormone positive metastatic breast cancers		• Pathological complete response as the end point of choice for neoadjuvant therapy • Targeted therapy response also quantified in terms of survival including RFS, DFS and OS	• Multiparametric radiomics[121] of contrast-enhanced T1, T2 and DWI MRI predicted pathologic response in 3 invasive breast cancer datasets • Peritumoural radiomics substantially improved neoadjuvant chemotherapy response prediction when combined with tumour features on contrast-enhanced MRI[59] • Radiomic texture features from breast ultrasound[125] predicted NAC response. • In HER2-positive breast cancer, integration of peritumoural and intratumoural radiomics from DCE-MRI[60] discovered intrinsic molecular subtypes with differing immune response and predicting response to HER2-targeted therapy • In hormone receptor-positive metastatic breast cancer treated with CDK4/6 inhibitors,

				radiomic features of liver metastasis on CT predicted response and OS .
Brain cancer				
Prognosis	• Primarily focused on the most aggressive variant of brain tumour, glioblastoma • Patients with glioblastoma receiving surgery followed by concomitant chemoradiotherapy	*Radiomics:* • Size, texture, intensity, volume mass deformation. • Tumour habitat (oedema, necrosis) and enhancing tumour *DL:* • CNNs especially specialized networks encompassing multiple sequences	• Prognostic outcome has included binary outcome like recurrence • Hazards analysis models have been developed in tandem with radiomics to predict time dependent outcome such as PFS, DFS and OS	• Volume, shape and texture from multiparametric MRI using a supervised principal component analysis was used to predict PFS and OS in glioblastoma[104] • A radiomic risk score[68] encompassing texture and entropy features especially from the peritumoural area could predict PFS in glioblastoma • A DL-based automatic neural network[106] on multiparametric MRI was validated on 532 patients from a completed clinical trial to predict survival, outperforming standard of care response criteria by over 50% • A novel four-input CNN using multiparametric MRI images projected along 3 spatial dimensions to form RGB images was stratified patients with glioblastoma on the basis of OS
Prostate cancer				

Prognosis	• Risk of recurrence following definitive surgery • Differentiating between aggressive and indolent tumours prostate cancers	*Radiomics:* • Not only texture, size but also novel shape features • Deep learning: • Made use of novel transfer learning approaches	• Biochemical recurrence after radical prostatectomy • Identifying aggressive tumours from those with good prognosis	• Radiomic textural heterogeneity and gradient orientation from T2-weighted MRI and ADC maps[107] predicted biochemical recurrence after radical prostatectomy • Hand-crafted radiomic features from multiparametric MRI, clinico-pathologic, and subjective imaging features predicted 3-year biochemical recurrence following radical prostatectomy[108] • DL transfer learning differentiated indolent and clinically significant prostate cancer on multiparametric MRI and outperformed standard of care PI-RADS scoring[109]

ADC, apparent diffusion coefficient; AI, artificial intelligence; CNN, convolutional neural network; DCE-MRI, dynamic contrast-enhanced MRI; DFS, disease-free survival; DL, deep learning; DM, distant metastases; DWI, diffusion-weighted imaging; FDG-PET, 2-deoxy-2-[^{18} F]fluoro-D-glucose PET; ICI, immune checkpoint inhibitor; LRR, locoregional recurrence; NSCLC, non-small-cell lung cancer; OS, overall survival; PFS, progression-free survival; PI-RADS, prostate imaging reporting and data system; RECIST, response evaluation criteria in solid tumors; RFS, recurrence-free survival; RGB, red green blue; SCLC, small-cell lung cancer; TILs, tumour-infiltrating lymphocytes.

Supplementary Table 2 | Overview of popular CNN architectures

Model	Year	Number of parameters	Number of layers	Contribution
LeNet	1998	0.06M	5	Proposed CNNs trained with back-propagation
AlexNet	2012	61M	8	Achieved superhuman performance in 2012 ImageNet challenge, launching modern age of computer vision. Popularized ReLU activation, multi-GPU training
VGG	2014	138M 143M	16 19	Increased the number of layers (depth) of CNNs over previous work.
GoogLeNet, Inception-V3, Inception-V4	2014, 2015, 2016	4M, 23.6M, 35M	22, 159, 70	Introduced inception unit, applying convolutional layers at multiple spatial scales in parallel. Enabled more powerful image classification and increased model depth with a fraction of the model parameters.
ResNet	2016	11M 22M 24M 43M 59M	18 34 50 101 152	Introduced the residual unit, which includes “skip connections” between several convolutional layers in series that combines deep features output by a convolutional layer with deep features from a few layers earlier in the network.
Inception-ResNet	2016	56M	572	Combined wide, multi-scale convolutions (inception block) with skip connections (residual unit)
DenseNet	2017	15.3M, 25.6M	190, 250	Extended the principles of ResNet by introducing dense blocks with skip connections between all layers.
SqueezeNet	2017	1.2M	18	Designed as a more compact replacement for AlexNet
EfficientNet (B0-B7)	2019	5.3–66M	18–36	Family of CNNs designed for increased efficiency by more balancing network depth, width, and resolution

CNN, convolutional neural network; GPU, graphics processing unit; M, million; ReLU, rectified linear unit; VGG, visual geometry group.

Supplementary references

178. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (eds. Navab, N., Hornegger, J., Wells, W. M. & Frangi, A. F.) 234–241 (Springer International Publishing, 2015). doi:10.1007/978-3-319-24574-4_28.

179. Baid, U. *et al.* Deep Learning Radiomics Algorithm for Gliomas (DRAG) Model: A Novel Approach Using 3D UNET Based Deep Convolutional Neural Network for Predicting Survival in Gliomas. in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries* (eds. Crimi, A. et al.) 369–379 (Springer International Publishing, 2019). doi:10.1007/978-3-030-11726-9_33.

180. Baldi, P. Autoencoders, Unsupervised Learning, and Deep Architectures. 14.

181. Zhu, Y. *et al.* MRI-based prostate cancer detection with high-level representation and hierarchical classification. *Medical Physics* **44**, 1028–1039 (2017).

182. Hochreiter, S. & Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **9**, 1735–1780 (1997).

183. Xu, Y. *et al.* Deep Learning Predicts Lung Cancer Treatment Response from Serial Medical Imaging. *Clin Cancer Res* **25**, 3266–3275 (2019).

184. Antropova, N., Huynh, B., Li, H. & Giger, M. L. Breast lesion classification based on dynamic contrast-enhanced magnetic resonance images sequences with long short-term memory networks. *J Med Imaging (Bellingham)* **6**, (2019).

185. Vaswani, A. *et al.* Attention Is All You Need. *arXiv:1706.03762 [cs]* (2017).

186. Wolf, T. *et al.* Transformers: State-of-the-Art Natural Language Processing. in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations* 38–45 (Association for Computational Linguistics, 2020). doi:10.18653/v1/2020.emnlp-demos.6.

187. Morid, M. A., Borjali, A. & Del Fiol, G. A scoping review of transfer learning research on medical image analysis using ImageNet. *Computers in Biology and Medicine* **128**, 104115 (2021).

188. Khan, A., Sohail, A., Zahoora, U. & Qureshi, A. S. A Survey of the Recent Architectures of Deep Convolutional Neural Networks. *Artif Intell Rev* **53**, 5455–5516 (2020).

189. Ciregan, D., Meier, U. & Schmidhuber, J. Multi-column deep neural networks for image classification. in *2012 IEEE Conference on Computer Vision and Pattern Recognition* 3642–3649 (2012). doi:10.1109/CVPR.2012.6248110.

190. Ciresan, D. C., Meier, U., Gambardella, L. M. & Schmidhuber, J. Deep Big Simple Neural Nets Excel on Handwritten Digit Recognition. *Neural Computation* **22**, 3207–3220 (2010).

191. Ciresan, D. C., Meier, U., Masci, J., Gambardella, L. M. & Schmidhuber, J. Flexible, High Performance Convolutional Neural Networks for Image Classification. 6.

192. Cireşan, D., Meier, U., Masci, J. & Schmidhuber, J. Multi-column deep neural network for traffic sign classification. *Neural Networks* **32**, 333–338 (2012).

193. Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet Classification with Deep Convolutional Neural Networks. in *Advances in Neural Information Processing Systems 25* (eds. Pereira, F., Burges, C. J. C., Bottou, L. & Weinberger, K. Q.) 1097–1105 (Curran Associates, Inc., 2012).

194. Simonyan, K. & Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv:1409.1556 [cs]* (2015).

195. Szegedy, C. *et al.* Going Deeper with Convolutions. *arXiv:1409.4842 [cs]* (2014).

196. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. Rethinking the Inception Architecture for Computer Vision. *arXiv:1512.00567 [cs]* (2015).

197. Szegedy, C., Ioffe, S., Vanhoucke, V. & Alemi, A. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. *arXiv:1602.07261 [cs]* (2016).

198. He, K., Zhang, X., Ren, S. & Sun, J. Deep Residual Learning for Image Recognition. *arXiv:1512.03385 [cs]* (2015).

199. Lehman, C. D. *et al.* Mammographic Breast Density Assessment Using Deep Learning: Clinical Implementation. *Radiology* **290**, 52–58 (2018).

200. Huang, G., Liu, Z., van der Maaten, L. & Weinberger, K. Q. Densely Connected Convolutional Networks. *arXiv:1608.06993 [cs]* (2018).

201. Tan, M. & Le, Q. V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *arXiv:1905.11946 [cs, stat]* (2020).

202. Suh, Y. J., Jung, J. & Cho, B.-J. Automated Breast Cancer Detection in Digital Mammograms of Various Densities via Deep Learning. *Journal of Personalized Medicine* **10**, 211 (2020).

203. Peng, J. *et al.* Residual convolutional neural network for predicting response of transarterial chemoembolization in hepatocellular carcinoma from CT imaging. *Eur Radiol* **30**, 413–424 (2020).

204. Hao, R., Namdar, K., Liu, L., Haider, M. A. & Khalvati, F. A Comprehensive Study of Data Augmentation Strategies for Prostate Cancer Detection in Diffusion-weighted MRI using Convolutional Neural Networks. *arXiv:2006.01693 [cs, eess, q-bio, stat]* (2020).

205. Goodfellow, I. *et al.* Generative Adversarial Nets. in *Advances in Neural Information Processing Systems 27* (eds. Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N. D. & Weinberger, K. Q.) 2672–2680 (Curran Associates, Inc., 2014).

206. Sandfort, V., Yan, K., Pickhardt, P. J. & Summers, R. M. Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. *Sci Rep* **9**, 16884 (2019).

207. Frid-Adar, M. *et al.* GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing* **321**, 321–331 (2018).

208. Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. SMOTE: Synthetic Minority Over-sampling Technique. *jair* **16**, 321–357 (2002).

209. Zhang, Y., Oikonomou, A., Wong, A., Haider, M. A. & Khalvati, F. Radiomics-based Prognosis Analysis for Non-Small Cell Lung Cancer. *Sci Rep* **7**, (2017).

210. Chitalia, R. *et al.* Imaging phenotypes of breast cancer heterogeneity in pre-operative breast Dynamic Contrast Enhanced Magnetic Resonance Imaging (DCE-MRI) scans predict 10-year recurrence. *Clin Cancer Res* (2019) doi:10.1158/1078-0432.CCR-18-4067.

211. Azizi, S. *et al.* Big Self-Supervised Models Advance Medical Image Classification. (2021).

212. Kamran, S. C. *et al.* CT-Based Radiomic Biomarker Features Predict Prognosis in Patients with Limited Stage Small Cell Lung Cancer. *International Journal of Radiation Oncology • Biology • Physics* **99**, S12–S13 (2017).