

<https://doi.org/10.1038/s44303-025-00076-0>

# A critical assessment of artificial intelligence in magnetic resonance imaging of cancer

Check for updates

Chengyue Wu<sup>1,2,3,4,5</sup>, Meryem Abbad Andaloussi<sup>6</sup>, David A. Hormuth II<sup>5,7</sup>, Ernesto A. B. F. Lima<sup>5,8</sup>, Guillermo Lorenzo<sup>5,9</sup>, Casey E. Stowers<sup>5</sup>, Sriram Ravula<sup>10</sup>, Brett Levac<sup>10</sup>, Alexandros G. Dimakis<sup>10</sup>, Jonathan I. Tamir<sup>5,10,11</sup>, Kristy K. Brock<sup>1,4,12</sup>, Caroline Chung<sup>4,13,14</sup> & Thomas E. Yankeelov<sup>1,5,7,11,15,16</sup> ✉

Given the enormous output and pace of development of artificial intelligence (AI) methods in medical imaging, it can be challenging to identify the true success stories to determine the state-of-the-art of the field. This report seeks to provide the magnetic resonance imaging (MRI) community with an initial guide into the major areas in which the methods of AI are contributing to MRI in oncology. After a general introduction to artificial intelligence, we proceed to discuss the successes and current limitations of AI in MRI when used for image acquisition, reconstruction, registration, and segmentation, as well as its utility for assisting in diagnostic and prognostic settings. Within each section, we attempt to present a balanced summary by first presenting common techniques, state of readiness, current clinical needs, and barriers to practical deployment in the clinical setting. We conclude by presenting areas in which new advances must be realized to address questions regarding generalizability, quality assurance and control, and uncertainty quantification when applying MRI to cancer to maintain patient safety and practical utility.

We seek to provide a critical assessment of the opportunities and limitations of artificial intelligence (AI) in the context of magnetic resonance imaging (MRI) of cancer and provide the reader with guidance on where to go for further information. Indeed, there have been many excellent publications on various aspects of AI in MRI (see, for example, references<sup>1–4</sup>). However, given the explosion of publications on the topic, it can be challenging to separate the hype from the actual successes and identify the current state-of-the-art of the field. In particular, it is difficult to identify what is currently available—and what the practical limitations are—for the application of AI in the MRI of cancer. In this contribution, we seek to respectfully provide a more balanced presentation of the utility and limitations of AI for MRI within clinical oncology. To address this problem, we begin by describing the key components of AI that are frequently employed in medical imaging, in general, and MRI, in particular. In Section 3, MRI acquisition and reconstruction is presented as a regression task that involves estimating the voxel values of an image from raw scanner measurements. In Section 4, we discuss the classification task of image segmentation, which entails categorizing every voxel as (for example) tumor or healthy tissue, and registration, which entails spatially aligning images to a common space. Sections 5 and 6 concern the classification tasks of making a diagnosis and prognosis from the MRI data, respectively, which require accounting for global characteristics (e.g., the location and size of tumors) to make a prediction.

## Key AI concepts for medical imaging What is artificial intelligence?

Artificial intelligence (AI) refers to the theory and development of computer systems that can perform tasks normally thought to require human intelligence including language, visual perception, and reasoning<sup>5</sup>. We focus on a branch of AI called machine learning (ML), which refers to training a statistical model on relevant data to perform a task. Deep learning is a sub-branch of ML that concerns the development of neural networks (NN), a special class of models that have demonstrated practical utility in medical imaging<sup>6–8</sup>.

## Common ML techniques for medical imaging

A key design consideration when developing ML models is the type(s) of data available for training, which can be split into supervised and unsupervised settings. Supervised datasets contain matched {sample, label} pairs; for example, MRI images annotated as containing or not containing a tumor. In this setting, the model learns to predict the label for each sample. However, annotation can be time-consuming and expensive as domain experts (e.g., radiologists) must label each individual data point<sup>9</sup>. In contrast, unsupervised datasets contain only samples, without any labels. For example, given a set of MR training images obtained with low spatial resolution, one may train a model to learn to increase the resolution (called super-resolution<sup>10,11</sup>).

A full list of affiliations appears at the end of the paper. ✉e-mail: [tey@utexas.edu](mailto:tey@utexas.edu)

without access to any high-resolution examples. After training, the model can be applied to enhance the quality of new images acquired at lower resolution, thus reducing the need for acquiring high-resolution data. While the unsupervised setting eases the burden of annotating data, the ability to generate useful labels is intimately connected to the ability to collect high-quality data that are relevant to the intended application. For example, if only images of knees are available to train a super-resolution model, the model is likely to perform poorly when used to enhance images of brains<sup>12</sup>.

The reader may find it helpful to refer to Table 1 as they consider the rest of the paper.

## AI-based image acquisition and reconstruction of MRI data

Image acquisition and reconstruction are inextricably linked in MRI. Image acquisition consists of both signal and spatial encoding through user-controllable radiofrequency and gradient waveforms. By modifying the different components of the acquisition process (pulse time delay, signal frequency, signal strength, signal phase, etc.), the speed of acquisition as well as the sensitivity of a given image to a particular tissue property is changed. The resulting image contrast can be described by:

$$x = f(\theta, q) \quad (1)$$

where  $x \in C^n$  is the (vectorized) image (containing  $n$  voxels) generated based on a spatiotemporal function  $f$  of both acquisition-controlled signal encoding parameters,  $\theta$ , and biophysical tissue parameters,  $q$ <sup>13</sup>. The acquired measurements (called *k-space*) are related to the image through the linear operator

$$y = A_\phi x + \eta \quad (2)$$

where  $A_\phi \in C^{m \times n}$  represents a (possibly multi-coil) sampled Fourier transform,  $\eta \in C^m$  is additive complex-valued Gaussian noise, and  $m$  is the number of acquired measurements with locations determined by the spatial encoding parameters  $\phi$ .

The measured raw signals (i.e.,  $y$ ) are not immediately ready for visualization as they represent Fourier components of the image. Therefore, optimization algorithms are used to reconstruct the image. The reconstruction algorithm can be represented by a function  $g(y)$  which either inverts the linear measurement process in Eq. (2) to estimate the image  $x$ , or inverts the non-linear measurement process in Eq. (1) to estimate the tissue parameters,  $q$ . These tasks are often challenging as the data are corrupted by additive noise and often subsampled ( $m \ll n$ ) to reduce scan time, thereby leading to an ill-posed inverse problem.

## Common machine learning techniques in image acquisition and reconstruction

Training machine learning models to assist in both acquisition and reconstruction procedures is often framed as an optimization problem in which we train a parameterized function  $f_w(\cdot)$ , often in the form of a deep neural network, to complete a specific task. In the supervised case, the function is optimized with respect to a training set of images (or scan parameters, or tissue parameters) given input  $k$ -space. This often takes the form of minimizing some loss function  $D(\cdot, \cdot)$  via large-scale optimization solvers in the following way:

$$w^* = \operatorname{argmin}_w \sum_{i=1}^N D(f_w(y_i), x_i) \quad (3)$$

where  $x_i$  and  $y_i$  represent the  $i$ th training sample and  $N$  is the size of the training set. In unsupervised learning, the same task is solved but with a minimization that only depends on the inputs  $y_i$ .

**Image acquisition.** The selection of signal encoding parameters  $\theta$  and spatial encoding parameters  $\phi$  can influence both the contrast of the

resulting image (Eq. (1)) and acquisition time (Eq. (2)). As image sensitivity to pathology is affected by acquisition parameters, it makes sense to optimize the image acquisition pipeline to increase sensitivity to relevant tissue contrast. This means finding the optimal radiofrequency flip angles, phases, timings, etc. for a desired contrast. Supervised learning has been used to find these acquisition strategies for optimizing contrast sensitivity<sup>14,15</sup> as well as optimizing the  $k$ -space measurements<sup>16–19</sup>. Here  $w$  in Eq. (3) is replaced by either or both of contrast ( $\theta$ ) and  $k$ -space locations ( $\phi$ ). For example,  $k$ -space trajectory optimization can be used to accelerate imaging by reducing the number of acquired  $k$ -space measurements for a given target spatial resolution. In this case, the loss function will take the form

$$\phi^* = \operatorname{argmin}_\phi \sum_{i=1}^N D(f_w(y_i(\phi)), x_i) \quad (4)$$

where  $y_i(\phi)$  represents a particular sampling trajectory in  $k$ -space (for example, specific phase encode lines) for the  $i$ th training sample, and  $f_w(y_i(\phi))$  represents a reconstruction algorithm that takes the subsampled  $k$ -space and outputs an image. When the reconstruction network and the sampling trajectory are both differentiable, this is straightforward to implement; however, for Cartesian sampling, this problem is combinatorial in nature. Therefore, greedy approaches or smooth approximations become necessary<sup>16–19</sup>. A similar approach can be used to update other scan parameters that influence image contrast<sup>15</sup>.

After Eq. (4) is solved, an optimized acquisition scheme can be represented as a new measurement model  $A_\phi$  and can be implemented on the scanner. When this sequence is used for acquisition, the image reconstruction problem can then be solved given the acquired measurements and the measurement model (i.e., Eq. (2)).

**Image reconstruction.** Classical image reconstruction relies on hand-crafted priors like  $L_1$ -wavelet regularization<sup>20</sup>. More recently, deep neural networks have contributed promising techniques for image reconstruction. AI-based techniques can be separated into several categories: end-to-end supervised, end-to-end unsupervised, and generative modeling. In the end-to-end supervised setting, the function  $f_w$  in Eq. (3) is trained with pairs of subsampled  $k$ -space data and images obtained from fully sampled data,  $(y_i, x_i)$ , and  $D(\cdot, \cdot)$  can represent any valid metric of distance (e.g., mean squared error). End-to-end unsupervised methods also train an NN except with access only to subsampled measurements,  $y_i$ , and no accompanying reference image,  $x_i$ . This is important in scenarios where it is only possible to collect large amounts of subsampled data (e.g., in dynamic imaging). End-to-end NNs for supervised and unsupervised reconstruction can take many forms, but unrolled optimization networks, which alternate between classical optimization steps (e.g., conjugate gradient descent) and forward passes through a NN, are quite common<sup>21,22</sup>.

Although end-to-end methods are powerful, performance can degrade with changes in how the measurements are taken at training versus at test time, which is referred to as a “test-time distribution shift”. Recently, the use of generative models for inverse problems has been shown to improve robustness to variations in acquisition schemes. These methods train a Bayesian prior on the distribution of fully sampled images,  $p(x)$ , and are therefore agnostic to distribution shifts in the likelihood,  $p(y|x)$ , which can change across scans and imaging protocols (i.e., the signal and spatial encoding parameters). Perhaps the most popular method for including generative priors in inverse problems is by way of score-based, or diffusion probabilistic models<sup>23,24</sup>.

Despite the distinction between supervised and unsupervised training, it is important to note that AI models are trained on data derived from raw sensors (i.e., the raw  $k$ -space measurements), which are typically acquired via multi-coil arrays and are often subsampled. Even when the data are fully sampled, they are noisy and must be first reconstructed into images. Therefore, there is no real notion of “ground-truth” images. New training

Table 1 | A taxonomy of machine learning techniques for medical imaging

Application	Dataset	Task	Notes	References
Acquisition (Section 3.1.1)	Supervised Unsupervised	Regression	Find the scan parameters that produce the best image after reconstruction. Dataset can be supervised (e.g., (parameters, image) pairs) or unsupervised (e.g., images with corresponding labels made by simulating acquisition scheme).	10–15
Reconstruction (Section 3.1.2)	Supervised Unsupervised	Regression	Recover human-interpretable image from raw scanner measurements. Supervised datasets (e.g., (measurement, image) pairs) can be used to train a model to directly map measurements to images. Unsupervised datasets can be used, for example, to train a generative model for MRI, then plug the model into a reconstruction algorithm.	17–20
Segmentation (Section 4.1.1)	Supervised	Fine Classification	Classify every voxel (for 3D data) or pixel (for 2D data) of an image.	28–35
Registration (Section 4.1.2)	Supervised Unsupervised	Regression	Align similar anatomical features between different images by transforming (e.g., shifting and rotating) pixels. Data with (image, transformation) pairs allow for simple supervised training. For unsupervised data with only images, we can train a model, for example, by optimizing similarity metrics between images.	36–41
Diagnosis (Section 5.1)	Supervised	Coarse Classification	Identify the type and/or presence of cancer (and optionally the prediction confidence) in an image.	29,60–67
Response/Prognosis Prediction (Section 6.1)	Supervised	Regression Coarse Classification	Predict a response or prognosis from imaging and other patient data. Data with (MRI data, we can regress specific numerical information (e.g., survival rate) or broadly categorize (e.g., as short-, medium-, or long-term survival)).	92–109

methods that account for the acquisition process have been proposed, both for end-to-end methods that are trained to take in acquired measurements and output reconstructed images, as well as for generative models (where the AI model is used as a statistical image prior to an iterative reconstruction). It is also possible to jointly train the acquisition parameters together with the reconstruction method, which may consist of deep neural networks<sup>14–17</sup>. For example, Aggarwal et al., formulate the optimization problem as

$$\phi^* w^* = \operatorname{argmin}_{\phi, w} \sum_{i=1}^N D(f_w(y_i(\phi)), x_i)$$

(5)

In other words, they jointly solve for the phase encode lines given by  $\phi$  and the neural network reconstruction given by  $f_w$ .

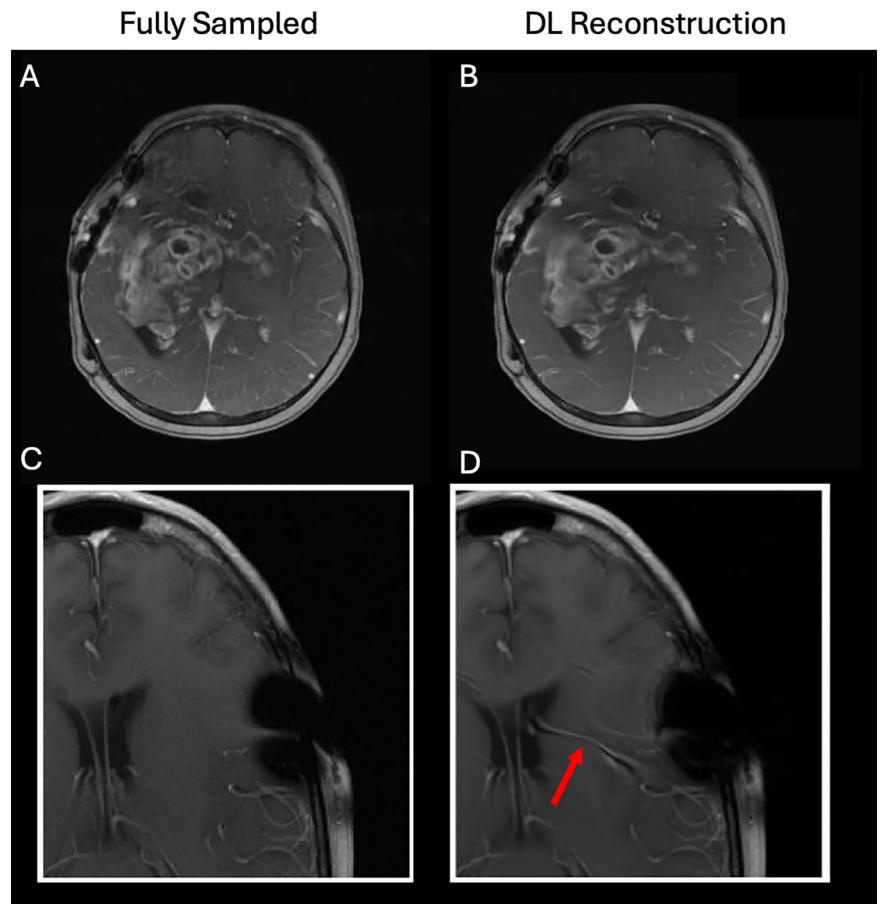
Current clinical needs for AI-based image acquisition and reconstruction

AI has shown the ability to maintain high image quality using fewer measurements than conventional reconstruction techniques, and can even produce diagnostic images for use in specific downstream clinical tasks<sup>25</sup>. For example, AI for the reconstruction of prospectively accelerated abdominal imaging produced non-inferior images compared to conventional reconstruction methods<sup>26</sup>. AI for image reconstruction has made its way into products for several scanner manufacturers. Two recent examples are GE’s AIR™ reconstruction protocol and Siemens’ Deep Resolve™ product line. However, it is important to emphasize that many of these products are for specialized cases (e.g., specific anatomy/contrast) and are not currently deployed for many specific use cases. An ability to generalize across imaging protocols is critical for widespread adoption. An additional concern for many longer scan sessions is patient motion, especially for pediatric populations. In general, this includes developing robust reconstruction methods that can handle dynamic imaging scenarios (e.g., contrast-enhanced imaging<sup>27</sup>) where it is often infeasible to collect the desired measurements from a single imaging volume. As many forms of pathology exhibit only subtle contrast changes, there is also a need for acquisition/reconstruction methods capable of resolving subtle variations in tissue properties, and thus reducing the use and dose of exogenous contrast agents containing Gadolinium<sup>28</sup>. Finally, as many new methods often train and measure performance based only on image quality metrics, the need for better quantifying the link between image quality and downstream diagnostic metrics (e.g., tumor classification) must be better understood and optimized for when designing new AI-based techniques.

Barriers to practical deployment in the clinical setting

Although AI has displayed marked improvements over classical reconstruction and acquisition techniques<sup>17,21–23</sup>, there are still many issues that must be addressed before clinical adoption can be pursued. For example, AI techniques need to be robust to variations in how data are collected due to differences in vendor, field strength, field inhomogeneity, and patient motion. Models must be trained on representative data for each of the variations in hardware and imaging protocols; if the scan protocol changes (or if the scan parameters change to accommodate a specific patient such as a larger field of view, different resolution, different echo time, larger fat saturation bands, etc.), then this “new” protocol could be out of the distribution that was included in the training set, and the subsequent performance will degrade as seen in Jalal et al.<sup>23</sup>. This can quickly devolve into a “model soup.” For example, the winning team in the FastMRI 2020 reconstruction challenge trained eight independent models, to account for field strength, scan anatomy, etc<sup>1</sup>. The interpretability must also be improved through uncertainty quantification. To demonstrate these points, Fig. 1 shows two example deep learning reconstructions (reproduced from the FastMRI 2020 reconstruction challenge<sup>29</sup>) where fully sampled scans were retrospectively subsampled to simulate faster scanning. In the first case, the DL reconstruction results in a faithful, and diagnostic image; however, in the second case, the DL reconstruction falsely hallucinates a blood vessel,

**Fig. 1 | Example reconstructions from the 2020 FastMRI Challenge, adapted from ref. 29.** The fully sampled images (A, C) were retrospectively subsampled to simulate 8× (top) and 4× (bottom) faster scans. In the top case, the DL reconstruction (B) is able to reproduce with high fidelity the lesion in the post-contrast  $T_1$ -weighted image, though with some blurring. In the bottom case, the DL reconstruction (D) hallucinated a false vessel (red arrow), perhaps due to the surgical staple artifact not being well-represented in the training set.



likely due to the unseen artifact caused by surgical staples. This demonstrates the enormous challenge, and potentially clinically confounding problems, of deploying these models in clinical settings<sup>30</sup>.

A potential avenue to addressing these problems is in the use of generative models for reconstruction: generative models have been shown to be less sensitive to changes in scan protocols and anatomy, largely due to the decoupling between the physical forward model and the statistical image prior. In other words, generative models are trained to learn the distribution of MR images independent of the physical imaging system parameters that produced those images. This means that the generative model can be used to reconstruct images from MR data acquired with other imaging schemes, so long as the images follow the model's learned distribution and the imaging parameters are known. Learning optimal sampling patterns for generative models could also be helpful as this would allow the acquisition to be more tailored to the specific imaging protocol, while still benefiting from the robustness offered by generative model-based reconstruction.

### Registering and segmenting MRI data via AI Common AI techniques in segmentation and registration

**Tumor and tissue segmentation.** Tumor and tissue segmentations are often used for surgical planning, radiotherapy design, and assessing treatment response<sup>31,32</sup>. Currently, manual or semi-automated segmentations often suffer from being resource-intensive and exhibiting unacceptable inter-observer variability<sup>31,33</sup>. While automated segmentation approaches enabled by AI may reduce both of these concerns, they require extensive training sets with expert-defined segmentations, and (due to its black-box nature) may fail without warning<sup>31,34</sup>. Given these limitations, AI-based approaches require human oversight to review and, if necessary, edit<sup>31,35</sup>. Many AI-automated approaches are based on voxel-based features and typically employ convolutional neural networks

(CNN; such as U-Nets) to identify healthy tissue (e.g., organs-at-risk for radiotherapy planning), cancer, and intra-tumoral sub-regions (e.g., necrosis, edema, enhancing lesions)<sup>32</sup>. Pal et al. introduced a method combining fuzzy c-means clustering and random forest algorithms and showed 99% accuracy for segmenting spine tumors<sup>36</sup>. Kundal et al. evaluated four CNN-Based methods for brain tumor segmentation. (CaPTk, 2DVNet, EnsembleUNets and ResNet50) EnsembleUNets achieved Dice scores of 0.93 and 0.85 and Hausdorff Distances of 18 and 17.5 for testing and validation, respectively<sup>37</sup>. While current approaches do not eliminate the need for human refinement, they can often provide an acceptable segmentation of organs at risk that can be manually refined when greater precision and accuracy are required<sup>38</sup>.

**Image registration.** Classical image registration methods consist of estimating the optimal transformation that maximizes the similarity between a set of images and the “target” image. Multiple combinations of transformations (rigid or deformable) and cost functions (sum square differences, normalized cross-correlation, or mutual information) exist. Osman et al. introduced a deformable CNN-based registration of 3D MRI scans for glioma patients called ConvUNet-DIR<sup>39</sup> that outperformed the Voxelmorph method with a mean dice score of 0.975 and similarity index of 0.908 compared to 0.969 and 0.893. However, each method has specific drawbacks such as being sensitive to artifacts or being limited in capturing local differences<sup>40</sup>. Learning-based methods, both supervised and unsupervised, attempt to overcome these difficulties. Supervised learning models such as BIRNet<sup>41</sup> and DeepFLASH<sup>42</sup> are trained using paired images (i.e., registered and un-registered images) with their corresponding transformation. Even though such methods can achieve accurate registration performance, finding the necessary training data (which includes before/after registered images and the associated



transformations) is difficult. Thus, unsupervised deep learning-based approaches are preferred<sup>43,44</sup>, traditionally focusing on optimizing similarity metrics and model architectures using gradient-based optimizers like stochastic gradient descent.

### Current clinical needs for AI-based image segmentation and registration

We assign a moderate state of readiness for DL-based registration of MRIs of brain cancer. In particular, locally registering voxels at the tumor boundary is challenging. Estienne et al.<sup>45</sup> introduced a joint 3D-CNN method for brain registration and tumor segmentation. The model was trained using the BraTS 2018 dataset<sup>46</sup> and then tested on over 200 pairs from the same dataset. The registration method was evaluated by computing the average distance between two ratios: (1) original to deformed tumor mask area, and (2) between brain volumes of the paired images. The method outperformed a previously established approach using whole tumor masks (part of the VoxelMorph package<sup>47</sup>) that employed an unsupervised learning-based inference algorithm that was initially applied to healthy brain MRIs<sup>48</sup>.

There are several areas where DL-based segmentation can meet clinical needs. First, segmenting a tumor from surrounding healthy tissues is central to radiotherapy planning. In fact, the time-intensive nature of manual segmentation has limited the widespread implementation of adaptive radiotherapy<sup>49</sup>. To address this challenge, AI tools employing CNNs to perform semi-automated and automated segmentation have been developed. Such tools often have reduced performance after treatment (i.e., surgery, radiotherapy, and chemotherapy), an increased focus should be placed on training these methods with post-treatment data<sup>50</sup> to enable accurate longitudinal segmentation of tumors. Clinical deployment of these automated segmentation tools seeks to accelerate the segmentation task within the workflow of adaptive radiotherapy. However, these algorithms have had mixed success due to their limited generalizability when applied to data they did not see during training, therefore limiting their widespread adoption. There are numerous potential reasons for the lack of success of AI-based segmentation algorithms including variability in the scan acquisition, patient-specific presentation of normal tissue and tumors, and the segmentation goals of the treating clinical team.

Tumor segmentation and registration are also necessary steps for quantitatively describing imaging features that represent the underlying biology (Fig. 2). The inconsistencies in MRI segmentation can arise from various factors including inter-rater variability among segmentations done by different readers and intra-rater variability among segmentations done by the same reader multiple times. Additional sources of inconsistency include variation in input data quality and resolution due to acquisition protocols, algorithmic bias across different segmentation methods and/or parameter settings, temporal changes in anatomical structure, or intensity

contrast of segmentation targets due to disease progress or intervention effects<sup>51–54</sup>. These inconsistencies introduce bias in the measurement of volume, morphology, and contrast of segmented regions, and influence the accurate and reproducible extraction of (for example) radiomic features<sup>55–58</sup>, which then restricts the reliability and generalizability of radiomics' clinical application in diagnosis and prognosis<sup>59,60</sup>. Therefore, it is critical to develop approaches to improve the consistency of segmentation to advance quantitative imaging.

### Shortcomings of current registration and segmentation approaches

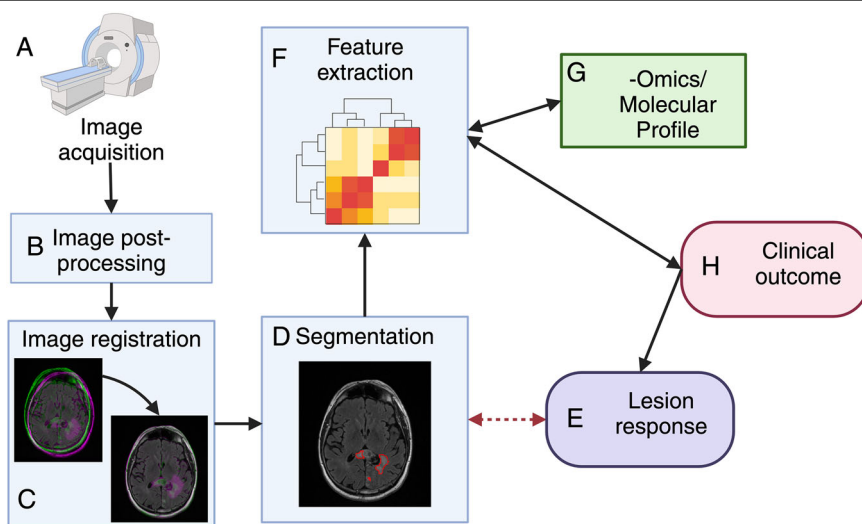
A fundamental shortcoming of current AI-based segmentation and registration approaches is limitations in the available data itself for training and validating these approaches. These limitations include scarcity of data in the target setting, variations in data quality, standardization of imaging protocols, or deviation of imaging protocols from those used to train previously validated approaches. For example in the setting of brain cancer, while there have historically been significant efforts to segment pre-operative tumors, a recent literature review of 180 articles observed only three papers that included post-operative imaging<sup>31</sup>. The lack of trained models or expert-segmented or curated data limits<sup>61</sup> the transferability of registration and segmentation approaches to new applications where treatment may alter image contrast or introduce large deformations and alterations in the anatomy. Another pitfall is the phenomena of data drift where changes in imaging protocols or other factors may lead to input data that falls outside of the distribution used for training<sup>62</sup> resulting in reduced accuracy of segmentations and registration.

### AI for diagnosis from MRI data

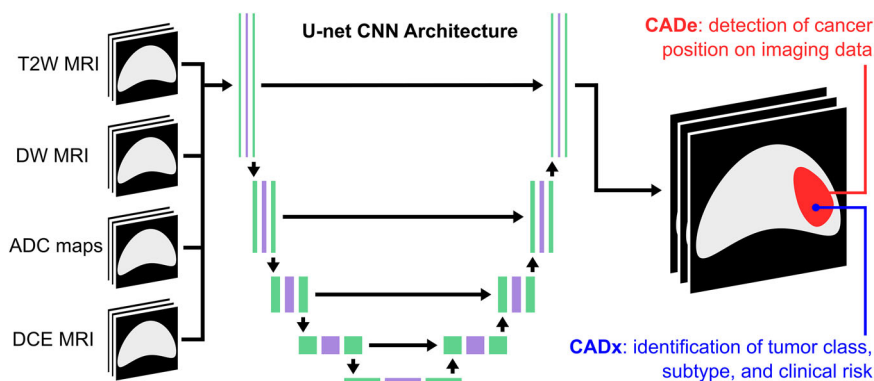
#### Common AI techniques for diagnosis via MRI

While computer-aided detection systems (CADE) leverage AI to identify the position of the tumor within an image, computer-aided diagnosis (CADx) systems exploit imaging features to characterize it quantitatively (see Fig. 3). These technologies have been constructed using AI approaches based on radiomics, ML<sup>32,63–65</sup> (e.g., random forests, support vector machines), and DL<sup>32,63–65</sup>. In particular, CNNs are the most common choice of architecture for DL-based CADE and CADx. For example, Chakrabarty et al.<sup>66</sup> trained a CNN with post-contrast,  $T_1$ -weighted data to perform differential diagnosis of six brain tumor types and healthy tissue, achieving AUCs over 0.95 in internal and external validation. Additionally, Saha et al.<sup>67</sup> achieved AUCs over 0.86 with a CNN-based CADE/CADx system for clinically significant prostate cancer that was informed by  $T_2$ -weighted, diffusion-weighted MRI (DW-MRI), apparent diffusion coefficient maps (ADC), and an anatomic prior of zonal cancer prevalence. While CNN training requires vast labeled datasets and computational resources, pre-trained CNNs (e.g., AlexNet,

**Fig. 2 | The flow chart depicts the downstream impact that quantitative imaging workflow can have on clinical applications.** Beginning with image acquisition (A), variability and uncertainty is propagated through each step and can affect the accuracy in assessing lesion response and/or anticipated clinical outcome (H). Image processing routines (B) are then used to quantify image features prior to image registration (C), however, image registration may also occur prior to post-processing. Once the images are co-registered, automatic or manual segmentation (D) identifies the tissue of interest. At this stage, the clinical response can be assessed by determining lesion response (E). Alternatively, the features (F) from imaging and -omics (G) can be identified to predict clinical outcome (H) or lesion response (E). Error at any of these steps can compound and result in a false classification of patient outcomes.



**Fig. 3 | AI for imaging-based cancer detection and diagnosis.** The figure illustrates the global process whereby MRI sequences (i.e., T<sub>2</sub>-weighted (T2W), DW-MRI, and DCE-MRI) are processed with a CNN to identify whether there is a tumor present, its localization, and the differential diagnosis (e.g., tumor class, tumor subtype, clinical risk). Several CNN architectures have been used to construct CADe and CADx systems for this type of MRI analysis<sup>66,67,69,91,94</sup>. This figure shows the use of a U-net due to its successful use for automatic tumor detection and segmentation, as well as in the context of cancer diagnosis.



GoogLeNet) can also be used in new datasets for cancer detection and diagnosis tasks (i.e., transfer learning)<sup>32,65,68,69</sup>. For instance, Antropova et al.<sup>70</sup> achieved an AUC of 0.89 in the diagnosis of malignant breast lesions on dynamic contrast-enhanced MRI (DCE-MRI) data with a CADx system that combined radiomics with CNN-based features, for which they used the pre-trained VGG19 model.

### Current clinical needs

There are ongoing efforts to leverage AI to analyze standard-of-care MRI data for cancer detection and diagnosis to reduce inter-reader variability, while also streamlining time-intensive diagnostic processes<sup>32,63–65,69</sup>. These methods attempt to enable timely and reliable assistance for MRI-informed clinical decision-making for diagnosis, monitoring, and treatment. For example, CADe/CADx systems can assist in determining an appropriate clinical management strategy for the identification of clinically-significant prostate cancer<sup>64,69</sup> on multiparametric MRI data. A fundamental challenge in neuro-oncology is the differential diagnosis of primary central nervous system tumor subtypes and brain metastases, which necessitate different treatments<sup>32,66</sup>. Current clinical needs in breast cancer diagnosis include early detection of high-risk disease, improvement of screening for intermediate-risk and dense breast populations, the differentiation between benign and malignant lesions, and the identification of specific subtypes<sup>63,65,70,71</sup>. These are all tasks that could potentially be assisted by leveraging AI methods, although there are several challenges to their development (see Section 5.3). Additionally, more recent efforts are tackling the integration of multimodal data (e.g., imaging, histopathology, biomarkers, omics) using AI to better inform cancer management.

### Barriers to practical deployment in the clinical setting

In Medicine, as opposed to many other fields of application of AI, understanding why a decision is made (e.g., a diagnosis) is as important as the decision itself. Thus, establishing model interpretability is a fundamental barrier in the clinical translation of AI-based techniques for decision-making in clinical oncology<sup>63,64,72,73</sup>. Beyond explaining the biophysical causes underlying AI-driven cancer detection and diagnosis, model interpretability is also required to extrapolate and interrogate model outcomes<sup>63,64</sup>, which would contribute to more informed clinical decisions. The medical AI community has proposed several approaches to address the lack of interpretability in MRI-informed AI models for cancer detection and diagnosis<sup>74–76</sup>, such as class activation methods<sup>66,67</sup>, knowledge-driven priors<sup>67,74</sup>, and integration of multimodal multiscale data<sup>77,78</sup>. Furthermore, recent approaches in the field of scientific machine learning have shown promise in improving the interpretability of AI models<sup>79</sup> and could therefore be developed for the detection and diagnosis of cancer on MRI data. For example, mechanistic feature engineering can identify biophysically-relevant inputs that characterize tumor biology<sup>80,81</sup>, physics-informed neural networks (PINNs)<sup>82,83</sup> include a (bio)physical model in the loss function, and biology-informed neural networks (BINNs) adapt the model architecture according to prior biological knowledge<sup>84,85</sup>.

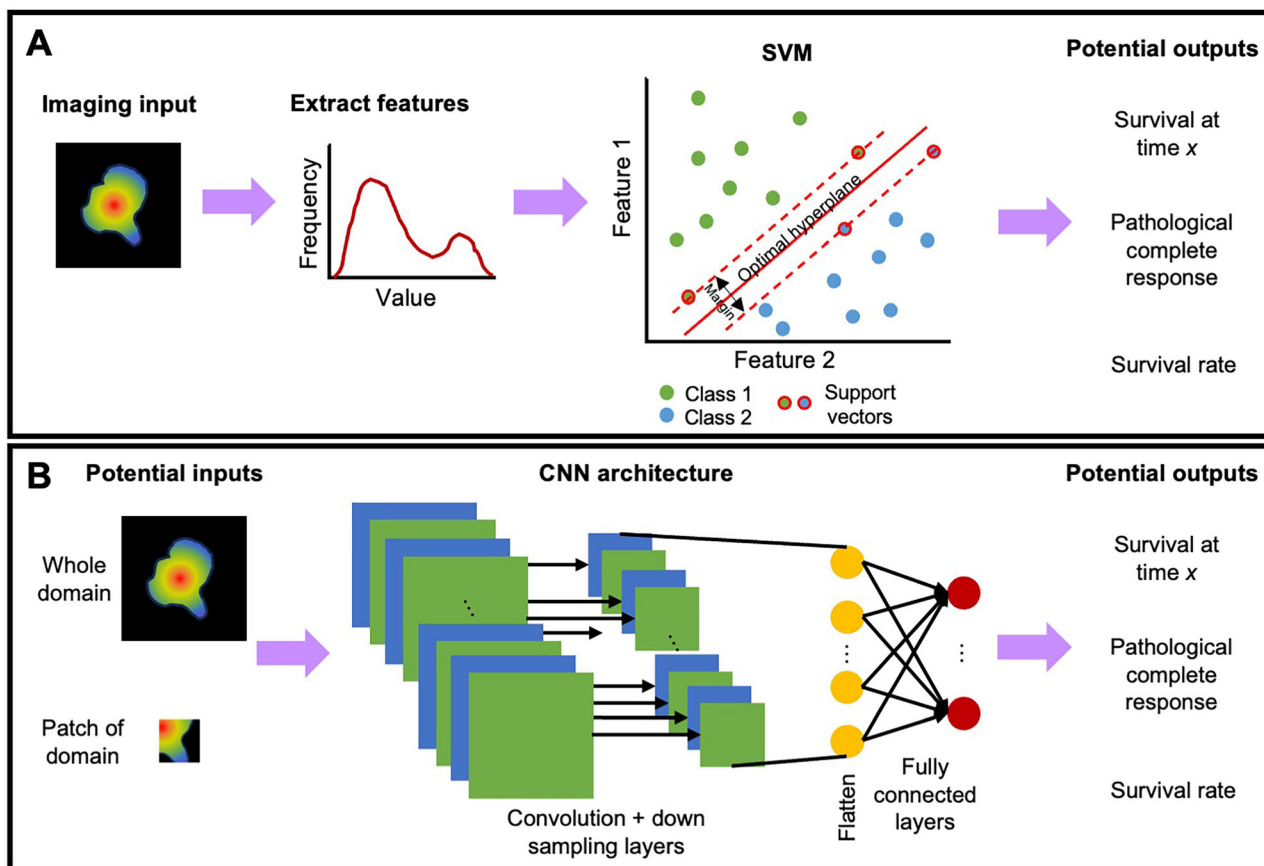
Despite the progress in the development of MRI-informed AI methods for cancer detection and diagnosis, these technologies need to address potentially critical pitfalls in their performance in tumor-specific scenarios. For example, the detection of malignant breast lesions using MRI-informed AI models can be affected by complex tumor geometries (e.g., non-mass tumors), the architectural and radiological features of surrounding healthy tissue (e.g., tissue density, background parenchymal enhancement), as well as signal distortion and movement of the tumor during DCE acquisition<sup>65,71,86,87</sup>. Additionally, the performance of CADe/CADx methods for prostate cancer can be affected by several well-established MRI confounders<sup>64,67,88–90</sup>, such as prostatitis and benign prostatic hyperplasia. Similarly, the different types of tumors that may develop in the brain can produce similar MRI signals that complicate their differential diagnosis<sup>32,66</sup>. To address these issues, future studies require (i) training and validation databases that balance the amount and diversity of confounding features and comorbidities (e.g., breasts with mass-like and non-mass geometries; prostates with cancer alone and combined with other prostatic pathologies; and diverse brain tumor cases), and (ii) training AI models to recognize confounding patterns (e.g., standard textural and deep learning features) to boost the performance of CADe/CADx technologies for cancer<sup>32,64–67,71,86–88</sup>.

Several validation, implementation, ethical, and data issues have also hindered the clinical translation of AI models for cancer detection and diagnosis. Firstly, the diagnostic performance of CADe/CADx systems still needs to be externally validated in large prospective clinical trials including diverse imaging acquisition methods, patient demographics, and reader expertise<sup>32,63–65,69,91,92</sup>. Additionally, two key obstacles to their practical clinical deployment are the lack of local data science support and the high computational cost, particularly during the training of DL models. The need for high-performance hardware and efficient algorithms demands a considerable financial investment, hindering widespread adoption in resource-constrained healthcare settings. As noted in Section 5.1, transfer learning is a promising strategy to address these computational limitations by leveraging pre-trained DL models<sup>32,65,68,69</sup>. Another fundamental issue is the limited availability of rigorously curated data for AI model development, which may require specific labels and non-standard preprocessing that can be costly and time-consuming<sup>63</sup>. Furthermore, biases in data collection processes and unrecognized biases in clinical practices from which the training data is gathered can lead to distorted outcomes, even perpetuating societal inequalities<sup>93</sup>. Data inaccessibility due to confidentiality concerns, potential breaches compromising patient confidentiality, and strict data privacy regulations limiting collaboration also constitute important challenges<sup>63,93</sup>. To address limitations in data transfer for multicenter collaborations, federated learning<sup>94</sup> is a potential strategy that relies on sharing model parameter updates rather than datasets.

### AI for predicting response from MRI data

#### Common AI techniques for predicting response

AI can be used to link imaging data to outcomes such as pathological response, time to recurrence, and overall survival. Commonly used machine



**Fig. 4 | Demonstration of how an SVM and a CNN can use imaging data to predict treatment response.** In Panel A, features related to histograms of relative cerebral blood volume<sup>97</sup> or peak height of a perfusion signal<sup>107</sup> are extracted from the imaging data. Clinical information such as patient age and genetic data can also be included as features. The goal of the SVM is to take  $N$  features and determine the  $(N-1)$ -dimensional hyperplanes that maximally separate (for example) patients into short, medium, or long survival, or complete response as determined by pathology. Panel B shows potential inputs to a CNN – either the whole image domain, imaging-

derived features, or a patch of the domain. Extracting patches from a domain can be used to increase the amount of training data, or to reduce computational burden when working with large images. These are then input to a CNN, here represented with convolution and down sampling layers feeding into a fully connected architecture. In general, multiple sets of convolution and down-sampling layers are used. The network output accomplishes the same goal as the SVM in Panel A; namely, separating inputs into classes such as responders and non-responders, or survival at a particular time.

learning methods for predicting response include support vector machine (SVM)<sup>95–97</sup>, regression<sup>98,99</sup>, random survival forest (RSF)<sup>100</sup>, clustering<sup>101,102</sup>, and CNNs<sup>99,103,104</sup>. Recently, more DL models<sup>105</sup> are established based on the transformer architecture or incorporating the attention mechanism<sup>106</sup>. Panels A and B of Fig. 4 demonstrate how SVM and CNN, respectively, can be used in prediction of response.

In brain cancer, combining clinical data with AI methods has been shown to be more effective in predicting survival outcomes than clinical methods alone<sup>107,108</sup>. SVM, for instance, has been used to predict treatment outcomes for gliomas<sup>96</sup> by employing both clinical and functional features<sup>96</sup>. An SVM trained by Emblem et al. found whole tumor relative cerebral blood volume was the optimal predictor of overall survival<sup>97</sup>. CNNs have been used for the assessment and prediction of outcomes. For example, Jang et al.<sup>104</sup> distinguished between pseudo-progression and progression in patients with GBM using CNNs with long short-term memory (LSTM)<sup>109,110</sup>, an ML algorithm used to train recurrent NNs. They compare two options for the CNN input data: (1) MRI (post-contrast  $T_1$ -weighted images) and clinical parameters, and (2) MRI data alone. The model trained on MRI and clinical data outperformed the MRI-only model, as the AUC for predicting progression versus pseudo-progression had values of 0.83 and 0.69, respectively. This demonstrates that the combination of AI, MRI, and clinical features might help in post-treatment decision-making for patients with GBM<sup>104</sup>.

In prostate cancer, MRI-based AI has enabled the prediction of recurrence after surgery and radiotherapy<sup>111</sup>. For example, Lee et al.<sup>99</sup>

developed a DL model trained on preoperative multiparametric MRI data ( $T_2$ -weighted, DW-MRI, and DCE-MRI) to predict long-term post-surgery recurrence-free survival. Using Cox models and Kaplan-Meier survival analysis, the features obtained from multiparametric MRI data via their DL model outperformed clinical and radiomics features, while the combination of clinical and DL features yielded the best predictive performance. Additionally, in breast cancer, prognostic CNN models trained on MRI data have also enabled the prediction of pathological complete response for neoadjuvant chemotherapy<sup>112</sup>.

### Current clinical needs

Common clinical challenges in oncology include the stratification of patients by treatment response, risk of relapse, and overall survival<sup>113,114</sup>. Identification of patient outcomes before treatment begins (or early during treatment) could help select, escalate, or de-escalate prescribed treatments, as well as tailor personalized follow-up schedules. It is important to note that analyses (AI-based and otherwise) leveraging multi-parametric MRI as compared to biopsy-based stratification also enable longitudinal whole-tumor coverage addressing the sampling bias limitations of biopsy<sup>113</sup>. Furthermore, many of these AI-based approaches leveraging imaging data achieve greater performance over standard clinical information (e.g., age, sex, extent of surgery) alone<sup>113,114</sup>.



## Barriers to practical deployment in the clinical setting

Although AI-based approaches have demonstrated greater performance over standard clinical information in some settings, these studies are typically done in specific cohorts with limited external validation and therefore have limited generalizability<sup>115</sup>. The variability in MR acquisition techniques between manufacturers, scanner types, protocols, and institutions can lead to substantial bias in training image-guided AIs without proper data harmonization<sup>116</sup>, such as that completed by Marzi et al.<sup>117</sup>, who were able to reduce site effects after data harmonization in a study of T<sub>1</sub>-weighted MRI data from 1740 healthy subjects at 36 sites using a harmonizer transformer as part of the preprocessing steps for a machine learning pipeline. Moreover, restrictions on inter-institution sharing of patient data<sup>118</sup> (due to concerns such as privacy) may restrict the verification of model generalizability. These limitations, alongside uncertainty in the AI models themselves, create substantial barriers to the widespread clinical adoption of AI models for response prediction. Federated learning could be a realistic strategy to bypass the complexity of inter-institutional data sharing. Instead of requiring assembling data from different institutions into a centralized large-scale dataset, federated learning enables training of the AI model on decentralized, private datasets in multiple sites independently, then only the trained parameters (instead of original data) will be shared between training sites to generate the globally tuned model<sup>119,120</sup>.

Efforts to construct large, publicly available MRI datasets for brain tumors are ongoing. However, careful considerations need to be made while using them in training, testing or validating AI-based model<sup>121</sup>. One significant issue is the potential for overlaps between different datasets (i.e., multiple datasets containing the same patient), which can reduce the number of uniquely available data. More specifically, the IvyGAP Radiomics dataset contains the pre-operative MRIs of the IvyGap dataset with additional segmentations and derived radiomics parameters. Moreover, the BraTS 2021 dataset contains data that was available in the previous BraTS challenges and other public datasets such as the TCGA-LGG (65 patients), TCGA GBM (102 patients), and Ivy Gap (30 patients). This may lead to redundancy and diminish the overall diversity of the training data<sup>121</sup>. Furthermore, these datasets have been published for more than a decade and some of them have undergone updates, adding a higher variability in protocols, scans quality, and evolving WHO classification. We identify similar issues in prostate cancer<sup>122</sup>, particularly when it comes to the age of the publicly available datasets (many being 10+ years old) and overlap between datasets (which may or may not be known between sets). Dataset sizes are also often small, especially once missing or low-quality data is removed.

Efforts for breast cancer data standardization are also in progress. For instance, Kilintzis et al.<sup>123</sup> attempted to produce a harmonized dataset from 5 publicly available datasets from the TCIA platform, including 2035 patients. More generally, Kosvira et al.<sup>124</sup> propose a new methodology for assessing the data quality of cancer imaging repositories by incorporating a three-step procedure that includes a data integration quality check tool to ensure compliance with quality requirements.

Federated learning could be a practical strategy to overcome the complexity of inter-institutional data sharing. Instead of assembling data from different institutions into a centralized large-scale dataset, federated learning enables the training of the AI model on decentralized, private datasets in multiple sites independently, then only the trained parameters (instead of original data) are shared between training sites to generate the globally tuned model. We discuss these points via the results presented in Rauniyar et al.<sup>119</sup> and Guan et al.<sup>120</sup>.

Beyond data variation, the cancer population itself is heterogeneous (as categorized by the cancer subtypes, staging, patient demographics, etc.), which leads to differences in therapy response between patients<sup>125</sup>. Moreover, novel therapies can be introduced into clinical care, the therapeutic regimens vary between institutions, and treatments can be refined, leading to significant differences in datasets collected at different times<sup>126</sup>. Overall, the trade-off between the sample size and homogeneity of accessible datasets to train AI models is a major barrier to the robust application of AI for response prediction. One method that could help combat generalizability

issues with small datasets is pretraining. For example, Yuan et al.<sup>127</sup> and Han et al.<sup>128</sup> pretrain convolutional neural networks using the ImageNet<sup>129</sup> dataset to successfully classify risk in prostate tumors and differentiate between long- and short-term glioma survivors, respectively. As an alternative to pretraining on ImageNet, Wen et al.<sup>130</sup> explore the possible benefits of pretraining using medical images. While their study ultimately found that ImageNet pretraining provided more accurate results, medical pretraining showed potential. As such, it could be useful to construct a large medical image database like ImageNet to further explore pretraining for medical problems using medical images. Zhang et al.<sup>131</sup> explore another possible workflow for fine-tuning a network designed to classify breast cancer molecular subtypes from DCE-MRI data. Rather than relying on a separate, large dataset for pre-training, they separate their data into a training set and two testing sets – A and B. They then compare (1) testing the network with both A and B, (2) fine-tuning with B and testing with A, and (3) fine-tuning with A and testing with B. They conclude that finetuning with this method increases accuracy. Thus, in a situation where the dataset of interest is very small, initial network training could be completed on large, publicly available datasets, then fine-tuning and testing could be completed using the dataset of interest.

Many AI-based algorithms lack interpretability, thereby also limiting their clinical adoption. Furthermore, most AI-based approaches for outcome prediction are deployed with set parameters that govern their sensitivity and specificity. The patient-specific use of these models may require their optimization to fit the specific wishes of the patient, their family, and their clinical team. This patient-specific optimization creates an additional barrier to clinical deployment. The integration of AI-based prediction algorithms into the clinical workflow faces additional challenges as it has not been well established and currently generates additional work for the clinical team. This limits their adoption to clinics that have a robust computational infrastructure, clinicians who are willing to spend extra time generating the data, and ready access to the necessary data for the algorithm.

## Discussion

AI-based methods have established some measure of success in several areas of MRI including image acquisition, reconstruction, registration, and segmentation, as well as assisting in diagnosis and prognosis. Some registration and segmentation techniques have even been approved by the FDA for clinical application<sup>69,91,92,132</sup>. However, while the application of AI techniques for MRI in cancer clearly has tremendous potential, three major challenges persist: (1) model generalizability, (2) model interpretability, and (3) establishing confidence in the output of an AI model. Indeed, these problems are common to many applications of AI in medicine.

The limited progress in being able to transport an AI method from one institute to another, and from one disease setting to another (i.e., generalizability), is heavily influenced by both the data and the devices employed to capture the data<sup>133</sup>. The wide variety of the available scanners (both from different manufacturers and then different models with a manufacturer), as well as the protocols they run, can further hamper generalizability as the data employed for training may not adequately sample all the acquisition scenarios encountered in the testing data set<sup>133,134</sup>. The lack of standardization in quality assurance and control (QA/QC) adds additional complexities<sup>135</sup>. Even when an AI method has had success in one clinical setting, it does not guarantee it will be successful when applied in another disease setting due to variations in patient characteristics and previously received treatments<sup>136</sup>. Thus, retraining the AI model is required and this can introduce nontrivial changes to the AI architecture or method of implementation. This is especially true in cancer, where the differences between diseases and the site of origin can introduce tremendous anatomical and physiological heterogeneities that may confound a previously trained AI model. It is increasingly recognized that each model should be evaluated with local data in the disease setting for which it is intended to be used<sup>137</sup>.

A fundamental issue with the majority of AI-based analysis is their limited “interpretability” in the sense of only providing a limited understanding of the relationships between model input and model output; that is,



there is limited insight into how the AI model gets from cause to effect. This is an active area of investigation<sup>72,138,139</sup> and is now well-recognized for being extremely important in domains that have high-consequence decisions like oncology<sup>140,141</sup>. For example, lack of interpretability generates problems when trying to identify optimal interventions for a particular patient where it is critical to understand why an AI model selects a particular therapeutic regimen over another. In particular, for an AI model built on population-based MRI data, does the training data set adequately capture the unique characteristics of the features in the individual's MRI data? One attractive way forward is to link mechanism-based modeling with AI methods and scientific machine learning in which known/established biological and physical laws can be explicitly incorporated in the AI algorithm<sup>142–144</sup>. This can have the additional benefit of increasing confidence in the method.

The challenge in establishing confidence is not only related to the lack of model generalizability and interpretability, but also concern over the ethical issues associated with applying AI techniques to healthcare data as well as maintaining patient privacy and data security. This is particularly true with medical imaging data in which detailed anatomical features can be rendered in 3D. Importantly, once introduced in the clinical workflow, the performance of the AI tool must be continuously monitored as the data inputs are adjusted with new imaging hardware, software, and methods of image acquisition and subsequent analysis. More generally, the Office of Science and Technology Policy has published a white paper entitled, “A Blueprint for an AI Bill of Rights”<sup>145</sup> designed to guide the safe, effective, and unbiased development and application of AI.

Beyond the issues related to generalizability, interpretability, and confidence enumerated above, there may be fundamental limitations to what AI can contribute to cancer imaging. For example, it is important to note that every AI-based method requires a training set and since cancer is notoriously heterogeneous across both space and time, there are fundamental limits to what a population-based method can achieve. Indeed, this problem manifests itself in everything from image reconstruction to predicting response if the specific details of the pathology under investigation are not well-characterized in the training set. For a disease as heterogeneous as cancer, and an imaging modality as flexible as MRI, this seems to be a difficult problem to overcome with an AI-only approach.

## Conclusion

AI has shown promise in accelerating image acquisition and reconstruction methods to maintain high spatial resolution in a fraction of the time typically required to obtain such an image. It has also been a useful tool for improving image segmentation and registration, as well as offering valuable results in both diagnostic and prognostic settings. However, questions remain about the interpretability and generalizability of these techniques, especially when a method is ported from one disease setting to another or, even, from one institution to another institution. The dearth of robust QA/QC and ethics-ensuring methods means that much work is left to be done to maintain patient safety given the current status of AI techniques in cancer imaging and healthcare. Finally, questions concerning fundamental limitations for any method that requires a large training dataset remain.

## Data availability

No datasets were generated or analyzed during the current study.

Received: 21 July 2024; Accepted: 17 March 2025;

Published online: 09 April 2025

## References

- Chen, Y. et al. AI-based reconstruction for Fast MRI—A systematic review and meta-analysis. *Proc. IEEE* **110**, 224–245 (2022).
- Shimron, E. & Perlman, O. AI in MRI: Computational frameworks for a faster, optimized, and automated imaging workflow. *Bioengineering* **10**, 492 (2023).
- Chen, W. et al. Artificial intelligence powered advancements in upper extremity joint MRI: A review. *Heliyon* **10**, e28731 (2024).
- Li, C. et al. Artificial intelligence in multiparametric magnetic resonance imaging: A review. *Med. Phys.* **49**, <https://doi.org/10.1002/mp.15936> (2022).
- McCarthy, J., Minsky, M. L., Rochester, N. & Shannon, C. E. A proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. *AI Mag.* **27**, 12 (2006).
- Ongie, G. et al. Deep learning techniques for inverse problems in imaging. *IEEE J. Sel. Areas Inf. Theory* **1**, 39–56 (2020).
- Shen, D., Wu, G. & Suk, H. I. Deep learning in medical image analysis. *Annu. Rev. Biomed. Eng.* **19**, 221–248 (2017).
- Lundervold, A. S. & Lundervold, A. An overview of deep learning in medical imaging focusing on MRI. *Z. F. ür. Med Phys.* **29**, 102–127 (2019).
- Karimi, D., Dou, H., Warfield, S. K. & Gholipour, A. Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis. *Med. Image Anal.* **65**, 101759 (2020).
- Iglesias, J. E. et al. SynthSR: A public AI tool to turn heterogeneous clinical brain scans into high-resolution T1-weighted images for 3D morphometry. *Sci. Adv.* **9**, eadd3607 (2023).
- De Leeuw Den Bouter, M. L. et al. Deep learning-based single image super-resolution for low-field MR brain images. *Sci. Rep.* **12**, 6362 (2022).
- Guan, H. & Liu, M. Domain adaptation for medical image analysis: a survey. *IEEE Trans. Biomed. Eng.* **69**, 1173–1185 (2022).
- Tamir, J. I. et al. Computational MRI with physics-based constraints: application to multicontrast and quantitative imaging. *IEEE Signal Process Mag.* **37**, 94–104 (2020).
- Dang, H. N. et al. MR-zero meets RARE MRI: Joint optimization of refocusing flip angles and neural networks to minimize  $T_2$ -induced blurring in spin echo sequences. *Magn. Reson Med* **90**, 1345–1362 (2023).
- Loktyushin, A. et al. MRzero - Automated discovery of MRI sequences using supervised learning. *Magn. Reson Med* **86**, 709–724 (2021).
- Aggarwal, H. K. & Jacob, M. J-MoDL: Joint model-based deep learning for optimized sampling and reconstruction. *IEEE J. Sel. Top. Signal Process* **14**, 1151–1162 (2020).
- Bahadir C. D., Dalca A. V. & Sabuncu M. R. Learning-Based Optimization of the Under-Sampling Pattern in MRI. In: Chung A. C. S., Gee J. C., Yushkevich P. A., Bao S., eds. *Information Processing in Medical Imaging*. Vol 11492. Lecture Notes in Computer Science. Springer International Publishing; 780–792. (2019).
- Ravula S., Levac B., Jalal A., Tamir J. I. & Dimakis A. G. Optimizing Sampling Patterns for Compressed Sensing MRI with Diffusion Generative Models. Published online. <https://doi.org/10.48550/ARXIV.2306.03284> (2023).
- Wang, G., Luo, T., Nielsen, J. F., Noll, D. C. & Fessler, J. A. B-spline parameterized joint optimization of reconstruction and K-Space Trajectories (BJORK) for accelerated 2D MRI. *IEEE Trans. Med. Imaging* **41**, 2318–2330 (2022).
- Lustig, M., Donoho, D. & Pauly, J. M. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magn. Reson. Med.* **58**, 1182–1195 (2007).
- Aggarwal, H. K., Mani, M. P. & Jacob, M. MoDL: Model-based deep learning architecture for inverse problems. *IEEE Trans. Med. Imaging* **38**, 394–405 (2019).
- Hammernik, K. et al. Learning a variational network for reconstruction of accelerated MRI data. *Magn. Reson. Med.* **79**, 3055–3071 (2018).
- Jalal A. et al. Robust Compressed Sensing MRI with Deep Generative Priors. In: Ranzato M., Beygelzimer A., Dauphin Y., Liang P. S., Vaughan J. W., eds. *Advances in Neural Information Processing Systems*. Vol 34. Curran Associates, Inc.; 14938–14954. [https://proceedings.neurips.cc/paper\\_files/paper/2021/file/7d6044e95a16761171b130dcb476a43e-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2021/file/7d6044e95a16761171b130dcb476a43e-Paper.pdf). (2021).

24. Luo, G., Blumenthal, M., Heide, M. & Uecker, M. Bayesian MRI reconstruction with joint uncertainty estimation using diffusion models. *Magn. Reson. Med.* **90**, 295–311 (2023).
25. Johnson, P. M. et al. Deep learning reconstruction enables prospectively accelerated clinical knee MRI. *Radiology* **307**, e220425 (2023).
26. Chen, F. et al. Variable-density single-shot fast spin-echo MRI with deep learning reconstruction by using variational networks. *Radiology* **289**, 366–373 (2018).
27. Yankeelov, T. & Gore, J. Dynamic contrast-enhanced magnetic resonance imaging in oncology: theory, data acquisition, analysis, and examples. *Curr. Med. Imaging Rev.* **3**, 91–107 (2007).
28. Gong, E., Pauly, J. M., Wintermark, M. & Zaharchuk, G. Deep learning enables reduced gadolinium dose for contrast-enhanced brain MRI. *J. Magn. Reson. Imaging* **48**, 330–340 (2018).
29. Muckley, M. J. et al. Results of the 2020 fastMRI challenge for machine learning MR image reconstruction. *IEEE Trans. Med. Imaging* **40**, 2306–2317 (2021).
30. Daye, D. et al. Implementation of clinical artificial intelligence in radiology: who decides and how? *Radiology* **305**, 555–563 (2022).
31. Hoebel, K. V. et al. Expert-centered evaluation of deep learning algorithms for brain tumor segmentation. *Radio. Artif. Intell.* **6**, e220231 (2024).
32. Cè, M. et al. Artificial Intelligence in brain tumor imaging: a step toward personalized medicine. *Curr. Oncol.* **30**, 2673–2701 (2023).
33. Egger, J. et al. GBM Volumetry using the 3D Slicer Medical Image Computing Platform. *Sci. Rep.* **3**, 1364 (2013).
34. Lambert, S. I. et al. An integrative review on the acceptance of artificial intelligence among healthcare professionals in hospitals. *Npj Digit. Med.* **6**, 111 (2023).
35. Baroudi, H. et al. Automated contouring and planning in radiation therapy: what is ‘Clinically Acceptable’? *Diagnostics* **13**, 667 (2023).
36. Pal, R. et al. Lumbar Spine Tumor Segmentation and Localization in T2 MRI Images Using AI. Published online. <https://doi.org/10.48550/ARXIV.2405.04023> (2024).
37. Kundal, K., Rao, K. V., Majumdar, A., Kumar, N. & Kumar, R. Comprehensive benchmarking of CNN-based tumor segmentation methods using multimodal MRI data. *Comput. Biol. Med.* **178**, 108799 (2024).
38. Eppenhof, K. A. J. & Pluim, J. P. W. Pulmonary CT registration through supervised learning with convolutional neural networks. *IEEE Trans. Med. Imaging* **38**, 1097–1105 (2019).
39. Osman, A. F. I., Al-Mugren, K. S., Tamam, N. M. & Shahine, B. Deformable registration of magnetic resonance images using unsupervised deep learning in neuro-/radiation oncology. *Radiat. Oncol.* **19**, 61 (2024).
40. Rudie, J. D. et al. Longitudinal assessment of posttreatment diffuse glioma tissue volumes with three-dimensional convolutional neural networks. *Radio. Artif. Intell.* **4**, e210243 (2022).
41. Teuwen, J., Gouw, Z. A. R. & Sonke, J. J. Artificial intelligence for image registration in radiation oncology. *Semin Radiat. Oncol.* **32**, 330–342 (2022).
42. Simonovsky M., Gutiérrez-Becker B., Mateus D., Navab N. & Komodakis N. A Deep Metric for Multimodal Registration. In: Ourselin S., Joskowicz L., Sabuncu M. R., Unal G., Wells W., eds. *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2016*. Vol 9902. Lecture Notes in Computer Science. Springer International Publishing 10–18 (2016).
43. Haskins, G. et al. Learning deep similarity metric for 3D MR-TRUS image registration. *Int. J. Comput. Assist. Radio. Surg.* **14**, 417–425 (2019).
44. Mohseni Salehi, S. S., Khan, S., Erdogmus, D. & Gholipour, A. Real-time deep pose estimation with geodesic loss for image-to-template rigid registration. *IEEE Trans. Med. Imaging* **38**, 470–481 (2019).
45. Estienne, T. et al. Deep learning-based concurrent brain registration and tumor segmentation. *Front. Comput. Neurosci.* **14**, 17 (2020).
46. Bakas S. et al. Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge. Published online November 5. <https://doi.org/10.17863/CAM.38755> (2018).
47. Balakrishnan, G., Zhao, A., Sabuncu, M. R., Guttag, J. & Dalca, A. V. VoxelMorph: A Learning Framework for Deformable Medical Image Registration. *IEEE Trans. Med. Imaging* **38**, 1788–1800 (2019).
48. Dalca, A. V., Balakrishnan, G., Guttag, J. & Sabuncu, M. R. Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces. *Med. Image Anal.* **57**, 226–236 (2019).
49. Pellicer-Valero, O. J. et al. Deep learning for fully automatic detection, segmentation, and Gleason grade estimation of prostate cancer in multiparametric magnetic resonance images. *Sci. Rep.* **12**, 2975 (2022).
50. Cardenas, C. E., Yang, J., Anderson, B. M., Court, L. E. & Brock, K. B. Advances in auto-segmentation. *Semin Radiat. Oncol.* **29**, 185–197 (2019).
51. Warfield, S. K., Zou, K. H. & Wells, W. M. Validation of image segmentation by estimating rater bias and variance. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **366**, 2361–2375 (2008).
52. Ramesh, K. K. D., Kumar, G. K., Swapna, K., Datta, D. & Rajest, S. S. A Review of Medical Image Segmentation Algorithms. *EAI Endorsed Trans. Pervasive Health Technol.* **7**, e6 (2021).
53. Jungo A. et al. On the Effect of Inter-observer Variability for a Reliable Estimation of Uncertainty of Medical Image Segmentation. In: Frangi A. F., Schnabel J. A., Davatzikos C., Alberola-López C., Fichtinger G., eds. *Medical Image Computing and Computer Assisted Intervention - MICCAI 2018*. Vol 11070. Lecture Notes in Computer Science. Springer International Publishing; 682–690. (2018).
54. Renard, F., Guedria, S., Palma, N. D. & Vuilleme, N. Variability and reproducibility in deep learning for medical image segmentation. *Sci. Rep.* **10**, 13724 (2020).
55. Tixier, F., Um, H., Young, R. J. & Veeraraghavan, H. Reliability of tumor segmentation in glioblastoma: Impact on the robustness of MRI-radiomic features. *Med. Phys.* **46**, 3582–3591 (2019).
56. Poirot, M. G. et al. Robustness of radiomics to variations in segmentation methods in multimodal brain MRI. *Sci. Rep.* **12**, 16712 (2022).
57. Koçak, B. Key concepts, common pitfalls, and best practices in artificial intelligence and machine learning: focus on radiomics. *Diagn. Inter. Radio.* **28**, 450–462 (2022).
58. Kocak, B., Durmaz, E. S., Ates, E. & Kilickesmez, O. Radiomics with artificial intelligence: a practical guide for beginners. *Diagn. Inter. Radio.* **25**, 485–495 (2019).
59. Chen, Z. et al. What matters in radiological image segmentation? Effect of segmentation errors on the diagnostic related features. *J. Digit. Imaging* **36**, 2088–2099 (2023).
60. Horvat, N., Papanikolaou, N. & Koh, D. M. Radiomics beyond the hype: a critical evaluation toward oncologic clinical use. *Radio. Artif. Intell.* **6**, e230437 (2024).
61. Fu, Y. et al. Deep learning in medical image registration: a review. *Phys. Med. Biol.* **65**, 20TR01 (2020).
62. Mårtensson, G. et al. The reliability of a deep learning model in clinical out-of-distribution MRI data: A multicohort study. *Med. Image Anal.* **66**, 101714 (2020).
63. Bi, W. L. et al. Artificial intelligence in cancer imaging: Clinical challenges and applications. *CA Cancer J. Clin.* **69**, 127–157 (2019).
64. Wildeboer, R. R., Van Sloun, R. J. G., Wijkstra, H. & Mischi, M. Artificial intelligence in multiparametric prostate cancer imaging with focus on deep-learning methods. *Comput. Methods Prog. Biomed.* **189**, 105316 (2020).
65. Jones, M. A., Islam, W., Faiz, R., Chen, X. & Zheng, B. Applying artificial intelligence technology to assist with breast cancer

- diagnosis and prognosis prediction. *Front. Oncol.* **12**, 980793 (2022).
66. Chakrabarty, S. et al. MRI-based identification and classification of major intracranial tumor types by using a 3D Convolutional Neural Network: A retrospective multi-institutional analysis. *Radio. Artif. Intell.* **3**, e200301 (2021).
67. Saha, A., Hosseinzadeh, M. & Huisman, H. End-to-end prostate cancer detection in bpMRI via 3D CNNs: Effects of attention mechanisms, clinical priori and decoupled false positive reduction. *Med. Image Anal.* **73**, 102155 (2021).
68. Giger, M. L. Machine learning in medical imaging. *J. Am. Coll. Radio.* **15**, 512–520 (2018).
69. Twilt, J. J., Van Leeuwen, K. G., Huisman, H. J., Fütterer, J. J. & De Rooij, M. Artificial intelligence based algorithms for prostate cancer classification and detection on magnetic resonance imaging: a narrative review. *Diagnostics* **11**, 959 (2021).
70. Antropova, N., Huynh, B. Q. & Giger, M. L. A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets. *Med. Phys.* **44**, 5162–5171 (2017).
71. Sardanelli, F. et al. The paradox of MRI for breast cancer screening: high-risk and dense breasts—available evidence and current practice. *Insights Imaging* **15**, 96 (2024).
72. Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* **1**, 206–215 (2019).
73. Hatherley, J., Sparrow, R. & Howard, M. The virtues of interpretable medical AI. *Camb. Q Health Ethics* **33**, 323–332 (2024).
74. Salahuddin, Z., Woodruff, H. C., Chatterjee, A. & Lambin, P. Transparency of deep neural networks for medical image analysis: A review of interpretability methods. *Comput. Biol. Med.* **140**, 105111 (2022).
75. Band, S. et al. Application of explainable artificial intelligence in medical health: A systematic review of interpretability methods. *Inf. Med. Unlocked* **40**, 101286 (2023).
76. Champendal, M., Müller, H., Prior, J. O. & Dos Reis, C. S. A scoping review of interpretability and explainability concerning artificial intelligence methods in medical imaging. *Eur. J. Radio.* **169**, 111159 (2023).
77. Lipkova, J. et al. Artificial intelligence for multimodal data integration in oncology. *Cancer Cell* **40**, 1095–1110 (2022).
78. Pavard, H., Zormpas-Petridis, K., Clayton, H., Burge, S. & Crispin-Ortuzar, M. Radiology and multi-scale data integration for precision oncology. *Npj Precis. Oncol.* **8**, 158 (2024).
79. Metcarr, J., Jutzeler, C. R., Macklin, P., Köhn-Luque, A. & Brünink, S. C. A review of mechanistic learning in mathematical oncology. *Front. Immunol.* **15**, 1363144 (2024).
80. Lorenzo, G. et al. Patient-specific, mechanistic models of tumor growth incorporating artificial intelligence and big data. *Annu. Rev. Biomed. Eng.* **26**, 529–560 (2024).
81. Bosque, J. J. et al. Metabolic activity grows in human cancers pushed by phenotypic variability. *iScience* **26**, 106118 (2023).
82. Raissi, M., Perdikaris, P. & Karniadakis, G. E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **378**, 686–707 (2019).
83. Zhang, R. Z. et al. Personalized predictions of Glioblastoma infiltration: Mathematical models, Physics-Informed Neural Networks and multimodal scans. *Med. Image Anal.* **101**, 103423 (2025).
84. Elmarakeby, H. A. et al. Biologically informed deep neural network for prostate cancer discovery. *Nature* **598**, 348–352 (2021).
85. Lagergren, J. H., Nardini, J. T., Baker, R. E., Simpson, M. J. & Flores, K. B. Biologically-informed neural networks guide mechanistic modeling from sparse experimental data. *PLOS Comput. Biol.* **16**, e1008462 (2020).
86. Witowski, J. et al. Improving breast cancer diagnostics with deep learning for MRI. *Sci. Transl. Med.* **14**, eabo4802 (2022).
87. Meyer-Base, A. et al. AI-enhanced diagnosis of challenging lesions in breast MRI: a methodology and application primer. *J. Magn. Reson. Imaging* **54**, 686–702 (2021).
88. Rosenkrantz, A. B. & Taneja, S. S. Radiologist, be aware: ten pitfalls that confound the interpretation of multiparametric prostate MRI. *Am. J. Roentgenol.* **202**, 109–120 (2014).
89. Cao, R. et al. Joint prostate cancer detection and Gleason score prediction in mp-MRI via FocalNet. *IEEE Trans. Med. Imaging* **38**, 2496–2506 (2019).
90. Bashkanov, O. et al. Automatic detection of prostate cancer grades and chronic prostatitis in biparametric MRI. *Comput. Methods Prog. Biomed.* **239**, 107624 (2023).
91. Wang, J. Y. et al. Stratified assessment of an FDA-cleared deep learning algorithm for automated detection and contouring of metastatic brain tumors in stereotactic radiosurgery. *Radiat. Oncol.* **18**, 61 (2023).
92. Jiang, Y., Edwards, A. V. & Newstead, G. M. Artificial Intelligence applied to breast MRI for improved diagnosis. *Radiology* **298**, 38–46 (2021).
93. Khan, B. et al. Drawbacks of artificial intelligence and their potential solutions in the healthcare sector. *Biomed. Mater. Devices* **1**, 731–738 (2023).
94. Pati, S. et al. Federated learning enables big data for rare cancer boundary detection. *Nat. Commun.* **13**, 7346 (2022).
95. Cortes, C. & Vapnik, V. Support-vector networks. *Mach. Learn.* **20**, 273–297 (1995).
96. Gutman, D. A. et al. MR imaging predictors of molecular profile and survival: multi-institutional study of the TCGA Glioblastoma data set. *Radiology* **267**, 560–569 (2013).
97. Emblem, K. E. et al. A generic support vector machine model for Preoperative Glioma Survival Associations. *Radiology* **275**, 228–234 (2015).
98. Cox, D. R. Regression models and life-tables. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **34**, 187–202 (1972).
99. Lee, H. W. et al. Novel multiparametric magnetic resonance imaging-based deep learning and clinical parameter integration for the prediction of long-term biochemical recurrence-free survival in prostate cancer after Radical Prostatectomy. *Cancers* **15**, 3416 (2023).
100. Ishwaran H., Kogalur U. B., Blackstone E. H. & Lauer M. S. Random survival forests. *Ann. Appl. Stat.* **2**, <https://doi.org/10.1214/08-AOAS169> (2008).
101. Xu, R. & Wunsch, D. C. Clustering algorithms in biomedical research: a review. *IEEE Rev. Biomed. Eng.* **3**, 120–154 (2010).
102. Andersen, E. K. F., Kristensen, G. B., Lyng, H. & Malinen, E. Pharmacokinetic analysis and k-means clustering of DCEMR images for radiotherapy outcome prediction of advanced cervical cancers. *Acta Oncol.* **50**, 859–865 (2011).
103. Anwar, S. M. et al. Medical image analysis using convolutional neural networks: a review. *J. Med. Syst.* **42**, 226 (2018).
104. Jang, B. S., Jeon, S. H., Kim, I. H. & Kim, I. A. Prediction of Pseudoprogression versus progression using machine learning algorithm in Glioblastoma. *Sci. Rep.* **8**, 12516 (2018).
105. Zhou, X. et al. Attention mechanism based multi-sequence MRI fusion improves prediction of response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *Radiat. Oncol.* **18**, 175 (2023).
106. Vaswani A., et al. Attention is All you Need. In: Guyon I., Luxburg U. V., Bengio S., et al., eds. *Advances in Neural Information Processing Systems*. Vol 30. Curran Associates, Inc.; [https://proceedings.neurips.cc/paper\\_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf). (2017).
107. Macyszyn, L. et al. Imaging patterns predict patient survival and molecular subtype in glioblastoma via machine learning techniques. *Neuro-Oncol.* **18**, 417–425 (2016).



108. Kickingeder, P. et al. Radiomic profiling of Glioblastoma: Identifying an imaging predictor of patient survival with improved performance over established clinical and radiologic risk models. *Radiology* **280**, 880–889 (2016).
109. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**, 1735–1780 (1997).
110. Van Houdt, G., Mosquera, C. & Nápoles, G. A review on the long short-term memory model. *Artif. Intell. Rev.* **53**, 5929–5955 (2020).
111. Li, H. et al. Machine learning in prostate MRI for prostate cancer: current status and future opportunities. *Diagnostics* **12**, 289 (2022).
112. Khan, N., Adam, R., Huang, P., Maldjian, T. & Duong, T. Q. Deep learning prediction of pathologic complete response in breast cancer using MRI and other clinical data: a systematic review. *Tomography* **8**, 2784–2795 (2022).
113. Cè M., et al. Artificial intelligence in breast cancer imaging: risk stratification, lesion detection and classification, treatment planning and prognosis—a narrative review. *Explor Target Anti-Tumor Ther.* Published online December **27**, 795–816. (2022)
114. Zhu, M. et al. Artificial intelligence in the radiomic analysis of glioblastomas: A review, taxonomy, and perspective. *Front Oncol.* **12**, 924245 (2022).
115. Zhang, B., Shi, H. & Wang, H. Machine learning and AI in cancer prognosis, prediction, and treatment selection: a critical approach. *J. Multidiscip. Health* **16**, 1779–1791 (2023).
116. Papadimitroulas, P. et al. Artificial intelligence: Deep learning in oncological radiomics and challenges of interpretability and data harmonization. *Phys. Med.* **83**, 108–121 (2021).
117. Marzi, C. et al. Efficacy of MRI data harmonization in the age of machine learning: a multicenter study across 36 datasets. *Sci. Data* **11**, 115 (2024).
118. Basu, K., Sinha, R., Ong, A. & Basu, T. Artificial intelligence: How is it changing medical sciences and its future? *Indian J. Dermatol* **65**, 365 (2020).
119. Rauniyar, A. et al. Federated learning for medical applications: a taxonomy, current trends, challenges, and future research directions. *IEEE Internet Things J.* **11**, 7374–7398 (2024).
120. Guan, H., Yap, P. T., Bozoki, A. & Liu, M. Federated learning for medical image analysis: A survey. *Pattern Recognit.* **151**, 110424 (2024).
121. Andaloussi M. A., Maser R., Hertel F., Lamoline F. & Husch A. D. Exploring Adult Glioma through MRI: A Review of Publicly Available Datasets to Guide Efficient Image Analysis. Published online. <https://doi.org/10.48550/ARXIV.2409.00109> (2024).
122. Sunoqrot, M. R. S., Saha, A., Hosseinzadeh, M., Elschot, M. & Huisman, H. Artificial intelligence for prostate MRI: open datasets, available applications, and grand challenges. *Eur. Radio. Exp.* **6**, 35 (2022).
123. Kilintzis, V. et al. Public data homogenization for AI model development in breast cancer. *Eur. Radio. Exp.* **8**, 42 (2024).
124. Kosvira, A., Filos, D. T., Fotopoulos, D. T. H., Tsavre, O. & Chouvarda, I. Toward ensuring data quality in multi-site cancer imaging repositories. *Information* **15**, 533 (2024).
125. Bedard, P. L., Hansen, A. R., Ratain, M. J. & Siu, L. L. Tumour heterogeneity in the clinic. *Nature* **501**, 355–364 (2013).
126. Booth, C. M., Karim, S. & Mackillop, W. J. Real-world data: towards achieving the achievable in cancer care. *Nat. Rev. Clin. Oncol.* **16**, 312–325 (2019).
127. Yuan, Y. et al. Prostate cancer classification with multiparametric MRI transfer learning model. *Med. Phys.* **46**, 756–765 (2019).
128. Han, W. et al. Deep transfer learning and radiomics feature prediction of survival of patients with high-grade Gliomas. *Am. J. Neuroradiol.* **41**, 40–48 (2020).
129. Deng J., et al. ImageNet: A large-scale hierarchical image database. In: 2009 *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE; 248–255 (2009).
130. Wen, Y., Chen, L., Deng, Y. & Zhou, C. Rethinking pre-training on medical imaging. *J. Vis. Commun. Image Represent.* **78**, 103145 (2021).
131. Zhang, Y. et al. Prediction of breast cancer molecular subtypes on DCE-MRI using convolutional neural network with transfer learning between two centers. *Eur. Radio.* **31**, 2559–2567 (2021).
132. US Food and Drug Administration. *Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices*. <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-ai-ml-enabled-medical-devices>. (2024).
133. Eche, T., Schwartz, L. H., Mokrane, F. Z. & Dercle, L. Toward generalizability in the deployment of artificial intelligence in radiology: role of computation stress testing to overcome underspecification. *Radio. Artif. Intell.* **3**, e210097 (2021).
134. Futoma, J., Simons, M., Panch, T., Doshi-Velez, F. & Celi, L. A. The myth of generalisability in clinical research and machine learning in health care. *Lancet Digit Health* **2**, e489–e492 (2020).
135. Adjeiwaah, M., Garpebring, A. & Nyholm, T. Sensitivity analysis of different quality assurance methods for magnetic resonance imaging in radiotherapy. *Phys. Imaging Radiat. Oncol.* **13**, 21–27 (2020).
136. Yang Y., Zhang H., Gichoya J. W., Katabi D. & Ghassemi, M. The limits of fair medical imaging AI in real-world generalization. *Nat. Med.* Published online June 28, <https://doi.org/10.1038/s41591-024-03113-4> (2024).
137. Albahri, A. S. et al. A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion. *Inf. Fusion* **96**, 156–191 (2023).
138. Linardatos, P., Papastefanopoulos, V. & Kotsiantis, S. Explainable AI: A review of machine learning interpretability methods. *Entropy* **23**, 18 (2020).
139. Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R. & Yu, B. Definitions, methods, and applications in interpretable machine learning. *Proc. Natl. Acad. Sci.* **116**, 22071–22080 (2019).
140. Banegas-Luna, A. J. et al. Towards the interpretability of machine learning predictions for medical applications targeting personalised therapies: a cancer case survey. *Int. J. Mol. Sci.* **22**, 4394 (2021).
141. Kuenzi, B. M. et al. Predicting drug response and synergy using a deep learning model of human cancer cells. *Cancer Cell* **38**, 672–684.e6 (2020).
142. Wysocka, M., Wysocki, O., Zufferey, M., Landers, D. & Freitas, A. A systematic review of biologically-informed deep learning models for cancer: fundamental trends for encoding and interpreting oncology data. *BMC Bioinforma.* **24**, 198 (2023).
143. Cuomo, S. et al. Scientific machine learning through physics-informed neural networks: where we are and what's next. *J. Sci. Comput* **92**, 88 (2022).
144. Jiang, Y. et al. Biology-guided deep learning predicts prognosis and cancer immunotherapy response. *Nat. Commun.* **14**, 5135 (2023).
145. White House Office of Science and Technology Policy. *Blueprint for an AI Bill of Rights*. <https://www.whitehouse.gov/ostp/ai-bill-of-rights/> (2022).

## Acknowledgements

We thank the National Cancer Institute for funding through 1R01CA260003, U24CA226110, U01CA253540, and P30CA016672. We thank the Cancer Prevention and Research Institute of Texas for funding through RP220225. We thank the National Science Foundation for funding through CCF-2239687 (CAREER), IFML 2019844, AF 1901292. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DE2137420. This project was supported by the Resources of the Image Guided Cancer Therapy Research Program at The University of Texas MD Anderson Cancer Center as well as the Joint Center for Computational Oncology (a collaboration between the



Oden Institute for Computational Engineering and Sciences, The University of Texas MD Anderson Cancer Center, and Texas Advanced Computing Center). The project was also supported by the Tumor Measurement Initiative at The University of Texas MD Anderson Cancer Center. We acknowledge the support from “la Caixa” Foundation (ID 100010434), fellowship code is LCF/BQ/PI23/11970033.

### Author contributions

C.W. – Project design, writing—original draft, writing—review and editing. M.A.A. – Project design, writing—original draft, writing—review and editing. D.A.H. – Project design, funding acquisition, investigation, methodology, writing—original draft, writing—review and editing. E.L. – Project design, investigation, methodology, writing—original draft, writing—review and editing. G.L. – Project design, investigation, methodology, writing—original draft, writing—review and editing, visualization. C.E.S. – Investigation, methodology, writing—original draft, writing—review and editing, visualization. S.R. – Visualization, writing—original draft, writing—review and editing, B.L. – Visualization, writing—original draft, writing—review and editing, A.G.D. – Conceptualization, writing—review and editing J.I.T. – Project design, funding acquisition, investigation, methodology, visualization, writing—original draft, writing—review and editing. K.K.B. – Conceptualization, writing—review and editing. C.C. – Conceptualization, writing—review and editing. T. E.Y. – Conceptualization, project design, funding acquisition, investigation, methodology, writing—original draft, writing—review and editing.

### Competing interests

Caroline Chung – Research funding to the institution from Siemens Healthineers and RaySearch Laboratories

### Additional information

**Correspondence** and requests for materials should be addressed to Thomas E. Yankeelov.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

<sup>1</sup>Department of Imaging Physics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. <sup>2</sup>Department of Breast Imaging, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. <sup>3</sup>Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. <sup>4</sup>Institute for Data Science in Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. <sup>5</sup>Oden Institute for Computational Engineering and Sciences, The University of Texas at Austin, Austin, TX, USA. <sup>6</sup>Faculty of Science, Technology and Medicine University of Luxembourg, Belvaux, Luxembourg. <sup>7</sup>Livestrong Cancer Institutes, The University of Texas at Austin, Austin, TX, USA. <sup>8</sup>Texas Advanced Computing Center, The University of Texas at Austin, Austin, TX, USA. <sup>9</sup>Health Research Institute of Santiago de Compostela, Santiago de Compostela, Spain. <sup>10</sup>Chandra Family Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX, USA. <sup>11</sup>Department of Diagnostic Medicine, The University of Texas at Austin, Austin, TX, USA. <sup>12</sup>Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. <sup>13</sup>Department of Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. <sup>14</sup>Department of Neuroradiology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. <sup>15</sup>Department of Biomedical Engineering, The University of Texas at Austin, Austin, TX, USA. <sup>16</sup>Department of Oncology, The University of Texas at Austin, Austin, TX, USA. ✉e-mail: [tey@utexas.edu](mailto:tey@utexas.edu)