Artificial Intelligence in Radiation Oncology: A Review of its Current Status and Potential Application for the Radiotherapy Workforce

Author names

Dr Craig Parkinson[a,*]

Mrs Catherine Matthams[b,*]

Dr Kieran Foley[b,#]

Dr Emiliano Spezi[a,#]

Affiliation

[a]School of Engineering, Cardiff University

[b]Velindre Cancer Centre

[*]Joint first author

[#]Joint last author

Declaration of interests : None

Artificial Intelligence in Radiation Oncology: A Review of its Current Status and Potential Application for the Radiotherapy Workforce.

Abstract

Radiation oncology is a continually evolving speciality. With the development of new imaging modalities and advanced imaging processing techniques, there is an increasing amount of data available to practitioners. Artificial intelligence (AI), has the potential to harness the availability of this data for improving patient outcomes, reducing toxicity, and easing clinical burdens. Within this review, we highlight both the potential utility of AI in radiation oncology and its potential problems, including the requirement of complexity of data, undefined core outcomes, and limited generalisability. Overcoming these problems, requires collaboration between AI experts and the radiotherapy workforce, the development of educational resources, and standardised reporting of AI studies to allow comparison of clinical acceptability between techniques.

AI in radiation oncology

In the UK, there are approximately 367,000 new cancers diagnosed every year. The incidence has increased by 12-16% since the 1990s and continues to rise [1]. Medical imaging has an increasingly vital role in the management of a patients' cancer. A greater incidence of cancer diagnoses means there is an increasing amount of imaging data available to clinicians [2]. The amount of data will continue to increase because advanced treatment options, such as adaptive radiotherapy (RT) planning [3]–[7], require a greater volume of image-guided planning. The availability of data is further increased by the quantitative mining of imaging [8]–[11]. The creation of linked data combined with artificial intelligence (AI) can profoundly change radiation oncology [12], with the potential to democratise the field [13].

AI is a subfield of computer science that concerns itself with the development of autonomous systems which replicate human functions [14]. Specifically, machine learning, a branch of AI and the focus of this article, is the process of solving practical problems by developing statistical models from a retrospectively gathered dataset [15]. These statistical models self-learn "knowledge" from collated data to make predictions [16]. Within machine learning sits deep learning which is the imitation of the workings of the human brain for use in decision making [17], via neurons in neural networks. Each neuron in the network learns a simple function, and the overall (more complex) function is defined by combining these simpler functions [18].

Within this review, AI is used as a reference to machine learning, and its potential, along with current problems in the field of Oncology, are considered from a technical position. The review aims to highlight these considerations to the radiotherapy workforce, particularly therapeutic radiographers, as there will be an increasing requirement for their familiarity within this area due to their unique position as the interface between imaging technology and patients.

The potential for AI

Easing clinical burdens

Advances in AI have led to an increase in efforts to automate tasks routinely undertaken by human observers. Automation has the potential to benefit patients through decreased costs, increased efficiency and a reduction in errors [19]. As an example, conformal brain radiotherapy (RT) planning can take between 2-4 hours [20]. Planning radiographers are proficient to undertake the contouring of Organs at Risk (OAR), whereas Advanced Practitioner and Consultant radiographers have a defined skillset for Gross Tumour Volume (GTV) contouring. Whilst OAR contouring is not subject to the same level of inter-and intra-observer variability as GTV or Metabolic Tumour Volume (MTV) delineation, both are time-consuming [21], [22]. Therefore, automating administrative and routine clinical tasks, such as OAR contouring, can reduce this clinical burden allowing for re-distribution of valuable healthcare resources. A further application of AI is the prioritisation of time-critical dependent scenarios aimed at improving patient outcomes [13], [23] and freeing up time for patients and human-centric tasks with the greatest need [24]. To give an example, AI has been implemented to prioritise patients at risk of developing colon cancer [25] and has been used for the segmentation of OAR [26].

Patient selection for clinical trials

The de-centralised nature of the NHS and inherent separation from academia [27] results in a disconnect between the clinical and research environments. Once clinical trials are open to recruitment, clinicians and research radiographers must actively screen to identify individual patients who match the inclusion and exclusion criteria. However, the development of a research picture archiving and communication system (PACS) [27], combined with AI techniques, has the potential to provide those recruiting with the ability to search, select and identify patients automatically. This would not only permit a more efficient and effective workflow, but would allow for more accurate and robust patient selection processes, therefore boosting recruitment into clinical trials. Federated learning is a machine learning technique that trains an algorithm across multiple centres, which are potentially geographically decentralised. Each centre holds localised data samples, and the learning algorithm travels to each centre which can be accomplished by a personalised health train (PHT) network infrastructure [28], [29]. A PHT is designed for communicating algorithms and results between centres [30] instead of being dependent on busy clinicians and radiographers to identify patients for currently available clinical trials.

A further advantage of research PACS is that federated approaches could be used to train and validate AI technologies. The increased availability of standardised data from a curated, anonymised, duplicate PACS combining clinical data, imaging, biological and outcome data from multiple centres would improve the generalisability of AI technologies. Linked data can generate new "knowledge" [12]. For example, Deist et al [28] trained a logistical regression model, on more than 20,000 patients to predict post-treatment two-year survival in lung cancer. The use of a PHT infrastructure overcame privacy concerns with data sharing and enabled data analysis from multiple centres in different countries.

Automated adaptation & optimisation of radiotherapy

AI also has the potential to automate RT planning, helping to remove variability within a patient's treatment. As previously mentioned, contouring [31], [32] can be automated using deep learning. Alternatively, RT plans can be automatically created by planning radiographers, using clustering analysis of previously defined plans or by using Pareto-guided navigation techniques [33], [34]. Pareto guided navigation allows for clinical experience to be considered in automated planning and the exploration of different 'trade-off' options. This is accomplished by a user manually assigning weights (importance) to each of the trade-off criteria which is specific to each cancer site. Further to this, the difficulty of adapting RT plans during treatment can be eased. An example of this is, intra-treatment auto-contouring [3], [35], [36] after chemotherapy and the adaptive RT planning of image-guided radiation therapy (IGRT) is being investigated within the PEARL clinical trial [37], [38]. AI-based auto contouring of the MTV, GTV and OAR is viable across a variety of imaging modalities and tumour sites [35], [36], [39]–[46].

Image-guided radiation therapy relies upon the acquisition of computed tomography (CT) and repeated cone-beam computed tomography (CBCT) imaging. CBCT & CT images are typically acquired by therapeutic radiographers using low-dose protocols, producing images with a low signal to noise ratio. However, Huber et al [47] have demonstrated that AI has the potential to improve image quality by automating noise removal from low-dose CT images. These techniques have the potential to optimise image resolution and signal to noise ratios whilst still delivering a low dose to the patient, therefore providing an additional layer of safety. AI technologies have already demonstrated reduced acquisition times for Magnetic Resonance Imaging (MRI) scans, and the ability to decrease CT and Positron Emission Tomography (PET) radiation doses [24]. AI techniques can also be used to alleviate imbalances within data distributions whilst training other AI models. For example, by generating synthetic images from an implicit distribution that follows real data distributions [48]. Kitchen et al, demonstrated the efficacy of generative adversarial networks for synthesising MRI prostate lesions [49] and Han et al demonstrated the efficacy of MRI synthesis using visual Turing tests [50].

Toxicity prediction

One of the goals of RT is to maximise the dose to the target whilst minimising radiation dose to the OAR to reduce short- and long-term toxicity. Short-term toxicity generally results in full recovery within weeks or months after completion of treatment, but predicting late effects, which are considered irreversible and progressive, could lead to improved treatment decision support.

Deist et al investigated the use a variety of machine learning techniques for toxicity prediction across brain, lung, and head & neck (H&N) cancer. They collated 12 datasets, consisting of 3496 patients, containing a combination of clinical, dosimetric and blood features. They found that no technique provided optimal performance across all datasets [51]. However, the empirical selection of a classifier led to a 2% improvement in the area under the curve (AUC). The AUC is derived from the Receiver Operating Characteristic (ROC) curve which plots true positive rate versus false positive rate. ROC and AUC provide very useful insight into model discrimination performance and are widely used for interpreting and comparing prediction models. Perfect prediction models have an AUC of 1, whereas models providing random predictions have an AUC of 0.5. Deist et al also investigated dosimetry and blood marker data, whilst in a separate patient cohort, Saednia et al [52] investigated body-surface temperature for predicting toxicity, producing an AUC of 0.87. Saednia et al, recruited 90 patients with thermal images acquired before RT and then weekly at fractions 5, 10 and 15. In comparison, Reddy et al [53] produced AUCs of 0.85, 0.82, 0.77 and 0.56 when predicting radiation dermatitis, moist desquamation, breast/chest wall pain and fatigue, in a prospective validation dataset of breast cancer patients. In oesophageal cancer [54], machine learning on 101 patients with PET/CT imaging produced a peak AUC of 0.63 for predicting radiation pneumonitis.

The different evaluation endpoints and target use problems of all the developed models mean that comparing model performance is challenging. The review by Isaksson et al [55] highlights that AI models have shown a high level of performance for predicting toxicity. However, the clinically limiting factor is the ability of a person to understand the cause of the decision [56]. The higher the interpretability of the model, the easier it is for someone to comprehend the decisions being made. It also highlights the disparity in the training datasets used for developing toxicity prediction-based AI models, with datasets ranging from sample sizes of 35 to more than 2000 patients, and a range of heterogeneous retrospective and prospective internal and external validation datasets used to assess models.

Current problems in AI for Oncology

Whilst AI has significant potential, there are currently several pitfalls that limit its implementation in the clinical environment. For example, significant errors have been reported for generating synthetic imaging using AI techniques. Wang et al [57] report that the mean absolute error (MAE) when generating synthetic CTs from MRI ranges from 40 to 70 Hounsfield Units (HU). However, the MAE of bone and air is more than 100 HU due to their indistinguishable contrast on MRI. Wang et al [57] also highlight that reported errors include the misalignment between CT and MR images, these errors propagate through training and lead to overestimations of the actual error. In PET imaging, the underestimation of the uptake in a lesion may result in misidentification and evaluation of treatment response [58].

Many AI models have been published [40], [42], [43], [59]–[63], for a variety of tasks including patient risk stratification, GTV, MTV, Clinical Target Volume (CTV), Planned Target Volume (PTV) and OAR delineation. However, the vast majority are not open-source or publicly available, making it difficult to externally validate them. Assessing AI model performance is further challenged by the requirement of high expertise levels to adjust the model configuration, with small errors leading to decreases in model performance [64]. Figure 1 shows the typical pathway to train and test an AI model and demonstrates some of the complex decisions that are required at each step, including data collection & processing, AI model tuning and validation. These decisions are typically based upon the designer's experience. The difficulty in model assessment has resulted in increasing interest, by researchers, in international competitions for assessing the performance of algorithms and models [65]. Also, the applicability of an AI model to the initial problem of an end-user is often limited [66], which highlights the essential involvement of the end-user at all stages of design. Additionally, the design and quality assurance, including the level of reporting is highly varied within these events. These challenges have led to the development of the best practice guidelines "Checklist for Artificial Intelligence in Medical Imaging" (CLAIM) [67] which aims to provide a standardised framework for reviewers and authors who report on AI technologies.

Reporting from international competitions is typically physics-based and uses quantitative metrics. When developed AI models generate similar results, the question of which AI technology is superior is raised. Reporting quantitative metrics may not assess the potential clinical utility of the technology. Therefore, there is a need for quantitative and qualitative assessment methodologies, as demonstrated by studies investigating modified Turing tests [62], [68]. First, proposed in the 1950s as the imitation game, a Turing test aims to answer the question of whether machines can think. By testing a machine's ability to exhibit behaviour equivalent to, or indistinguishable from that of a human [69]. Within the fields of radiology and radiation oncology, this has been achieved by asking viewers to assess the quality of contours and whether a contour has been drawn 'by a machine' or 'by a human' [62], [68]. If a viewer is unable to determine the difference, the machine/algorithm is considered to have passed the test.

A further barrier to the adoption of AI in the clinical setting is that AI is often considered to be a black box, meaning that clinicians, radiologists, radiographers, other medical professionals, and the public are not able to verify and fully trust the algorithm. This leads to concerns that incorrect data can be inputted into the system, generating inaccurate results. By combining different AI technologies, e.g. deep-learning combined with classical machine-learning, we can potentially improve the traceability of decisions made by these technologies [70]. It is important to recognise that malfunctioning AI systems can create new safety hazards, making accurate inputs as important as the accuracy of the algorithm itself. It is therefore vital that health and care records are integrated and digitised to enhance data integrity and curation [24]. This includes the combination of patient imaging, genomic data, and patient records.

In 2018, Doe et al [71] discussed how the Digital Imaging and Communication in Medicine (DICOM) standard provided different annotation mechanisms for encoding, transporting and querying imaging results. These mechanisms relied heavily on human guidance. Whereas AI requires semantically meaningful payloads and interactions. To improve data integrity, tools provided within the DICOM standard (e.g., the DICOM Segmentation Image Module DICOM-Seg) can be used for generating annotations in a meaningful way. DICOM-Seg provides anatomical labelling stored with the rasterised segmentation, enabling curated nomenclature of structures, including the GTV, CTV, PTV and OAR, thereby allowing data curation to occur as part of routine clinical practice. However, the availability of these mechanisms is reliant upon uptake by clinicians and implementation by vendors [72]. Figure 2 demonstrates coded semantics of segmented regions, including anatomical location, type of segmentation, the algorithm used for segmentation and visualisation parameters (colour). The MTV, GTV, CTV and PTV outlined include the mandible, oral cavity, parotid glands, spinal cord, and external outline of a patient with head & neck cancer.

Additionally, those working with AI must be trained on its expected function and therefore be able to anticipate its dysfunction. A vital component of using AI is the availability of override systems. For example, having the ability to manually alter the MTV or GTV contours. It is also vital these override systems are intuitive to use. Furthermore, it has been highlighted that regulation may not provide sufficient protection to patients, especially when conflicts of interest exist [19]. Therefore, AI technologies must embody the principle of medical care: ‘do no harm’, with practitioners maintaining vigilance or validation of advice received from AI technologies [24]. Whilst the field of AI in radiation oncology, is battling these pitfalls and potential problems, AI still has the potential to dramatically improve workflows and processes. The question has changed from will AI replace clinical practitioners, to how can practitioners use AI to benefit patients? [73].

Discussion.

AI has great potential in radiation oncology for toxicity prediction, automated RT planning and optimisation, clinical trial patient selection, and easing of clinical burdens. However, there is a requirement for resources to educate and train the radiotherapy workforce, in data provenance, curation, and integration as well as the ethics of AI development and interpretation of AI technologies [24]. The right level of expertise could be achieved, as an example, by introducing the principles of clinical data science and AI to therapeutic radiographers at an earlier stage within their careers. Further, providing ongoing continued professional development would allow for the integration of knowledge between the profession and AI experts. This will help to reduce the black-box effect of AI, therefore increasing trust within the technologies. Additionally, providing these training resources will help radiologists and clinicians in identifying additional areas in which AI can potentially be used. The application of these technologies can help enable clinicians to harness patterns within local, regional, and global populations which are not necessarily visible to the human eye [8], [17]. However, the patient benefit should be the driving factor behind these technologies, with patients being involved from the design process to the implementation of technologies, embodying the principle of ‘do no harm’ [24].

Uniform reporting of challenge events and technical papers is also required. Conformance to CLAIM guidance [67] can assist the radiotherapy workforce to compare AI technologies. Furthermore, there is a significant need to assess the clinical utility of these technologies through prospective [55] and qualitative studies [74]. Standardised endpoints should also be evaluated between AI studies, a common scenario in clinical studies using patient outcomes. Whilst AI still faces significant challenges, such as lack of generalisability, small datasets, and limited prospective validation studies, there is hope for the eventual clinical adoption of these technologies. Netherton et al [75], using hype curves, predict 20% of clinics will be using AI within the next few years. However, this will require transparent and standardised reporting of AI model performance, commercial vendor integration of technologies with current NHS information technology systems and improved connectivity between healthcare providers [75].

Conclusion

AI continues to have the potential to harness the increasing amount of data available to clinical practitioners. However, there is still a significant disconnect between the availability of educational resources for healthcare professionals, the ethics of AI development and the interpretation of AI technologies.

Declaration of interests: None
Funding Source: This work was funded by Cancer Research Wales

References

[1] Cancer Research UK, “Cancer incidence for all cancers combined,” *Cancer Research UK*. [Online]. Available: https://www.cancerresearchuk.org/health-professional/cancer-statistics/incidence/all-cancers-combined. [Accessed: 05-Jan-2021].

[2] T. Lustberg *et al.*, “Big Data in radiation therapy: Challenges and opportunities,” *Br. J. Radiol.*, vol. 90, no. 1069, 2017.

[3] S. Hargreaves *et al.*, “Interim 18F-FDG Positron Emission Tomography/Computed Tomography During Chemoradiotherapy in the Management of Cancer Patients: a Response,” *Clin. Oncol.*, vol. 31, no. 9, pp. 669–670, Sep. 2019.

[4] K. K. Brock, “Adaptive Radiotherapy: Moving Into the Future,” *Semin. Radiat. Oncol.*, vol. 29, no. 3, pp. 181–184, Jul. 2019.

[5] M. Vanderhoek, S. B. Perlman, and R. Jeraj, “Impact of the Definition of Peak Standardized Uptake Value on Quantification of Treatment Response,” *J. Nucl. Med.*, vol. 53, no. 1, pp. 4–11, 2012.

[6] S. B. Mali, “Adaptive Radiotherapy for Head Neck Cancer,” *J. Maxillofac. Oral Surg.*, vol. 15, no. 4, pp. 549–554, Dec. 2016.

[7] M. Antico *et al.*, “Real-time adaptive planning method for radiotherapy treatment delivery for prostate cancer patients, based on a library of plans accounting for possible anatomy configuration changes,” *PLoS One*, vol. 14, no. 2, p. e0213002, Feb. 2019.

[8] R. J. Gillies, P. E. Kinahan, and H. Hricak, “Radiomics: Images Are More than Pictures, They Are Data,” *Radiology*, vol. 278, no. 2, pp. 563–577, Feb. 2016.

[9] A. Zwanenburg *et al.*, “The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping,” *Radiology*, vol. 295, no. 2, pp. 328–338, May 2020.

[10] A. Zwanenburg *et al.*, “Results from the Image Biomarker Standardisation Initiative,” in *ESTRO 37*, 2018.

[11] A. Depeursinge *et al.*, “Standardised convolutional filtering for radiomics,” *Arxiv*, no. June,

2020.

[12] J. Kazmierska *et al.*, “From multisource data to clinical decision aids in radiation oncology: The need for a clinical data science community,” *Radiother. Oncol.*, vol. 153, pp. 43–54, 2020.

[13] S. D. O’Connor and M. Bhalla, “Should Artificial Intelligence Tell Radiologists Which Study to Read Next?,” *Radiol. Artif. Intell.*, vol. 3, no. 2, p. e210009, Mar. 2021.

[14] R. Pfeifer and C. Scheier, *Understanding Intelligence*. The MIT Press, 2001.

[15] A. Burkov, *The hundred-page machine learning book*. Andriy Burkov, 2019.

[16] S. Raschka and V. Mirjalili, *Python Machine Learning*, Third Edit. Packt Publishing, 2019.

[17] A. Hosny, C. Parmar, J. Quackenbush, L. H. Schwartz, and H. J. W. L. Aerts, “Artificial intelligence in radiology,” *Nat. Rev. Cancer*, vol. 18, no. 8, pp. 500–510, Aug. 2018.

[18] J. Kelleher, *Deep Learning*. The MIT Press, 2019.

[19] J. Mongan and M. Kohli, “Artificial Intelligence and Human Life: Five Lessons for Radiology from the 737 MAX Disasters,” *Radiol. Artif. Intell.*, vol. 2, no. 2, p. e190111, Mar. 2020.

[20] L. S. Chou, C. Y. Lin, Y. H. Wu, J. H. Ji, J. A. Liang, and T. H. Wang, “A Simple and Automated VMAT Treatment Planning for Hippocampal Avoidance Whole Brain Radiotherapy,” *Int. J. Radiat. Oncol.*, vol. 108, no. 3, p. e335, 2020.

[21] H. Tang *et al.*, “Clinically applicable deep learning framework for organs at risk delineation in CT images,” *Nat. Mach. Intell.*, vol. 1, no. 10, pp. 480–491, Oct. 2019.

[22] H. Zaidi and I. El Naqa, “PET-guided delineation of radiation therapy treatment volumes: a survey of image segmentation techniques,” *Eur. J. Nucl. Med. Mol. Imaging*, vol. 37, no. 11, pp. 2165–2187, Nov. 2010.

[23] T. J. O’Neill *et al.*, “Active Reprioritization of the Reading Worklist Using Artificial Intelligence Has a Beneficial Effect on the Turnaround Time for Interpretation of Head CT with Intracranial Hemorrhage,” *Radiol. Artif. Intell.*, vol. 3, no. 2, p. e200024, Mar. 2021.

[24] E. Topol, “Preparing the healthcare workforce to deliver the digital future The Topol Review. An independent report on behalf of the Secretary of State for Health and Social Care,” *Nhs*, no. February, p. 102, 2019.

[25] M. Downing, “Barts Health using AI to prioritise care for colon cancer patients,” 2020. [Online]. Available: https://www.bartshealth.nhs.uk/news/barts-health-using-ai-to-prioritise-care-for-high-risk-colon-cancer-patients-8867. [Accessed: 29-Mar-2021].

[26] L. Kratzke *et al.*, “DirectORGANS 2.0.” [Online]. Available: https://www.siemens-healthineers.com/en-uk/radiotherapy/ct-for-rt/somatom-go-sim. [Accessed: 13-Apr-2021].

[27] S. J. Doran *et al.*, “Informatics in Radiology: Development of a Research PACS for Analysis of Functional Imaging Data in Clinical Research and Clinical Trials,” *RadioGraphics*, vol. 32, no. 7, pp. 2135–2150, Nov. 2012.

[28] T. M. Deist *et al.*, “Distributed learning on 20 000+ lung cancer patients – The Personal Health Train,” *Radiother. Oncol.*, vol. 144, pp. 189–200, 2020.

[29] T. Skripcak *et al.*, “Creating a data exchange strategy for radiotherapy research: Towards federated databases and anonymised public datasets,” *Radiother. Oncol.*, vol. 113, no. 3, pp. 303–309, 2014.

[30] J. Van Soest *et al.*, “Using the personal health train for automated and privacy-preserving analytics on vertically partitioned data,” *Stud. Health Technol. Inform.*, vol. 247, pp. 581–585, 2018.

[31] S. K. Vinod, M. Min, M. G. Jameson, and L. C. Holloway, “A review of interventions to reduce inter-observer variability in volume delineation in radiation oncology,” *J. Med. Imaging Radiat. Oncol.*, vol. 60, no. 3, pp. 393–406, 2016.

[32] Z. Liu *et al.*, “Development and Validation of a Deep Learning Algorithm for Auto-delineation of Clinical Target Volume and Organs at Risk in Cervical Cancer Radiotherapy,” *Radiother. Oncol.*, 2020.

[33] P. A. Wheeler *et al.*, “Evaluating the application of Pareto navigation guided automated radiotherapy treatment planning to prostate cancer,” *Radiother. Oncol.*, vol. 141, pp. 220–226, 2019.

[34] P. A. Wheeler *et al.*, “Utilisation of Pareto navigation techniques to calibrate a fully automated radiotherapy treatment planning solution,” *Phys. Imaging Radiat. Oncol.*, vol. 10, no. January, pp. 41–48, 2019.

[35] C. Parkinson *et al.*, “Machine-learned target volume delineation of 18 F-FDG PET images after one cycle of induction chemotherapy,” *Phys. Medica*, vol. 61, no. April, pp. 85–93, 2019.

[36] C. Parkinson *et al.*, “EP-1126: Target volume delineation of PET post one cycle of induction chemotherapy in oropharyngeal cancer,” *Radiother. Oncol.*, vol. 127, pp. S634–S635, Apr. 2018.

[37] C. Trials, “PEARL PET-based Adpative Radiotherapy Clinical Trial.” [Online]. Available: https://clinicaltrials.gov/ct2/show/NCT03935672.

[38] BBC, “PEARL Trial.” [Online]. Available: https://www.bbc.co.uk/news/uk-wales-39566782. [Accessed: 19-Feb-2019].

[39] B. Berthon *et al.*, “Head and neck target delineation using a novel PET automatic segmentation algorithm,” *Radiother. Oncol.*, vol. 122, no. 2, pp. 242–247, 2017.

[40] B. Berthon, C. Marshall, M. Evans, and E. Spezi, “ATLAAS: An automatic decision tree-based learning algorithm for advanced image segmentation in positron emission tomography,” *Phys. Med. Biol.*, vol. 61, no. 13, pp. 4855–4869, 2016.

[41] M. Hatt, C. Cheze le Rest, A. Turzo, C. Roux, and D. Visvikis, “A Fuzzy Locally Adaptive Bayesian Segmentation Approach for Volume Determination in PET,” *IEEE Trans. Med. Imaging*, vol. 28, no. 6, pp. 881–893, Jun. 2009.

[42] H. Fayad *et al.*, “The first MICCAI challenge on PET tumor segmentation,” *Med. Image Anal.*, vol. 44, no. December, pp. 177–195, 2017.

[43] M. Hatt *et al.*, “Classification and evaluation strategies of auto-segmentation approaches for PET: Report of AAPM task group No. 211,” *Med. Phys.*, vol. 44, no. 6, pp. e1–e42, May 2017.

[44] Z. Zhong *et al.*, “3D fully convolutional networks for co-segmentation of tumors on PET-CT images,” *Proc. - Int. Symp. Biomed. Imaging*, vol. 2018-April, no. Isbi, pp. 228–231, 2018.

[45] C. G. B. Yogananda *et al.*, “A fully automated deep learning network for brain tumor segmentation,” *Tomography*, vol. 6, no. 2, pp. 186–193, 2020.

[46] P. R. M. Study *et al.*, “A Deep Learning-Based Automated CT Segmentation of Prostate Cancer Anatomy for Radiation Therapy,” 2020.

[47] N. R. Huber, A. D. Missert, and B. J. Erickson, “Magician’s Corner: 7. Using Convolutional Neural Networks to Reduce Noise in Medical Images,” *Radiol. Artif. Intell.*, vol. 2, no. 5, p. e200036, Sep. 2020.

[48] S. Kazeminia *et al.*, “GANs for medical image analysis,” *Artif. Intell. Med.*, vol. 109, p. 101938, Sep. 2020.

[49] A. Kitchen and J. Seah, “Deep Generative Adversarial Neural Networks for Realistic Prostate Lesion MRI Synthesis,” Jul. 2017.

[50] C. Han *et al.*, “GAN-based synthetic brain MR image generation,” in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 2018, pp. 734–738.

[51] T. M. Deist *et al.*, “Machine learning algorithms for outcome prediction in (chemo)radiotherapy: An empirical comparison of classifiers,” *Med. Phys.*, vol. 45, no. 7, pp. 3449–3459, Jul. 2018.

[52] K. Saednia *et al.*, “Quantitative Thermal Imaging Biomarkers to Detect Acute Skin Toxicity From Breast Radiation Therapy Using Supervised Machine Learning,” *Int. J. Radiat. Oncol.*, vol. 106, no. 5, pp. 1071–1083, Apr. 2020.

[53] J. Reddy, W. D. Lindsay, C. G. Berlind, C. A. Ahern, and B. D. Smith, “Applying a Machine Learning Approach to Predict Acute Toxicities During Radiation for Breast Cancer Patients,” *Int. J. Radiat. Oncol.*, vol. 102, no. 3, p. S59, Nov. 2018.

[54] J. P. Hart *et al.*, “Radiation Pneumonitis: Correlation of Toxicity With Pulmonary Metabolic Radiation Response,” *Int. J. Radiat. Oncol.*, vol. 71, no. 4, pp. 967–971, Jul. 2008.

[55] L. J. Isaksson, M. Pepa, M. Zaffaroni, G. Marvaso, and D. Alterio, “Machine Learning-Based Models for Prediction of Toxicity Outcomes in Radiotherapy,” vol. 10, no. June, 2020.

[56] T. Miller, “Explanation in artificial intelligence: Insights from the social sciences,” *Artif. Intell.*, vol. 267, pp. 1–38, Feb. 2019.

[57] T. Wang *et al.*, “A review on medical imaging synthesis using deep learning and its clinical applications,” *J. Appl. Clin. Med. Phys.*, vol. 22, no. 1, pp. 11–36, Jan. 2021.

[58] T. Wang *et al.*, “Machine learning in quantitative PET: A review of attenuation correction and low-count image reconstruction methods,” *Phys. Medica*, vol. 76, pp. 294–306, Aug. 2020.

[59] K. G. Foley *et al.*, “Development and validation of a prognostic model incorporating texture analysis derived from standardised segmentation of PET in patients with oesophageal cancer,” *Eur. Radiol.*, vol. 28, no. 1, pp. 428–436, Jan. 2018.

[60] B. H. Menze *et al.*, “The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS),” *IEEE Trans. Med. Imaging*, vol. 34, no. 10, pp. 1993–2024, 2015.

[61] F. Wang, R. Jiang, L. Zheng, C. Meng, and B. Biswal, “3d u-net based brain tumor segmentation and survival days prediction,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2020, vol. 11992 LNCS, pp. 131–141.

[62] E. Spezi *et al.*, “Metabolic tumour volume segmentation for oesophageal cancer on hybrid PET/CT using convolutional network architecture,” *Eur. J. Nucl. Med. Mol. Imaging*, vol. 47, no. S1, pp. 1–753, Sep. 2020.

[63] S. Bakas *et al.*, “Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS

challenge," *arXiv*, 2018.

[64] L. Maier-Hein *et al.*, "Why rankings of biomedical image analysis competitions should be interpreted with care," *Nat. Commun.*, vol. 9, no. 1, p. 5217, Dec. 2018.

[65] "Challenges," *The Medical Image Computing and Computer Assisted Intervention Society*. [Online]. Available: http://www.miccai.org/events/challenges/. [Accessed: 06-Apr-2021].

[66] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "Automated design of deep learning methods for biomedical image segmentation," *arXiv*, pp. 1–55, 2019.

[67] J. Mongan, L. Moy, and C. E. Kahn, "Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A Guide for Authors and Reviewers," *Radiol. Artif. Intell.*, vol. 2, no. 2, p. e200029, Mar. 2020.

[68] M. J. Gooding *et al.*, "Comparative evaluation of autocontouring in clinical practice: A practical method using the Turing test," *Med. Phys.*, vol. 45, no. 11, pp. 5105–5115, 2018.

[69] A. M. TURING, "I.—COMPUTING MACHINERY AND INTELLIGENCE," *Mind*, vol. LIX, no. 236, pp. 433–460, Oct. 1950.

[70] S. Al-qazzaz *et al.*, "Image classification-based brain tumour tissue segmentation," *Multimed. Tools Appl.*, 2020.

[71] D. Clunie, "Standardizing AI Annotations - The DICOM Way," in *Conference on Machine Intelligence in Medical Imaging*, 2018.

[72] C. Herz *et al.*, "dcmqi : An Open Source Library for Standardized Communication of Quantitative Image Analysis Results Using DICOM," *Cancer Res.*, vol. 77, no. 21, pp. e87–e90, Nov. 2017.

[73] C. P. Langlotz, "Will Artificial Intelligence Replace Radiologists?," *Radiol. Artif. Intell.*, vol. 1, no. 3, p. e190058, May 2019.

[74] M. J. Gooding, "Assessment of thoracic auto-contouring using a modified Turing Test," 2017.

[75] T. J. Netherton, C. E. Cardenas, D. J. Rhee, L. E. Court, and B. M. Beadle, "The emergence of artificial intelligence within radiation oncology treatment planning," *Oncol.*, vol. 77030, pp. 124–134, 2020.

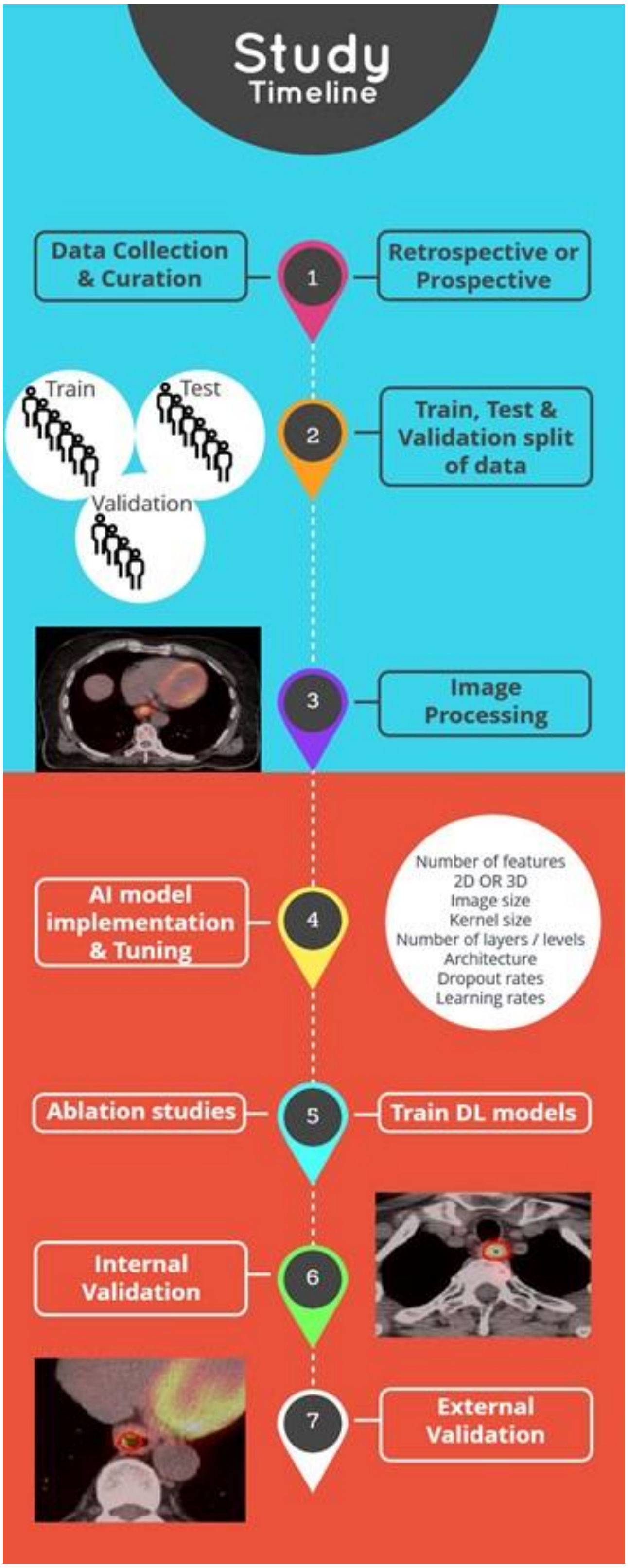

Figure 1: A typical pathway of the decision required when developing AI algorithms. Decisions required include how to collect data, split data, and process it. Then decisions on the type of AI model required, and its parameters.

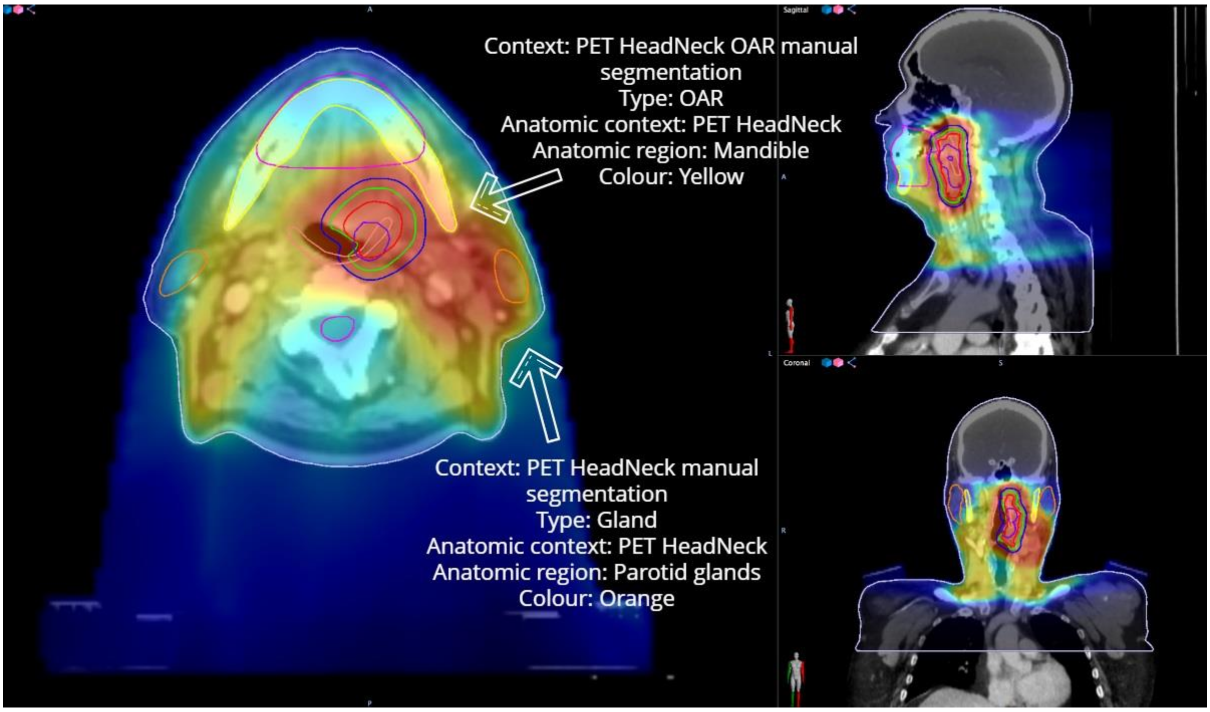

Figure 2: A planned H&N cancer patient, the metabolic tumour volume (purple), gross tumour volume (red), clinical target volume (green) and planned target volume (blue) have been manually segmented. Also, the mandible (yellow), oral cavity (pink), parotid glands (orange), spinal cord (pink) and the patient outline (white) have been segmented. DICOM-Seg allows for contextual information to be stored within the structure.