Automated Tumor Segmentation in Radiotherapy

Ricky R Savjani[1,2] MD/PhD, Michael Lauria[3] MS, Supratik Bose[2] PhD, Jie Deng[1] MD/PhD, Ye Yuan[4] MD/PhD, Vincent Andrearczyk[5] PhD

University of California, Los Angeles, Department of Radiation Oncology[1] and Graduate Program in Physics and Biology in Medicine[3], Los Angeles, California

Varian Medical Systems, A Siemens Healthineers Company, Applied Research[2], Palo Alto, CA

New York University, Department of Radiation Oncology[4], New York, New York

University of Applied Sciences and Arts Western Switzerland[5], Sierre, Valais, Switzerland

Abstract

Autosegmentation of gross tumor volumes (GTVs) holds promise to decrease clinical demand and to provide consistency across clinicians and institutions for radiation treatment planning. Additionally, autosegmentation can enable imaging analyses such as radiomics to construct and deploy large studies with thousands of patients. Here, we review modern results that utilize deep learning approaches to segment tumors in five major clinical sites: brain, head and neck, thorax, abdomen, and pelvis. We focus on approaches that inch closer to clinical adoption, highlighting winning entries in international competitions, unique network architectures, and novel ways of overcoming specific challenges. We also broadly discuss the future of GTV autosegmentation and the remaining barriers that must be overcome before widespread replacement or augmentation of manual contouring.

Introduction

A critical component of radiotherapy planning involves segmentation of both target volumes and organs at risk (OARs). This process utilizes a significant portion of physician and staff time away from patients to contour structures prior to dosimetric treatment planning. Accurate segmentation depends crucially on the underlying imaging to guide the segmentation, which gives rise to the potential to automate the entire process: autosegmentation.

Segmentation of OARs is a time-consuming process for radiotherapy planning, and autosegmentation holds promise to ease the clinical demand and bolster contour consistency. However, the sheer variability in patient anatomy, positioning, implants, catheters/stents, metal artifacts, and physiological state is enormous and continues to challenge autosegmentation. A comprehensive historical perspective on autosegmentation of OARs from atlas-based segmentation to deep-learning based approaches has been recently summarized in a comprehensive book: *Auto-Segmentation for Radiation Oncology: State of the Art*,[1] which focuses on the 2017 AAPM Thoracic Auto-segmentation Challenge dataset. OAR autosegmentation has been gaining significant traction with several institutions clinically integrating OAR autosegmentation and products being deployed by industry. Given the additional thorough reviews[1–4] of OAR autosegmentation and its relative maturity, here we instead focus on autosegmentation of gross tumor volumes (GTVs).

GTV segmentation boils fundamentally down to selecting which voxels contain tumors and which do not. However, there is significant variability in this task - physicians hold varying training experiences, adopt unique preferences, incorporate differing amounts of clinical information into the contour, and perform patient-individualized tradeoffs between tumor control and toxicity. These decisions are often baked into the contour and not always represented in the underlying imaging alone. In manual segmentation, inter-observer variability can be significantly impacting both clinical treatment and radiomic features and predictive power.[5,6] Additionally, information from multi-modal imaging is often needed to help define the tumor volumes (see the review in this series on multimodal image registration. Autosegmentation of GTVs can play an important role in not only decreasing the clinical demand but also in providing consistency and standardization across providers and departments.

We focus in this review on advances in GTV autosegmentation made in five key sites: brain, head and neck, thorax, abdomen, and pelvis (**Figure 1**). We will discuss innovations and models designed specifically for autosegmentation in these areas.

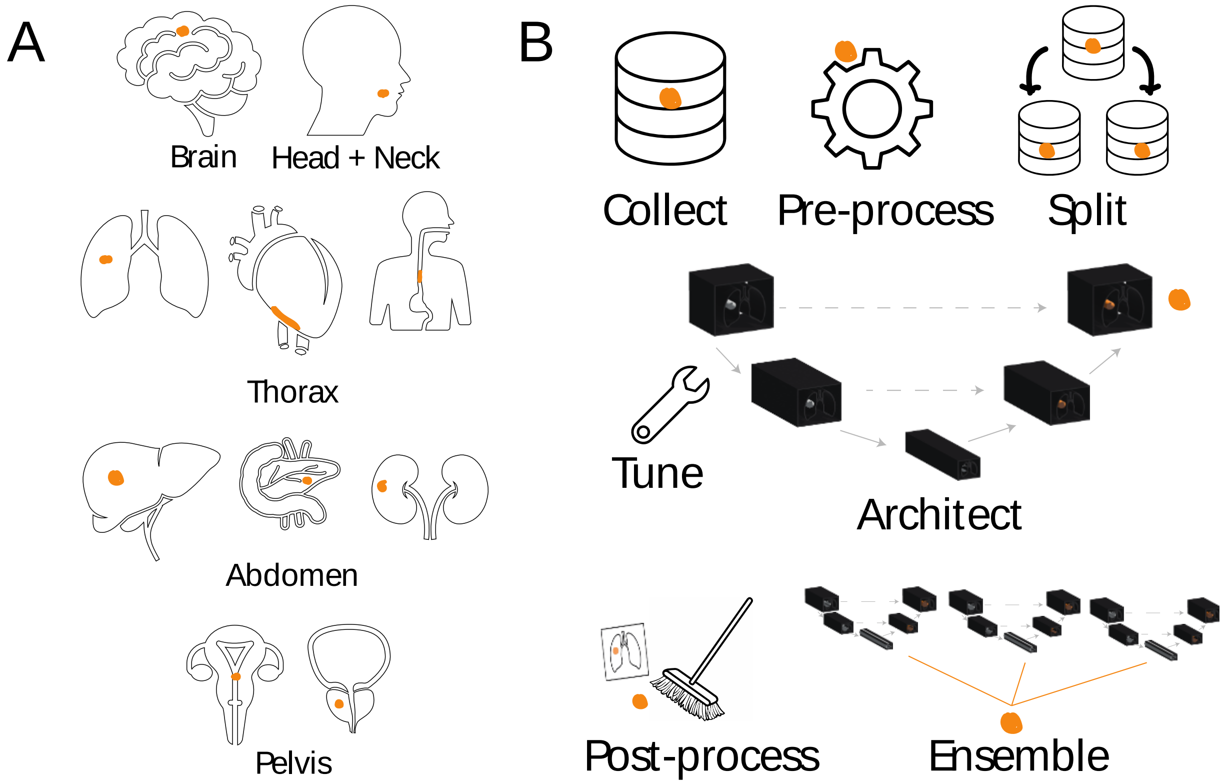

Figure 1. **A.** In this review, we highlight major advances in tumor autosegmentation for the five clinical sites: brain, head and neck, thorax (lung, heart, and esophagus), abdomen (liver, pancreas, kidney), and pelvis (prostate and cervix). **B.** Successful autosegmentation models rely on several steps including: data collection and curation, pre-processing and data ingestion, splitting datasets into train/validate/test sets, hyperparameter optimization and tuning, architecting networks, post-processing and visualization, and aggregating outputs from ensembles of networks.

Table 1. Overview of top-performing tumor segmentation models by site, highlighting novel architects, key innovations, outstanding challenges, datasets used, and best Dice scores.

	Top-performing Architects	Key Innovations	Challenges Remaining	Key Dataset(s)	Dice Score (%)
Brain					
Glioblastoma	Densely Connected CNNs	Sparsification training	Robustness	BraTS	0.89
Brain Metastases	3D U-Net	T1 w - T1 wo subtraction maps	Lesions < 6 mm	BraTS	0.75
Intracranial Multi-class	3D U-Net	Add 10% of institutional data to previously trained models to boost performance	Validation	BraTS	0.77
Head and Neck					
Oropharynx	Squeeze-and-excitation layers	Adaptive weighting of channel-wise features (e.g., PET and CT)	Ground truth, no bounding box, MRI datasets	HECKTOR	0.76
Elective Nodes	U-Net	Computer vision pre-processing	External validation, comparison to vender models	MDA	0.90
Thorax					
Heart	Multi-atlas-based approaches	4DCT and multi-atlas-based approaches	Substructures for radioablation planning, coronary vessels, motion management	Australian Breast Cancer patients (n = 20)	0.92

Lung	CNNs, ResNets	Adaptive CNNs	SABR-specific models and consistency of contours	TCIA, MSKCC, LIDC	0.82
Esophagus	Two-stream deep fusion framework, multi-branch decoders, Attention-based U-Nets	Progressive semantically - nested network: DeepTarget	Low tissue contrast	Chang Gung Memorial Hospital (n = 148)	0.79

Abdomen

Kidney	nnU-net	Automated preprocessing and model architecture decisions	larger, more varied datasets	KiTS19	0.85
Pancreas	Square-window CNN, U-Net	CE endoscopic US images have helped	motion management, image quality	Medical College of Wisconsin (n = 40)	0.73
Liver	Adversarial networks, Spatial feature fusion CNNs,	Arterial phase imaging	diversity of datasets	LiTS and 3DIRCADb	0.84

Pelvis

Prostate	Physician style-aware network, multiple decoders	cater to preferences/styles	post-prostatectomy bed	UC Irvine (n = 242) mpMRI Prostate	0.94
Cervix	U-Nets, fine-tuning	compare to resident-level performance	Generalizability	First Affiliated Hospital of Anhui Medical University in China (n = 125)	0.86

Deep-learning State-of-the-Art in Medical Image Segmentation

Prior to diving into the site-specific models, there have been several intuitions developed from deep-learning worth discussing upfront. Radiological imaging as input to deep learning models differs markedly from more conventional photographs and images used to train large neural networks. Importantly, medical imaging data are often multimodal (combinations of X-ray, CT, contrast enhancement, MR sequences, PET, and Ultrasound), non-isotropic (voxels can have different slice thicknesses), three-dimensional (with various reconstructions), and fixed in viewpoint (patients tend to be scanned in certain standard positions like supine). Researchers from single institutions may have access to only their own unique patient dataset, but medical imaging data generally is subject to domain shifts[7] in which different hospitals and even different scanners or protocols on the same scanner can introduce significant variation in the resultant imaging. Carefully selecting and validating data at a multi-institutional level is imperative to generate robust models with clinical relevance.

Prior to deep learning, computer vision and machine learning were utilized to attempt autosegmentation. These approaches often required knowledge of image properties to guide manual selection of parameters such as contrast-based thresholding, definition of edge detectors, or cluster determination. They tended to work well for particular datasets or patients but often did not generalize well to different centers and held an upward limit to their utility. Deep learning offered a different approach in which parameters could be derived from training with the data and optimizing weights of neural networks. The evolution of deep learning approaches for medical imaging segmentation has been elegantly reviewed recently.[8,9]

The most commonly used architectures utilize convolutional neural networks (CNNs), with most adopting U-Nets[10] or V-Nets[11]. Briefly, these architectures have a downsampling path in which images are compressed into higher level semantic features with increasing depth. The upsampling path then brings the images back into input resolution, and skip-connections allow the network to bring information across the downsampling path directly to the upsampling path. To date, most top-performing tumor segmentation architectures use a flavor of a U-Net at their core.[12–15] Many features and components of the network can be further customized to optimize performance. Several top performing models also employ model ensembling in which multiple models trained separately with various splits of the data or distinct model configurations are combined to vote for the most likely GTV.[16,17] Further, instead of hand-crafting features, frameworks like neural architecture search (NAS) explore a broad gamut of architectures and find the optimal configurations.[18–20] NAS and even hyperparameter optimization[21] tend to require hundreds of GPU hours and often leverage multi-GPU and/or multi-node hardware to train several possible networks simultaneously to find optimal performing networks.

However, the network architecture itself is not the only decision to be made. Beyond the network, decisions on pre-processing, training scheme (e.g., data splits, hyperparameters, loss functions, optimizers, data augmentation), post-processing and ensembling are key attributes that need to be carefully selected alongside the network architecture. These decisions typically are hand-chosen and can vary significantly across datasets and tasks. nnU-Net recently innovated this process by creating data fingerprints which aim to automate these preprocessing decisions directly from the data or from fixed parameters that have been shown to work robustly.[22] nnU-Net at its core uses a U-Net, but it adapts all other decisions to create one framework to segment any medical imaging task. Out-of-the-box, nnU-Net has been shown to score highly in several competitions without any fine-tuning. Several entries in modern challenges today are now using nnU-Net as a baseline and adding features to it or replacing components.

During training and on validation, models need to be evaluated with a loss function. There has been a deep investigation using nnU-Net and varying loss functions on a variety of segmentation tasks.[23] Interestingly, no single loss function was able to work robustly across datasets. Instead, a combination of Dice and other loss functions tended to perform best. However, as we will see, most studies train and report a single loss function (e.g., Dice, Hausdoff distance, etc.). Additionally, the evaluation metric depends critically upon the goal, and physician review may be necessary for widespread clinical adoption of autosegmentation models.[24]

Although this review is aimed for radiation oncology departments, we pull together here information and studies across several different disciplines including in radiology, international tumor segmentation challenges, AI-based conferences, and the literature more broadly to provide a well-rounded perspective on the state of autosegmentation for each of these 5 sites.

Autosegmentation of GTVs (and in AI models generally) is challenged by the limited number of available datasets, bias in the training data, differences in image acquisition protocols, and a trade-off between accuracy and complexity in deep neural networks. There is a rich history of segmentation approaches for each site, which we cannot fully capture here. Rather, we focus mostly on works published over the last two years (2020 – 2021).

Site-specific Advances in Autosegmentation

Brain

Glioblastoma: Glioblastoma multiforme (GBM) represents one of the most challenging tumors to contour. Clinically, GBMs are typically contoured post-operatively, and thus pre- and post-imaging as well as multimodal MRI are needed to delineate the tumor bed. Recent efforts have used densely connected CNNs to segment the resection cavity of GBMs.[25] However, failures compared to manual contours persisted, especially in areas with signal inhomogeneities like the ventricles and subarachnoid spaces, where the model failed to differentiate resection cavity from normal anatomy. Gross GBM tumors can also be important to contour to ensure coverage of initial pre-surgical lesions. Significant heterogeneity in multi-modal imaging exists, including missing acquisitions of particular sequences. Sparsification training can simulate missing MR sequences during training and has been shown to improve autosegmentation of gross GBMs, allowing for implementation on more heterogeneous data acquisitions.[26] The Brain Tumor Segmentation (BraTS) challenge has been proposed yearly since 2012 for multimodal MR GBM segmentation. Most recently in 2020 results, nnU-net was used to achieve the top performing scores with a Dice of 0.8895 for the whole tumor, emphasizing that preprocessing decisions can play an instrumental role in autosegmentation.[27]

Brain Metastases: Autosegmentation of brain metastases poses some unique challenges in that intracranial metastases can often have multiple lesions on initial presentation, as well as have a high propensity to develop new lesions on follow-up imaging. Several implementations of various flavors of 3D U-Nets for the identification of brain metastases have recently been published.[28–31] Most of these implementations focus on utilizing the T1-weighted MR imaging with contrast, which best isolates the tumor and is most heavily used in manual contouring. Some studies also compute subtraction maps between T1-weighted contrast volumes and T1-weighted non-contrast volumes, and use all three as inputs to the model.[28] Importantly, Zhou and colleagues utilized DL-based single shot detectors to output bounding boxes and confidence measures of individual lesions.[32] Detection is generally a different class of diagnostic problem than segmentation, but for brain metastases detection of small lesions can be instrumental. They noted excellent performance on identifying lesions greater than 6 mm, detecting all lesions with few false positives; however, for lesions less than 6 mm, results were markedly worse. This is an important area of research, as stereotactic radiosurgery is being increasingly used to treat small lesions as soon as they radiographically appear or grow. False negatives tend to be lesions less than 3 mm, subtle lesions, or lesions near the dura/vessels, whereas false positives are more extra-axial, within bone, or developmental anomalies.[28]

Multi-class: One recent framework has also been able to classify tissues into different intracranial tumor types (low and high-grade gliomas, brain metastases, meningiomas, pituitary adenomas, and acoustic neuromas).[33] Although this may not be essential for radiotherapy planning, one framework or model that can robustly classify and identify multiple different lesion types holds great clinical value. Further, a recurrent theme in deep learning is overfitting onto the training set and the need for a variety of multi-institutional data. Recent work has shown that a 3D U-Net trained to identify a variety of neurologic abnormalities (including various tumors) on T2 FLAIR imaging does not generalize well to an independent institution not used in training. However, if a modest amount of training data (10%) is included that closely matches the distribution and characteristics of the test set, the AI model can perform significantly better on the test set.[34] This is an interesting concept that might allow institutions to take previously trained models from public repositories and retrain them including a small amount of data from their own institution that could be more readily available.

Head and Neck

Autosegmentation of the head and neck is of particular interest due to the necessary clinical tradeoff between tumor control and radiation-induced toxicity. It is clear that multimodal imaging is necessary for both the manual delineation of head and neck GTVs, as well as in autosegmentation. PET imaging reflects the metabolic tumor response, indicating the active tumor region and is robust to metal artifact, whereas CT focuses on morphological tissue properties. A recent quantitative review of segmentation approaches for GTVs in the head and neck for both primary tumor and nodal GTVs demonstrated the superiority of using multi-modal (PET and CT) imaging over CT alone, as well demonstrating superiority of a 2D CNN compared to classical thresholding and machine learning approaches.[35] CNN models that used multi-channel PET and CT achieved Dice scores of 0.74, compared to 0.66 (CT) and 0.68 (PET) alone.

Oropharynx: Oropharyngeal cancers (OPCs) are globally the most common primary head and neck cancer. The Medical Imaging Computing and Computer Assisted Intervention Society (MICCAI) has hosted and run the Head and NeCK TumOr (HECKTOR) segmentation challenge in which fused PET and CT imaging were provided as a challenge for autosegmentation in 2020[17] and has extended this competition in 2021[36]. The winning submission achieved a Dice score of 0.7591 on a hold-out test set using a squeeze-and-excitation (SE) normalization, which adaptively weights channel-wise features (here, the PET and CT imaging).[12] A similar approach also obtained the best score on the enriched test set in 2021 (Dice of 0.7785), confirming the results. Most of the top scoring submissions used multi-modal 3D U-Nets or ResNets of various flavors, with a few top submissions employing generative adversarial networks (GANs). GTV autosegmentation would allow for prediction of clinical outcomes on large populations of data, and validation studies directly comparing outcomes predictions from manual contours to autosegmentations are showing increasingly comparable results.[37] Further, a multi-task architecture that jointly trains both GTV autosegmentation and clinical outcomes (radiomics) data with a common encoder in an end-to-end fashion has recently shown to have greater predictive power, as well as an ability to predict clinical outcomes without requiring a segmentation as input at all.[38]

However, physiologic PET avidity and image registration from PET to CT Simulation scans do pose significant challenges. To overcome these challenges, multi-parametric MRI is now being routinely obtained prior to treatment for diagnosis and planning. Additionally, MR-linacs are increasingly being used for OPC and head and neck radiotherapy. Multi-parametric MR using anatomical (T1-weighted, T2-weighted) combined with functional (apparent diffusion coefficient (ADC), volume transfer constant, and extracellular volume fraction) imaging has been recently used to train a multi-channel U-Net for autosegmentation of OPCs.[39] These early results show promise and were retrospectively indistinguishable to physicians compared to manual contours. Similar approaches and results have been seen at other institutions, with anatomical MR alone.[40] Further work is under investigation to evaluate dosimetric and potential clinical outcomes and toxicity impacts of these autosegmentations. Still, MR presents challenges in segmentation due to metal artifacts and poses clinical challenges due to availability and contraindications in certain patient populations.

Gross and Elective Nodal Irradiation: Gross nodal metastases that are PET avid or meet size or morphology specifications have also been successfully contoured with autosegmentation U-Net and V-Net architectures, without explicit distinction from the primary tumors.[41,42] Ongoing efforts are also underway in autosegmentation that distinguish nodal GTVs from primary GTVs.[43] CTV neck nodal contouring, while not gross tumor, remains the most time-consuming aspect of head and neck contouring. Recent work from MD Anderson has automated the contouring of CTV neck nodal levels using computer-vision volume of interest identification and U-Nets.[44] This work is particularly appealing clinically due to the reduction of time spent and variability rendered in manual contouring.

Head and neck tumor sites outside of the oropharynx are less well studied due to the relatively lower prevalence. Similar approaches using 3D Unets have been tried with success for salivary gland tumors[45], nasopharyngeal carcinomas[46,47], and thyroid nodules on diagnostic scanning [48].

Thorax

Many advances have been made to enable GTV autosegmentation of thoracic anatomy through machine and deep learning techniques. This review will primarily focus on the current state of autosegmentation of GTVs in the heart, lungs, and esophagus.

Heart: While radiation therapy is not the standard care for cardiac tumors[49], cardiac radioablation is a treatment that would benefit from the fast, accurate target volume delineation that automatic techniques have to offer. Autosegmentation of the entire heart has been performed with a high degree of accuracy thanks to atlas-based approaches. Finnegan et al. recently achieved a mean Dice of 0.923 ± 0.01 using a multi-atlas-based approach with 4DCTs[50,51]. However, for radioablation, the substructures are important to contour but success has been varied. Using an atlas-based approach, Ferrugia et al. determined that while larger substructures like the great vessels and heart chambers could be successfully autosegmented, the coronary arteries and heart valves had too much segmentation variability to be applied clinically[52]. This conclusion matched that of similar previous studies[53,54]. Results could be improved with motion management techniques to raise the quality of those smaller substructures, as well as through additional datasets.

Lung: A thorough review of the advancements in deep learning-based autosegmentation of GTVs in the lungs was published in July 2021 by Liu et al[15]. Much of the research cited in the review involves novel techniques inspired by the CNN architecture. For example, Wang et al. designed the patient specific A-net for contouring non-small cell lung cancer (NSCLC) tumors seen in MRI.[55] The network was trained on previous weekly MRI images and tested on current weekly images, yielding an average Dice of 0.82 ± 0.10 when comparing the contours to those contoured manually by radiation oncologists. Zhang et al. modified a ResNet to segment the GTV of NSCLC patients on CT images.[56] The modification fused shallow surface features with the deep semantic features to generate dense pixel outputs, and this led to an average Dice of 0.73. The review also includes the full resolution residual neural network (FRRN) proposed by Pohlen et al.[57], which passes full resolution of features to each layer, and the later modification to multiple resolutions in the multiple resolution residually connected network (MRRN) by Jiang et al.[58] These developments improved the ability to recover the input image resolution and increased robustness of results. Finally, the efforts to develop multi-modality techniques are recognized, especially the work of Zhao et al. in combining sub-segmentation branches that handled CT and PET images with a V-Net and later fused the modalities, providing an average Dice of 0.85 ± 0.08.[59] Another review of target volume contouring in radiation therapy by Mercieca et al. suggested that a large database of contours with a common protocol, peer review, and acceptable local control and toxicities could alleviate many of the issues with learning-based autosegmentation[60], which have also been recently highlighted specifically for lung GTVs[61]. Few studies have also specifically focused on lung stereotactic ablative radiotherapy (SABR) GTV contours to train models, as the SABR contours overall may differ.[62] However, the current studies are encouraging for the future of deep learning-based lung tumor autosegmentation for routine clinical use.

Esophagus: Esophageal tumors can be trickier to segment than NSCLC tumors due to the lack of contrast from the surrounding tissue, and thus could be a great beneficiary of deep learning techniques. Recent studies have tried combatting the low contrast with the use of PET/CT. Jin et al. provides a thorough analysis of autosegmentation of esophageal tumors using a two-stream chained deep fusion framework for CT and PET and a progressive semantically-nested network, an approach they call DeepTarget, including comparison to a wide variety of state-of-the-art approaches from other groups.[63] With a mean Dice of 0.790 ± 0.095, their technique outperformed DenseUNet, progressive holistically nested neural networks, and several other cited fusion approaches. In an effort to simplify the workflow, Yousefi et al. developed a Dilated Dense Attention U-Net to automatically segment esophageal tumors in CT only.[64] They successfully obtained comparable results with a mean Dice of 0.79 ± 0.20. The group highlighted an enriched dataset containing a wider variety of tumors, air pockets, foreign bodies, etc. to improve results in the future. Recent work has also used two distinct decoders (multi-branch) to segment separately distal and proximal esophageal lymph node GTVs based on OAR distance-based gating.[65]

Abdomen

Kidney: Autosegmentation of tumors in the kidneys was put to the test during the 2019 KiTS19 Challenge[66] in which teams were given common training and testing data to try to achieve the best Dice in kidney and GTV segmentations. It was anticipated that garnishing the nnU-net[67] would yield the highest score, but the winning team used the original architecture and focused on pre-processing to achieve a tumor segmentation Dice of 0.851.[68] As has been a consistent theme in autosegmentation, the future direction of this challenge includes a larger and more varied training dataset to reduce bias.

Pancreas: Autosegmentation of pancreatic tumors has been seemingly more difficult. In multi-parametric MRI, a square-window CNN-based approach yielded average Dice of 0.73 ± 0.09, though very notably this was comparable to a Dice between two separate observer contours.[69] Another interesting study utilized contrast-enhanced endoscopic ultrasound images of pancreatic tumors along with a U-Net to accomplish the autosegmentation task.[70] Instead of Dice, Intersection over Union (IoU) was used to evaluate the results, which included a mean IoU of 0.77 and minimum and maximum values of 0.39 and 0.91, respectively. This indicates that the use of deep learning offers encouraging results, but further developments are needed to obtain consistent accuracy suitable for clinical implementation. Improvements in motion management, image quality, and network architecture have been cited as key steps to enabling more accurate results.

Liver: There have been a relatively large number of studies pertaining to autosegmentation of tumors of the liver with increasing levels of success. Most of the published work utilized two publicly available CT datasets: 2017 LiTS[71] and 3DIRCADb[72]. One example is the SegNet-based study performed by Almotairi et al.[73] SegNet is an encoder-decoder network with a pixel-wise classification layer. Using the 3DIRCADb dataset, tumor segmentations were achieved with superior accuracy to many previously applied techniques including random forest[74], cascaded fully convolutional neural networks[75], CNN[76], hierarchical convolution[77], and others. For three test cases, the IoUs were all above 0.90. An example of a study using the LiTS dataset is Liu et al, in which a Spatial Feature Fusion Convolutional Network was presented to segment tumors[78]. This approach included output extraction at every convolutional block and skip-connections in the down-sampling phase to efficiently transfer spatial information to later layers. Feature fusion blocks were used to merge spatial features, and fully connected 3D conditional random fields were applied to refine segmentations. With this technique, the mean Dice per case achieved for liver tumors was 0.59 and the mean Dice when considering all cases as an entire volume, or Dice global, was 0.75. This study also included many previously developed techniques for comparison and showed superior results. Two impressive studies were performed with data outside of the two typically used datasets. The first, by Xu et al.[79], utilized arterial phase images to provide additional information to the segmentations performed on portal venous phase images with a network architecture inspired by the VoxResNet[80]. With this approach, a DPC and DG of 0.78 and 0.87 were achieved, respectively. The other study by Zhao et al. used a united adversarial learning framework with several novel techniques to segment tumors in multi-modality non-contrast MRI[81]. These features included an edge dissimilarity feature pyramid module, a fusion and selection channel, coordinate sharing with padding, and a multi-phase radiomics guided discriminator to use radiomics features to enhance the autosegmentation results. This study achieved mean Dice and IoU of 0.84 ± 0.02 and 0.81 ± 0.03, respectively. Through the recent success of these studies, it is evident that adding more image information has been helpful in improving results. Additionally, improving the diversity of datasets and tumor types, along with developing the network architectures and adding useful modifications, are promising ways to further improve liver tumor autosegmentation in the future.

Pelvis

Prostate: Radiotherapy for intact prostate typically involves treating the entire prostate, thus autosegmentation approaches for the entire gland would serve well clinically for radiotherapy. Results of a CNN model applied to a single institution showed that 65% of contours (both the CTV and OARs) required only minor edits, saving an average of 12 minutes per case for physicians.[82] However, 35% of contours required major edits, and no autosegmentations were created that did not require any editing even though CTV Dice scores were high at 0.89. Further work has been done recently to segment out the transitional zone, peripheral zone, and the prostate cancer lesion itself, which may be useful as considerations for boosting gross prostate disease within the gland are evolving.[83,84]

After prostatectomy, resection cavities are more complex and give rise to more physician preferences. Recent work from UT Southwestern attempted to build a physician-style aware (PSA) network that could learn different preference styles first with a CNN and then use an encoder paired with multiple decoders that represented particular physician styles.[85,86] The study found no major associated clinical outcomes in biochemical recurrence or toxicity associated with physician styles, and the autosegmentation of post-op beds can be tailored to individual physician styles. These style aware approaches may increase adoption of autosegmenation into clinical practice.

New frameworks like Ethos (Varian, A Siemens Healthineers Company, Palo Alto, CA) can potentially allow daily adaptive treatment enabled by automated contouring on CBCT. Evaluations of this approach have shown that on Ethos automated CBCT can generate clinically acceptable contours without any editing and with reductions in OAR dose in 24 of 25 patients.[87] However, one patient did require significant edits in the auto CBCT contour, highlighting that these contours still require physician review and potential editing. Further, such systems have only been tested for intact prostate CTV with seminal vesicles - more work is needed for nodal involvement and post-prostatectomy treatments.

Cervix: Autosegmentation for cervical cancer has also been gaining attention. An interesting comparison was made against a U-Net model vs. a single resident physician learning to contour CTVs for cervical cancer compared to an attending physician for 125 patients.[88] The U-Net autosegmentation model was able to achieve comparable levels of segmentation performance as measured by Dice and Haursdorf distance compared to the resident physician. Recent work showing fine-tuning a model previously trained at another institution also can improve generalizability.[89] Further, adversarial networks with multi-institutional data and scored in three stages (objective performance, subjective physician assessment, and Turing test) showed promising results priming the stage for clinical adoption.[90] Additionally, U-Net models have also been trained to segment and reconstruct the applicators for brachytherapy with promising Dice and HD scores, as well as low tip and shaft location errors.[91]

Discussion

The state of GTV of autosegmentation is constantly evolving for multiple tumor types, marching towards clinical utilization. Hosted challenges and competitions have pushed forward methodologies and architectures to improve accuracies across multiple tumor sites. However, outside of such constrained challenges, it remains difficult to compare performances across different sites and studies. Evaluation metrics like Dice depend upon tumor volumes, datasets contain various consistency of ground truth segmentations and the number of patients, and imaging quality vary significantly from institutions and studies performed. Ideally, we might plot a metric across all tumor sites in the body to understand where we excel and which GTV tumors need improvement, but such a depiction would bury the intricacies and challenges associated with each GTV type.

Despite the great advances discussed above for each of these sites, widespread adoption of clinical GTV autosegmentation remains limited. Given that GTVs will be targeted with the highest dose of radiation, physicians certainly carry the responsibility that the appropriate volume is contoured. For clinical contouring, the physician will remain instrumental to the oversight and editing, even for autosegmentation models. Recent work has shown that even when ML models perform objectively well and even would be selected retrospectively, there can be significant differences when evaluating prospectively (i.e., when actually deciding to use the ML/DL model to treat a patient).[92] Physicians likely have an individualized preference and comfort level, and approaches like style-awareness[85] may help achieve more widespread clinical adoption. Nonetheless, autosegmentation aims to improve consistency and can enable large scale analyses like radiomics that can remove the need for manual physician segmentation and extract features within the regions of interest. Recent work is revealing that GTVs contoured with autosegmentation can have comparable predictive power to manual annotations.[37]

Several other clinical challenges still remain. Industry and research institutions may wish to commercialize their algorithms, which requires regulatory oversight and FDA approvals, a lengthy and costly process.[93] Further, there is a common theme of pitfalls in applying AI to medical imaging, as has been highlighted repeatedly in lessons learned from DL attempts in Covid-19 classification.[94] GTV autosegmentation must learn from those mistakes and not repeat them to avoid negative attention on the approach. Additionally, there are growing concerns with data privacy and HIPAA compliance - while data sharing is theoretically ideal, many institutions have strict regulatory policies on data governance and sharing. Approaches like federated learning[95] can allow individual institutions to retain their data but share only model parameters/weights to centralized servers to train with data at many institutions securely. Swarm learning[95] goes one step further and removes the centralized server by invoking edge computing and block-chain coordination. These approaches may help institutions retain their data but participate in training large segmentation models across tens of thousands of patients. Lastly, there is no doubt that autosegmentation output will require manual physician editing for when it fails for individual cases or when a physician may desire to override the output. Several approaches exist in detecting out-of-distribution cases and poor segmentations, including recently using variational autoencoders[96]. Further, techniques that allow the physician to just click a few areas rather than recontouring the whole structure such as DeepGrow[97] and Gated Graph Propagator[98] may help enhance clinical adoption of entire autosegmentation frameworks. For radiotherapy, there is also growing interest in methodologies combining registration and segmentation into a single framework, especially for adaptive radiotherapy treatment deliveries.

Beyond adoption, there is increasing attention on the interpretability of AI models. Classification tasks undergo sanity checks to ensure relevant features are being used, for example with saliency maps such as in Grad-CAM[99]. Saliency maps are not particularly useful for autosegmentation (and generally shouldn't be used as a means for medical imaging segmentation[100]); however, there are emerging approaches that attempt to increase explainability for autosegmentation. Global features can be captured with concept vectors and used to probe how much a model may be associated or correlated with each concept, which has been applied for histopathological identification of breast tumors[101] and radiomics analyses[102]. Additionally, deep CNNs have been studied with probed with effective receptive fields[103], showing that local information tends to still be preserved in deep layers neural nets, and the overall shape of the attention of network layers is Gaussian, yielding a foveal attentional representation akin to the human retina. Another important consideration for autosegmentation models is uncertainty - where are models less confident about their predictions on which voxels are indeed GTV? Predictive uncertainty can be dissected into constituent parts: aleatoric uncertainty (arising from noisy data) and epistemic uncertainty (confidence in model parameter weights and whether the right model was selected for the task).[104] Understanding and assessing where these uncertainties arise from and communicating them to clinicians can increase trust in AI-based autosegmentation models.[104,105]

Conclusions

Here, we highlight significant progress made on autosegmentation for five key tumor sites for radiation therapy: brain, head and neck, thorax, abdomen, and pelvis. Many of these studies have been objectively evaluated, tested retrospectively in clinical settings, and put to the Turing test. However, most of these implementations are not a part of routine treatment planning yet. While the field is advancing network design and architectures, we must, in parallel, evaluate these models prospectively in the clinic. With physician involvement, autosegmentation can be added as a new brush in the contouring toolbox, and physicians can start fluidly working with it. Clinical feedback will also likely inform how to iterate and improve autosegmentation models, rather than just objective metrics like Dice scores. We try to capture here the state-of-the-art in GTV autosegmentation and highlight the path ahead for more widespread clinical adoption and integration.

References

[1] G. Sharp, J. Yang, M.J. Gooding, Auto-Segmentation for Radiation Oncology: State of the Art, 1st ed., 2021. https://doi.org/10.1201/9780429323782.
[2] T. Vrtovec, D. Močnik, P. Strojan, F. Pernuš, B. Ibragimov, Auto-segmentation of organs at risk for head and neck radiotherapy planning: From atlas-based to deep learning methods, Med Phys. 47 (2020) e929–e950. https://doi.org/10.1002/mp.14320.
[3] H. Guo, J. Wang, X. Xia, Y. Zhong, J. Peng, Z. Zhang, W. Hu, The dosimetric impact of deep learning-based auto-segmentation of organs at risk on nasopharyngeal and rectal cancer, Radiat Oncol. 16 (2021) 113. https://doi.org/10.1186/s13014-021-01837-y.
[4] J. Wong, V. Huang, D. Wells, J. Giambattista, J. Giambattista, C. Kolbeck, K. Otto, E.P. Saibishkumar, A. Alexander, Implementation of deep learning-based auto-segmentation for radiotherapy planning structures: a workflow study at two cancer centers, Radiat Oncol. 16 (2021) 101. https://doi.org/10.1186/s13014-021-01831-4.
[5] C. Haarburger, G. Müller-Franzes, L. Weninger, C. Kuhl, D. Truhn, D. Merhof, Radiomics feature reproducibility under inter-rater variability in segmentations of CT images, Sci Rep-Uk. 10 (2020) 12688. https://doi.org/10.1038/s41598-020-69534-6.
[6] R. Liu, H. Elhalawani, A.S.R. Mohamed, B. Elgohari, L. Court, H. Zhu, C.D. Fuller, Stability analysis of CT radiomic features with respect to segmentation variation in oropharyngeal cancer, Clin Transl Radiat Oncol. 21 (2019) 11–18. https://doi.org/10.1016/j.ctro.2019.11.005.
[7] H. Guan, M. Liu, Domain Adaptation for Medical Image Analysis: A Survey, Ieee T Bio-Med Eng. PP (2021) 1–1. https://doi.org/10.1109/tbme.2021.3117407.
[8] I.R.I. Haque, J. Neubert, Deep learning approaches to biomedical image segmentation, Informatics Medicine Unlocked. 18 (2020) 100297. https://doi.org/10.1016/j.imu.2020.100297.
[9] X. Liu, L. Song, S. Liu, Y. Zhang, A Review of Deep-Learning-Based Medical Image Segmentation Methods, Sustainability-Basel. 13 (2021) 1224. https://doi.org/10.3390/su13031224.
[10] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, ArXiv:1505.04597 [Cs]. (2015). http://arxiv.org/abs/1505.04597.
[11] F. Milletari, N. Navab, S.-A. Ahmadi, V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation, Arxiv. (2016).
[12] A. Iantsen, D. Visvikis, M. Hatt, Head and Neck Tumor Segmentation, First Challenge, HECKTOR 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Proceedings, Lect Notes Comput Sc. (2021) 37–43. https://doi.org/10.1007/978-3-030-67194-5_4.
[13] E. Michael, H. Ma, H. Li, F. Kulwa, J. Li, Breast Cancer Segmentation Methods: Current Status and Future Potentials, Biomed Res Int. 2021 (2021) 1–29. https://doi.org/10.1155/2021/9962109.
[14] G. Samarasinghe, M. Jameson, S. Vinod, M. Field, J. Dowling, A. Sowmya, L. Holloway, Deep learning for segmentation in radiation therapy planning: a review, J Med Imag Radiat On. 65 (2021) 578–595. https://doi.org/10.1111/1754-9485.13286.
[15] X. Liu, K.-W. Li, R. Yang, L.-S. Geng, Review of Deep Learning Based Automatic Segmentation for Lung Cancer Radiotherapy, Frontiers Oncol. 11 (2021) 717039. https://doi.org/10.3389/fonc.2021.717039.
[16] A. Wadhwa, A. Bhardwaj, V.S. Verma, A review on brain tumor segmentation of MRI images, Magn Reson Imaging. 61 (2019) 247–259. https://doi.org/10.1016/j.mri.2019.05.043.

[17] V. Andrearczyk, V. Oreiller, M. Jreige, M. Vallières, J. Castelli, H. Elhalawani, S. Boughdad, J.O. Prior, A. Depeursinge, Head and Neck Tumor Segmentation, First Challenge, HECKTOR 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Proceedings, Lect Notes Comput Sc. (2021) 1–21. https://doi.org/10.1007/978-3-030-67194-5_1.
[18] Z. Zhu, C. Liu, D. Yang, A. Yuille, D. Xu, V-NAS: Neural Architecture Search for Volumetric Medical Image Segmentation, 2019 Int Conf 3d Vis 3dv. 00 (2019) 240–248. https://doi.org/10.1109/3dv.2019.00035.
[19] F. Wang, Neural Architecture Search for Gliomas Segmentation on Multimodal Magnetic Resonance Imaging, Arxiv. (2020).
[20] Q. Yu, D. Yang, H. Roth, Y. Bai, Y. Zhang, A.L. Yuille, D. Xu, C2FNAS: Coarse-to-Fine Neural Architecture Search for 3D Medical Image Segmentation, Arxiv. (2019).
[21] J.L. Berral, O. Aranda, J.L. Dominguez, J. Torres, Distributing Deep Learning Hyperparameter Tuning for 3D Medical Image Segmentation, Arxiv. (2021).
[22] F. Isensee, P.F. Jaeger, S.A.A. Kohl, J. Petersen, K.H. Maier-Hein, nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation, Nat Methods. (2020) 1--9. https://doi.org/10.1038/s41592-020-01008-z.
[23] J. Ma, J. Chen, M. Ng, R. Huang, Y. Li, C. Li, X. Yang, A.L. Martel, Loss Odyssey in Medical Image Segmentation, Med Image Anal. (2021) 102035. https://doi.org/10.1016/j.media.2021.102035.
[24] M.V. Sherer, D. Lin, S. Elguindi, S. Duke, L.-T. Tan, J. Cacicedo, M. Dahele, E.F. Gillespie, Metrics to evaluate the performance of auto-segmentation for radiation treatment planning: a critical review, Radiother Oncol. 160 (2021) 185–191. https://doi.org/10.1016/j.radonc.2021.05.003.
[25] E. Ermiş, A. Jungo, R. Poel, M. Blatti-Moreno, R. Meier, U. Knecht, D.M. Aebersold, M.K. Fix, P. Manser, M. Reyes, E. Herrmann, Fully automated brain resection cavity delineation for radiation target volume definition in glioblastoma patients using deep learning, Radiat Oncol. 15 (2020) 100. https://doi.org/10.1186/s13014-020-01553-z.
[26] R.S. Eijgelaar, M. Visser, D.M.J. Müller, F. Barkhof, H. Vrenken, M. van Herk, L. Bello, M.C. Nibali, M. Rossi, T. Sciortino, M.S. Berger, S. Hervey-Jumper, B. Kiesel, G. Widhalm, J. Furtner, P.A.J.T. Robe, E. Mandonnet, P.C.D.W. Hamer, J.C. de Munck, M.G. Witte, Robust Deep Learning–based Segmentation of Glioblastoma on Routine Clinical MRI Scans Using Sparsified Training, Radiology Artif Intell. 2 (2020) e190103. https://doi.org/10.1148/ryai.2020190103.
[27] F. Isensee, P.F. Jaeger, P.M. Full, P. Vollmuth, K.H. Maier-Hein, nnU-Net for Brain Tumor Segmentation, Arxiv. (2020).
[28] J.D. Rudie, D.A. Weiss, J.B. Colby, A.M. Rauschecker, B. Laguna, S. Braunstein, L.P. Sugrue, C.P. Hess, J.E. Villanueva-Meyer, 3D U-Net Convolutional Neural Network for Detection and Segmentation of Intracranial Metastases, Radiology Artif Intell. (2021) e200204. https://doi.org/10.1148/ryai.2021200204.
[29] E. Grøvik, D. Yi, M. Iv, E. Tong, D. Rubin, G. Zaharchuk, Deep learning enables automatic detection and segmentation of brain metastases on multisequence MRI, J Magn Reson Imaging. 51 (2020) 175–182. https://doi.org/10.1002/jmri.26766.
[30] E. Dikici, J.L. Ryu, M. Demirer, M. Bigelow, R.D. White, W. Slone, B.S. Erdal, L.M. Prevedello, Automated Brain Metastases Detection Framework for T1-Weighted Contrast-Enhanced 3D MRI, Ieee J Biomed Health. 24 (2019) 2883–2893.

https://doi.org/10.1109/jbhi.2020.2982103.
[31] Z. Zhou, J.W. Sanders, J.M. Johnson, M. Gule-Monroe, M. Chen, T.M. Briere, Y. Wang, J.B. Son, M.D. Pagel, J. Ma, J. Li, MetNet: Computer-aided segmentation of brain metastases in post-contrast T1-weighted magnetic resonance imaging, Radiother Oncol. 153 (2020) 189–196. https://doi.org/10.1016/j.radonc.2020.09.016.
[32] Z. Zhou, J.W. Sanders, J.M. Johnson, M.K. Gule-Monroe, M.M. Chen, T.M. Briere, Y. Wang, J.B. Son, M.D. Pagel, J. Li, J. Ma, Computer-aided Detection of Brain Metastases in T1-weighted MRI for Stereotactic Radiosurgery Using Deep Learning Single-Shot Detectors, Radiology. 295 (2020) 407–415. https://doi.org/10.1148/radiol.2020191479.
[33] S. Chakrabarty, A. Sotiras, M. Milchenko, P. LaMontagne, M. Hileman, D. Marcus, MRI-based Identification and Classification of Major Intracranial Tumor Types by Using a 3D Convolutional Neural Network: A Retrospective Multi-institutional Analysis, Radiology Artif Intell. 3 (2021) e200301. https://doi.org/10.1148/ryai.2021200301.
[34] A.M. Rauschecker, T. Gleason, P. Nedelec, M.T. Duong, D. Weiss, E. Calabrese, J. Colby, L.P. Sugrue, J.D. Rudie, C.P. Hess, Interinstitutional Portability of a Deep Learning Brain MRI Lesion Segmentation Algorithm, Radiology Artif Intell. (2021). https://doi.org/10.1148/ryai.2021200152.
[35] A.R. Groendahl, I.S. Knudtsen, B.N. Huynh, M. Mulstad, Y.M. Moe, F. Knuth, O. Tomic, U.G. Indahl, T. Torheim, E. Dale, E. Malinen, C.M. Futsaether, A comparison of methods for fully automatic segmentation of tumors and involved nodes in PET/CT of head and neck cancers, Phys Medicine Biology. 66 (2021) 065012. https://doi.org/10.1088/1361-6560/abe553.
[36] V. Andrearczyk, V. Oreiller, S. Boughdad, C.C.L. Rest, H. Elhalawani, M. Jreige, J.O. Prior, M. Vallières, D. Visvikis, M. Hatt, A. Depeursinge, Overview of the HECKTOR challenge at MICCAI 2021: Automatic Head and Neck Tumor Segmentation and Outcome Prediction in PET/CT images., LNCS Challenges. (2022).
[37] P. Fontaine, V. Andrearczyk, V. Oreiller, J. Castelli, M. Jreige, J.O. Prior, A. Depeursinge, Multimodal Learning for Clinical Decision Support, 11th International Workshop, ML-CDS 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, October 1, 2021, Proceedings, Lect Notes Comput Sc. (2021) 59–68. https://doi.org/10.1007/978-3-030-89847-2_6.
[38] V. Andrearczyk, P. Fontaine, V. Oreiller, J. Castelli, M. Jreige, J.O. Prior, A. Depeursinge, Predictive Intelligence in Medicine, 4th International Workshop, PRIME 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, October 1, 2021, Proceedings, Lect Notes Comput Sc. (2021) 147–156. https://doi.org/10.1007/978-3-030-87602-9_14.
[39] K.A. Wahid, S. Ahmed, R. He, L.V. van Dijk, J. Teuwen, B.A. McDonald, V. Salama, A.S.R. Mohamed, T. Salzillo, C. Dede, N. Taku, S.Y. Lai, C.D. Fuller, M.A. Naser, Evaluation of Deep Learning-Based Multiparametric MRI Oropharyngeal Primary Tumor Auto-Segmentation and Investigation of Input Channel Effects: Results from a Prospective Imaging Registry, Clin Transl Radiat Oncol. (2021). https://doi.org/10.1016/j.ctro.2021.10.003.
[40] R.R. Outeiral, P. Bos, A. Al-Mamgani, B. Jasperse, R. Simões, U.A. van der Heide, Oropharyngeal primary tumor segmentation for radiotherapy planning on magnetic resonance imaging using deep learning, Phys Imaging Radiat Oncol. 19 (2021) 39–44. https://doi.org/10.1016/j.phro.2021.06.005.
[41] V.A.V. Oreiller1, V. Oreiller, 2 Martin Valli`eres3, 4 Joel Castelli5, 7 Hesham Elhalawani8 Mario Jreige2 Sarah Boughdad2 John O. Prior2 Adrien, Automatic Segmentation of Head and

Neck Tumors and Nodal Metastases in PET-CT scans, MIDL. (2020).
[42] Y.M. Moe, A.R. Groendahl, M. Mulstad, O. Tomic, U. Indahl, E. Dale, E. Malinen, C.M. Futsaether, Deep learning for automatic tumour segmentation in PET/CT images of patients with head and neck cancers, Arxiv. (2019).
[43] V. Andrearczyk, Segmentation and classification of head and neck nodal metastases and primary tumors in PET/CT (under review), ISBI. (2022).
[44] C.E. Cardenas, B.M. Beadle, A.S. Garden, H.D. Skinner, J. Yang, D.J. Rhee, R.E. McCarroll, T.J. Netherton, S.S. Gay, L. Zhang, L.E. Court, Generating High-Quality Lymph Node Clinical Target Volumes for Head and Neck Cancer Radiation Therapy Using a Fully Automated Deep Learning-Based Approach, Int J Radiat Oncol Biology Phys. 109 (2021) 801–812. https://doi.org/10.1016/j.ijrobp.2020.10.005.
[45] E. Prezioso, S. Izzo, F. Giampaolo, F. Piccialli, G.D. Orabona, R. Cuocolo, V. Abbate, L. Ugga, L. Califano, Predictive Medicine for Salivary gland tumours identification through Deep Learning, Ieee J Biomed Health. PP (2021) 1–1. https://doi.org/10.1109/jbhi.2021.3120178.
[46] Y. Qi, J. Li, H. Chen, Y. Guo, Y. Yin, G. Gong, L. Wang, Computer-aided diagnosis and regional segmentation of nasopharyngeal carcinoma based on multi-modality medical images, Int J Comput Ass Rad. 16 (2021) 871–882. https://doi.org/10.1007/s11548-021-02351-y.
[47] S. Li, J. Xiao, L. He, X. Peng, X. Yuan, The Tumor Target Segmentation of Nasopharyngeal Cancer in CT Images Based on Deep Learning Methods, Technol Cancer Res T. 18 (2019) 1533033819884561. https://doi.org/10.1177/1533033819884561.
[48] W. Li, S. Cheng, K. Qian, K. Yue, H. Liu, Automatic Recognition and Classification System of Thyroid Nodules in CT Images Based on CNN, Comput Intel Neurosc. 2021 (2021) 1–11. https://doi.org/10.1155/2021/5540186.
[49] C.-F. Chen, M.-H. Lin, K.-A. Chu, W.-S. Liu, S.-H. Hsiao, R.-S. Lai, Effective cardiac radiotherapy relieved life-threatening heart failure caused by advanced small cell lung cancer with cardiac metastasis: a case report, Journal of Thoracic Disease. 10 (2018) E250.
[50] R. Finnegan, J. Dowling, E.-S. Koh, S. Tang, J. Otton, G. Delaney, V. Batumalai, C. Luo, P. Atluri, A. Satchithanandha, Feasibility of multi-atlas cardiac segmentation from thoracic planning CT in a probabilistic framework, Physics in Medicine & Biology. 64 (2019) 85006.
[51] R.N. Finnegan, L. Orlandini, X. Liao, J. Yin, J. Lang, J. Dowling, D. Fontanarosa, Feasibility of using a novel automatic cardiac segmentation algorithm in the clinical routine of lung cancer patients, Plos One. 16 (2021) e0245364.
[52] M. Farrugia, H. Yu, A.K. Singh, H. Malhotra, Autosegmentation of cardiac substructures in respiratory-gated, non-contrasted computed tomography images, World Journal of Clinical Oncology. 12 (2021) 95.
[53] N. Maffei, L. Fiorini, G. Aluisio, E. D’Angelo, P. Ferrazza, V. Vanoni, F. Lohr, B. Meduri, G. Guidi, Hierarchical clustering applied to automatic atlas based segmentation of 25 cardiac sub-structures, Physica Medica. 69 (2020) 70–80.
[54] A. McWilliam, J. Khalifa, E.V. Osorio, K. Banfill, A. Abravan, C. Faivre-Finn, M. van Herk, Novel methodology to investigate the effect of radiation dose to heart substructures on overall survival, International Journal of Radiation Oncology* Biology* Physics. 108 (2020) 1073–1081.
[55] C. Wang, N. Tyagi, A. Rimner, Y.-C. Hu, H. Veeraraghavan, G. Li, M. Hunt, G. Mageras, P. Zhang, Segmenting lung tumors on longitudinal imaging studies via a patient-specific adaptive convolutional neural network, Radiotherapy and Oncology. 131 (2019) 101–107.

[56] F. Zhang, Q. Wang, H. Li, Automatic segmentation of the gross target volume in non-small cell lung cancer using a modified version of resNet, Technology in Cancer Research & Treatment. 19 (2020) 1533033820947484.
[57] T. Pohlen, A. Hermans, M. Mathias, B. Leibe, Full-resolution residual networks for semantic segmentation in street scenes, in: 2017: pp. 4151–4160.
[58] J. Jiang, Y. Hu, C.-J. Liu, D. Halpenny, M.D. Hellmann, J.O. Deasy, G. Mageras, H. Veeraraghavan, Multiple resolution residually connected feature streams for automatic lung tumor segmentation from CT images, IEEE Transactions on Medical Imaging. 38 (2018) 134–144.
[59] X. Zhao, L. Li, W. Lu, S. Tan, Tumor co-segmentation in PET/CT using multi-modality fully convolutional neural network, Physics in Medicine & Biology. 64 (2018) 15011.
[60] S. Mercieca, J.S. Belderbos, M. van Herk, Challenges in the target volume definition of lung cancer radiotherapy, Translational Lung Cancer Research. 10 (2021) 1983.
[61] S. Mercieca, J.S.A. Belderbos, M. van Herk, Challenges in the target volume definition of lung cancer radiotherapy, Transl Lung Cancer Res. 10 (2020) 1983–1998. https://doi.org/10.21037/tlcr-20-627.
[62] J. Wong, V. Huang, J.A. Giambattista, T. Teke, C. Kolbeck, J. Giambattista, S. Atrchian, Training and Validation of Deep Learning-Based Auto-Segmentation Models for Lung Stereotactic Ablative Radiotherapy Using Retrospective Radiotherapy Planning Contours, Frontiers Oncol. 11 (2021) 626499. https://doi.org/10.3389/fonc.2021.626499.
[63] D. Jin, D. Guo, T.-Y. Ho, A.P. Harrison, J. Xiao, C.-K. Tseng, L. Lu, DeepTarget: Gross tumor and clinical target volume segmentation in esophageal cancer radiotherapy, Medical Image Analysis. 68 (2021) 101909.
[64] S. Yousefi, H. Sokooti, M.S. Elmahdy, I.M. Lips, M.T.M. Shalmani, R.T. Zinkstok, F.J. Dankers, M. Staring, Esophageal Tumor Segmentation in CT Images Using a Dilated Dense Attention Unet (DDAUnet), IEEE Access. 9 (2021) 99235–99248.
[65] Z. Zhu, D. Jin, K. Yan, T.-Y. Ho, X. Ye, D. Guo, C.-H. Chao, J. Xiao, A. Yuille, L. Lu, Medical Image Computing and Computer Assisted Intervention – MICCAI 2020, 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part VII, Lect Notes Comput Sc. (2020) 753–762. https://doi.org/10.1007/978-3-030-59728-3_73.
[66] N. Heller, F. Isensee, K.H. Maier-Hein, X. Hou, C. Xie, F. Li, Y. Nan, G. Mu, Z. Lin, M. Han, G. Yao, Y. Gao, Y. Zhang, Y. Wang, F. Hou, J. Yang, G. Xiong, J. Tian, C. Zhong, J. Ma, J. Rickman, J. Dean, B. Stai, R. Tejpaul, M. Oestreich, P. Blake, H. Kaluzniak, S. Raza, J. Rosenberg, K. Moore, E. Walczak, Z. Rengel, Z. Edgerton, R. Vasdev, M. Peterson, S. McSweeney, S. Peterson, A. Kalapara, N. Sathianathen, N. Papanikolopoulos, C. Weight, The state of the art in kidney and kidney tumor segmentation in contrast-enhanced CT imaging: Results of the KiTS19 challenge, Med Image Anal. 67 (2021) 101821. https://doi.org/10.1016/j.media.2020.101821.
[67] F. Isensee, J. Petersen, A. Klein, D. Zimmerer, P.F. Jaeger, S. Kohl, J. Wasserthal, G. Koehler, T. Norajitra, S. Wirkert, nnu-net: Self-adapting framework for u-net-based medical image segmentation, ArXiv Preprint ArXiv:1809.10486. (2018).
[68] F. Isensee, K.H. Maier-Hein, An attempt at beating the 3D U-Net, ArXiv Preprint ArXiv:1908.02182. (2019).
[69] Y. Liang, D. Schott, Y. Zhang, Z. Wang, H. Nasief, E. Paulson, W. Hall, P. Knechtges, B.

Erickson, X.A. Li, Auto-segmentation of pancreatic tumor in multi-parametric MRI using deep convolutional neural networks, Radiotherapy and Oncology. 145 (2020) 193–200.
[70] Y. Iwasa, T. Iwashita, Y. Takeuchi, H. Ichikawa, N. Mita, S. Uemura, M. Shimizu, Y.-T. Kuo, H.-P. Wang, T. Hara, Automatic Segmentation of Pancreatic Tumors Using Deep Learning on a Video Image of Contrast-Enhanced Endoscopic Ultrasound, Journal of Clinical Medicine. 10 (2021) 3589.
[71] P. Bilic, P.F. Christ, E. Vorontsov, G. Chlebus, H. Chen, Q. Dou, C.-W. Fu, X. Han, P.-A. Heng, J. Hesser, The liver tumor segmentation benchmark (lits), ArXiv Preprint ArXiv:1901.04056. (2019).
[72] L. Soler, A. Hostettler, V. Agnus, A. Charnoz, J. Fasquel, J. Moreau, A. Osswald, M. Bouhadjar, J. Marescaux, 3D image reconstruction for comparison of algorithm database: A patient specific anatomical and medical image database, IRCAD, Strasbourg, France, Tech. Rep. (2010).
[73] S. Almotairi, G. Kareem, M. Aouf, B. Almutairi, M.A.-M. Salem, Liver tumor segmentation in CT scans using modified SegNet, Sensors. 20 (2020) 1516.
[74] G. Chlebus, H. Meine, J.H. Moltz, A. Schenk, Neural network-based automatic liver tumor segmentation with random forest-based candidate filtering, ArXiv Preprint ArXiv:1706.00842. (2017).
[75] P.F. Christ, M.E.A. Elshaer, F. Ettlinger, S. Tatavarty, M. Bickel, P. Bilic, M. Rempfler, M. Armbruster, F. Hofmann, M. D'Anastasi, Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields, in: Springer, 2016: pp. 415–423.
[76] W. Li, Automatic segmentation of liver tumor in CT images with deep convolutional neural networks, Journal of Computer and Communications. 3 (2015) 146.
[77] Y. Yuan, Hierarchical convolutional-deconvolutional neural networks for automatic liver and tumor segmentation, ArXiv Preprint ArXiv:1710.04540. (2017).
[78] T. Liu, J. Liu, Y. Ma, J. He, J. Han, X. Ding, C. Chen, Spatial feature fusion convolutional network for liver and liver tumor segmentation from CT images, Medical Physics. 48 (2021) 264–272.
[79] Y. Xu, M. Cai, L. Lin, Y. Zhang, H. Hu, Z. Peng, Q. Zhang, Q. Chen, X. Mao, Y. Iwamoto, PA-ResSeg: A phase attention residual network for liver tumor segmentation from multiphase CT images, Medical Physics. (2021).
[80] H. Chen, Q. Dou, L. Yu, J. Qin, P.-A. Heng, VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images, NeuroImage. 170 (2018) 446–455.
[81] J. Zhao, D. Li, X. Xiao, F. Accorsi, H. Marshall, T. Cossetto, D. Kim, D. McCarthy, C. Dawson, S. Knezevic, United adversarial learning for liver tumor segmentation and detection of multi-modality non-contrast MRI, Medical Image Analysis. 73 (2021) 102154.
[82] E. Cha, S. Elguindi, I. Onochie, D. Gorovets, J.O. Deasy, M. Zelefsky, E.F. Gillespie, Clinical implementation of deep learning contour autosegmentation for prostate radiotherapy, Radiother Oncol. 159 (2021) 1–7. https://doi.org/10.1016/j.radonc.2021.02.040.
[83] C.-C. Lai, H.-K. Wang, F.-N. Wang, Y.-C. Peng, T.-P. Lin, H.-H. Peng, S.-H. Shen, Autosegmentation of Prostate Zones and Cancer Regions from Biparametric Magnetic Resonance Images by Using Deep-Learning-Based Neural Networks, Sensors. 21 (2021) 2709. https://doi.org/10.3390/s21082709.

[84] M. Bardis, R. Houshyar, C. Chantaduly, K. Tran-Harding, A. Ushinsky, C. Chahine, M. Rupasinghe, D. Chow, P. Chang, Segmentation of the Prostate Transition Zone and Peripheral Zone on MR Images with Deep Learning, Radiology Imaging Cancer. 3 (2021) e200024. https://doi.org/10.1148/rycan.2021200024.
[85] A. Balagopal, H. Morgan, M. Dohopolski, R. Timmerman, J. Shan, D.F. Heitjan, W. Liu, D. Nguyen, R. Hannan, A. Garant, N. Desai, S. Jiang, PSA-Net: Deep learning–based physician style–aware segmentation network for postoperative prostate cancer clinical target volumes, Artif Intell Med. 121 (2021) 102195. https://doi.org/10.1016/j.artmed.2021.102195.
[86] A. Balagopal, D. Nguyen, H. Morgan, Y. Weng, M. Dohopolski, M.-H. Lin, A.S. Barkousaraie, Y. Gonzalez, A. Garant, N. Desai, R. Hannan, S. Jiang, A deep learning-based framework for segmenting invisible clinical target volumes with estimated uncertainties for post-operative prostate cancer radiotherapy, Med Image Anal. 72 (2021) 102101. https://doi.org/10.1016/j.media.2021.102101.
[87] M. Moazzezi, B. Rose, K. Kisling, K.L. Moore, X. Ray, Prospects for daily online adaptive radiotherapy via ethos for prostate cancer patients without nodal involvement using unedited CBCT auto-segmentation, J Appl Clin Med Phys. 22 (2021) 82–93. https://doi.org/10.1002/acm2.13399.
[88] Z. Wang, Y. Chang, Z. Peng, Y. Lv, W. Shi, F. Wang, X. Pei, X.G. Xu, Evaluation of deep learning-based auto-segmentation algorithms for delineating clinical target volume and organs at risk involving data for 125 cervical cancer patients, J Appl Clin Med Phys. 21 (2020) 272–279. https://doi.org/10.1002/acm2.13097.
[89] Y. Chang, Z. Wang, Z. Peng, J. Zhou, Y. Pi, X.G. Xu, X. Pei, Clinical application and improvement of a CNN-based autosegmentation model for clinical target volumes in cervical cancer radiotherapy, J Appl Clin Med Phys. (2021). https://doi.org/10.1002/acm2.13440.
[90] Z. Liu, W. Chen, H. Guan, H. Zhen, J. Shen, X. Liu, A. Liu, R. Li, J. Geng, J. You, W. Wang, Z. Li, Y. Zhang, Y. Chen, J. Du, Q. Chen, Y. Chen, S. Wang, F. Zhang, J. Qiu, An Adversarial Deep-Learning-Based Model for Cervical Cancer CTV Segmentation With Multicenter Blinded Randomized Controlled Validation, Frontiers Oncol. 11 (2021) 702270. https://doi.org/10.3389/fonc.2021.702270.
[91] H. Hu, Q. Yang, J. Li, P. Wang, B. Tang, X. Wang, J. Lang, Deep learning applications in automatic segmentation and reconstruction in CT-based cervix brachytherapy, J Contemp Brachyther. 13 (2021) 325–330. https://doi.org/10.5114/jcb.2021.106118.
[92] C. McIntosh, L. Conroy, M.C. Tjong, T. Craig, A. Bayley, C. Catton, M. Gospodarowicz, J. Helou, N. Isfahanian, V. Kong, T. Lam, S. Raman, P. Warde, P. Chung, A. Berlin, T.G. Purdie, Clinical integration of machine learning for curative-intent radiation treatment of patients with prostate cancer, Nat Med. (2021) 1–7. https://doi.org/10.1038/s41591-021-01359-w.
[93] E. Wu, K. Wu, R. Daneshjou, D. Ouyang, D.E. Ho, J. Zou, How medical AI devices are evaluated: limitations and recommendations from an analysis of FDA approvals, Nat Med. (2021) 1--3. https://doi.org/10.1038/s41591-021-01312-x.
[94] M. Roberts, D. Driggs, M. Thorpe, J. Gilbey, M. Yeung, S. Ursprung, A.I. Aviles-Rivero, C. Etmann, C. McCague, L. Beer, J.R. Weir-McCall, Z. Teng, E. Gkrania-Klotsas, J.H.F. Rudd, E. Sala, C.-B. Schonlieb, Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans, Nat Mach Intell. 3 (2021) 199--217. https://doi.org/10.1038/s42256-021-00307-0.

[95] S. Warnat-Herresthal, H. Schultze, K.L. Shastry, S. Manamohan, S. Mukherjee, V. Garg, R. Sarveswara, K. Händler, P. Pickkers, N.A. Aziz, S. Ktena, F. Tran, M. Bitzer, S. Ossowski, N. Casadei, C. Herr, D. Petersheim, U. Behrends, F. Kern, T. Fehlmann, P. Schommers, C. Lehmann, M. Augustin, J. Rybniker, J. Altmüller, N. Mishra, J.P. Bernardes, B. Krämer, L. Bonaguro, J. Schulte-Schrepping, E.D. Domenico, C. Siever, M. Kraut, M. Desai, B. Monnet, M. Saridaki, C.M. Siegel, A. Drews, M. Nuesch-Germano, H. Theis, J. Heyckendorf, S. Schreiber, S. Kim-Hellmuth, C.-19 A.S. (COVAS), P. Balfanz, T. Eggermann, P. Boor, R. Hausmann, H. Kuhn, S. Isfort, J.C. Stingl, G. Schmalzing, C.K. Kuhl, R. Röhrig, G. Marx, S. Uhlig, E. Dahl, D. Müller-Wieland, M. Dreher, N. Marx, J. Nattermann, D. Skowasch, I. Kurth, A. Keller, R. Bals, P. Nürnberg, O. Rieß, P. Rosenstiel, M.G. Netea, F. Theis, S. Mukherjee, M. Backes, A.C. Aschenbrenner, T. Ulas, D.C.-19 O.I. (DeCOI), A. Angelov, A. Bartholomäus, A. Becker, D. Bezdan, C. Blumert, E. Bonifacio, P. Bork, B. Boyke, H. Blum, T. Clavel, M. Colome-Tatche, M. Cornberg, I.A.D.L.R. Velázquez, A. Diefenbach, A. Dilthey, N. Fischer, K. Förstner, S. Franzenburg, J.-S. Frick, G. Gabernet, J. Gagneur, T. Ganzenmueller, M. Gauder, J. Geißert, A. Goesmann, S. Göpel, A. Grundhoff, H. Grundmann, T. Hain, F. Hanses, U. Hehr, A. Heimbach, M. Hoeper, F. Horn, D. Hübschmann, M. Hummel, T. Iftner, A. Iftner, T. Illig, S. Janssen, J. Kalinowski, R. Kallies, B. Kehr, O.T. Keppler, C. Klein, M. Knop, O. Kohlbacher, K. Köhrer, J. Korbel, P.G. Kremsner, D. Kühnert, M. Landthaler, Y. Li, K.U. Ludwig, O. Makarewicz, M. Marz, A.C. McHardy, C. Mertes, M. Münchhoff, S. Nahnsen, M. Nöthen, F. Ntoumi, J. Overmann, S. Peter, K. Pfeffer, I. Pink, A.R. Poetsch, U. Protzer, A. Pühler, N. Rajewsky, M. Ralser, K. Reiche, S. Ripke, U.N. da Rocha, A.-E. Saliba, L.E. Sander, B. Sawitzki, S. Scheithauer, P. Schiffer, J. Schmid-Burgk, W. Schneider, E.-C. Schulte, A. Sczyrba, M.L. Sharaf, Y. Singh, M. Sonnabend, O. Stegle, J. Stoye, J. Vehreschild, T.P. Velavan, J. Vogel, S. Volland, M. von Kleist, A. Walker, J. Walter, D. Wieczorek, S. Winkler, J. Ziebuhr, M.M.B. Breteler, E.J. Giamarellos-Bourboulis, M. Kox, M. Becker, S. Cheran, M.S. Woodacre, E.L. Goh, J.L. Schultze, Swarm Learning for decentralized and confidential clinical machine learning, Nature. 594 (2021) 265–270. https://doi.org/10.1038/s41586-021-03583-3.
[96] V. Sandfort, K. Yan, P.M. Graffy, P.J. Pickhardt, R.M. Summers, Use of Variational Autoencoders with Unsupervised Learning to Detect Incorrect Organ Segmentations at CT, Radiology Artif Intell. 3 (2021) e200218. https://doi.org/10.1148/ryai.2021200218.
[97] T. Sakinis, F. Milletari, H. Roth, P. Korfiatis, P. Kostandy, K. Philbrick, Z. Akkus, Z. Xu, D. Xu, B.J. Erickson, Interactive segmentation of medical images through fully convolutional neural networks, ArXiv. (2019).
[98] C.-H. Chao, Y.-C. Cheng, H.-T. Cheng, C.-W. Huang, T.-Y. Ho, C.-K. Tseng, L. Lu, M. Sun, Radiotherapy Target Contouring with Convolutional Gated Graph Neural Network, Arxiv. (2019).
[99] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: Visual Explanations from Deep Networks Via Gradient-Based Localization, 2017 Ieee Int Conf Comput Vis Iccv. (2017) 618–626. https://doi.org/10.1109/iccv.2017.74.
[100] N. Arun, N. Gaw, P. Singh, K. Chang, M. Aggarwal, B. Chen, K. Hoebel, S. Gupta, J. Patel, M. Gidwani, J. Adebayo, M.D. Li, J. Kalpathy-Cramer, Assessing the Trustworthiness of Saliency Maps for Localizing Abnormalities in Medical Imaging, Radiology Artif Intell. 3 (2021). https://doi.org/10.1148/ryai.2021200267.
[101] M. Graziani, V. Andrearczyk, H. Müller, Understanding and Interpreting Machine Learning in Medical Image Computing Applications, First International Workshops, MLCN 2018, DLF

2018, and iMIMIC 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16-20, 2018, Proceedings, Lect Notes Comput Sc. (2018) 124–132. https://doi.org/10.1007/978-3-030-02628-8_14.
[102] H. Yeche, J. Harrison, T. Berthier, Interpretability of Machine Intelligence in Medical Image Computing and Multimodal Learning for Clinical Decision Support, Second International Workshop, iMIMIC 2019, and 9th International Workshop, ML-CDS 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 17, 2019, Proceedings, Lect Notes Comput Sc. (2019) 12–20. https://doi.org/10.1007/978-3-030-33850-3_2.
[103] W. Luo, Y. Li, R. Urtasun, R. Zemel, Understanding the Effective Receptive Field in Deep Convolutional Neural Networks, Arxiv. (2017).
[104] B. McCrindle, K. Zukotynski, T.E. Doyle, M.D. Noseworthy, A Radiology-focused Review of Predictive Uncertainty for AI Interpretability in Computer-assisted Segmentation, Radiology Artif Intell. (2021) e210031. https://doi.org/10.1148/ryai.2021210031.
[105] K. Hoebel, V. Andrearczyk, A.L. Beers, J.B. Patel, K. Chang, A. Depeursinge, H. Mueller, J. Kalpathy-Cramer, An exploration of uncertainty information for segmentation quality assessment, Medical Imaging 2020 Image Process. (2020) 55. https://doi.org/10.1117/12.2548722.