


ONCOLOGY

Open Access



Development and evaluation of two open-source nnU-Net models for automatic segmentation of lung tumors on PET and CT images with and without respiratory motion compensation

Montserrat Carles^{1*} , Dejan Kuhn^{2,3}, Tobias Fechter^{2,3}, Dimos Baltas^{2,3}, Michael Mix⁴, Ursula Nestle^{3,5,6}, Anca L. Grosu^{3,5}, Luis Martí-Bonmatí¹, Gianluca Radicioni^{3,5} and Eleni Gkika^{3,5}

Abstract

Objectives In lung cancer, one of the main limitations for the optimal integration of the biological and anatomical information derived from Positron Emission Tomography (PET) and Computed Tomography (CT) is the time and expertise required for the evaluation of the different respiratory phases. In this study, we present two open-source models able to automatically segment lung tumors on PET and CT, with and without motion compensation.

Materials and methods This study involved time-bin gated (4D) and non-gated (3D) PET/CT images from two prospective lung cancer cohorts (Trials 108237 and 108472) and one retrospective. For model construction, the ground truth (GT) was defined by consensus of two experts, and the nnU-Net with 5-fold cross-validation was applied to 560 4D-images for PET and 100 3D-images for CT. The test sets included 270 4D- images and 19 3D-images for PET and 80 4D-images and 27 3D-images for CT, recruited at 10 different centres.

Results In the performance evaluation with the multicentre test sets, the Dice Similarity Coefficients (DSC) obtained for our PET model were $DSC(4D-PET) = 0.74 \pm 0.06$, improving 19% relative to the DSC between experts and $DSC(3D-PET) = 0.82 \pm 0.11$. The performance for CT was $DSC(4D-CT) = 0.61 \pm 0.28$ and $DSC(3D-CT) = 0.63 \pm 0.34$, improving 4% and 15% relative to DSC between experts.

Conclusions Performance evaluation demonstrated that the automatic segmentation models have the potential to achieve accuracy comparable to manual segmentation and thus hold promise for clinical application. The resulting models can be freely downloaded and employed to support the integration of 3D- or 4D- PET/CT and to facilitate the evaluation of its impact on lung cancer clinical practice.

Clinical relevance statement We provide two open-source nnU-Net models for the automatic segmentation of lung tumors on PET/CT to facilitate the optimal integration of biological and anatomical information in clinical practice. The models have superior performance compared to the variability observed in manual segmentations by the different experts for images with and without motion compensation, allowing to take advantage in the clinical practice of the more accurate and robust 4D-quantification.

Gianluca Radicioni and Eleni Gkika contributed equally to this publication.

*Correspondence:

Montserrat Carles

montserrat_carles@iislafe.es

Full list of author information is available at the end of the article



© The Author(s) 2024, corrected publication 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Key Points

- Lung tumor segmentation on PET/CT imaging is limited by respiratory motion and manual delineation is time consuming and suffer from inter- and intra-variability.
- Our segmentation models had superior performance compared to the manual segmentations by different experts.
- Automating PET image segmentation allows for easier clinical implementation of biological information.

Keywords Lung cancer, Positron emission tomography, Computed tomography, Deep learning, Respiratory motion

Introduction

In lung cancer, the ability of Positron Emission Tomography (PET) with [¹⁸F]fluoro-2-deoxy-D-glucose (FDG) to exploit the biochemical differences between normal and neoplastic tissue [1] has been proven to be a valuable tool for tumor detection [2], staging [3], treatment planning [4], monitoring [5] and outcome prediction [6, 7]. Based on the sensitivity and specificity implied by FDG-PET imaging in NSCLC [8], the integration of FDG imaging in radiation therapy (RT) clinical practice, which has been conventionally CT-only based, has been recommended [9]. Recent studies have evaluated the role of [¹⁸F]FDG PET/CT in the precise definition not only of the target volume, but also of subvolumes aiming at a further dose escalation [10–13]. However, the evaluation of the lesion by multimodality PET and Computed Tomography (CT) imaging presents challenges due to respiratory movement [14]. Tumor motion due to the various breathing cycles involved during PET scan acquisition results in inaccurate quantification of tracer distribution, including erroneous estimation of the shape, volume and location of the lesion. Moreover, the different acquisition times of the two studies (few seconds for a CT and several minutes for PET) can result in a spatial mismatch, with artefacts originated by the use of CT images for attenuation correction in PET image reconstruction. Different strategies have been proposed for the management of respiratory motion in PET/CT systems [15–17], such as retrospectively respiratory gated (4D)-PET/CT [18]. As a result of this data processing, an improvement in the image quality and accuracy of estimation of the tracer concentration and distribution should be obtained by compensating for motion effects [19, 20]. The definition of the target volume in RT includes safety margins around the tumor given by geometrical uncertainties (as for example, respiratory motion [21, 22]). The aim of these safety margins is to minimise tumor underdosage. 4D-protocols have been proven to decrease geometrical uncertainties due to respiratory motion and consequently, their corresponding safety margins could be reduced. Therefore, 4D-protocols would lead to a

smaller target volume definition and the dose delivered at the organs at risk in the vicinity could be minimised, facilitating dose escalation. In addition, 4D-protocols have been demonstrated to significantly improve the reproducibility of radiomics features [23, 24].

Manual contouring is the method most widely employed for tumor segmentation in RT target delineation, monitoring and radiomics. For these applications, higher accuracy and reproducibility are required beyond what is required in diagnosis. Disadvantages related to the manual approach are to be very laborious and time consuming, especially for the number of images involved in 4D-protocols and to lead to a significant inter- and intra-variability even among experts. An alternative approach consists in relying on automatic or semi-automatic segmentation methods. PET segmentation methods have been proposed [25], ranging from simple uptake thresholding to very elaborate probabilistic models. Applications of artificial intelligence (AI, machine/deep learning) are extremely wide and promising [26, 27]. Superiority of these methods in image segmentation has been shown in many studies [28]. Although the number of automatic AI segmentation algorithms for lung cancer is increasing, most of them focus on CT or MR images [29, 30]. In addition, the AI segmentation models already published for PET have rarely involved 4DPET data [31, 32].

In this study, our primary aim was to develop an AI model for lung tumor segmentation on PET images. In order to facilitate the implementation of the PET segmentation in the clinical workflow, a secondary aim was to develop also an AI model for tumor segmentation on CT images, which is the image modality most commonly employed. In addition, to facilitate the simultaneous implementation of both models, the same open-source convolutional-neuronal-network (nnU-Net) was employed. Finally, in order to maximise the probability that the resulting models have good performance independently of imaging systems and protocols (generalizability), image data was collected from ten different institutions with different involved protocols with (4D) and without (3D) respiratory motion compensation.

Materials and methods

Patient cohorts and data sets

Our study included two prospective (PET-Plan and STRIPE) and one retrospective (RC) cohort of lung cancer patients. All patients gave written informed consent according to institutional and federal guidelines. The institutional ethics committee approved the study protocol (EK-Nr 21-1228-S1-retro, EK Nr 108237, EK 113/12, EK-Nr Nr. 108472).

The main clinical characteristics of each trial are described in the following subsections.

PET-Plan trial

The randomised controlled PET-Plan trial (ARO 2009-09, NCT00697333, Deutsche Krebshilfe, German Cancer Aid Organisation, Nr 108237) involved patients older than 18 years with histologically or cytologically proven inoperable stage II or III NSCLC. Eligibility criteria also included having an Eastern Cooperative Oncology Group performance status of less than 3; having adequate pulmonary, cardiac, renal and haematological function and being suitable for chemoradiotherapy. The PET-Plan trial was conducted in 24 centres in Germany, Austria and Switzerland, in accordance with the Declaration of Helsinki. Study design, procedures and main outcome results have been published elsewhere [9].

STRIPE trial

The prospective monocentre phase II STRIPE trial (Deutsche Krebshilfe, German Cancer Aid Organisation, Nr 108472) involved patients with pulmonary lesions with a maximum diameter of 5 cm (early stage NSCLC or less than 2 pulmonary metastases of a controlled primary tumor), refusing surgery or inoperable due to comorbidities. The study was performed at Medical University Freiburg and the design, procedures and main outcome results have been published elsewhere [33].

Retrospective cohort

For the retrospective cohort (RC), patients with primary early-stage (N)SCLC or isolated pulmonary metastases were recruited at the Medical University Freiburg between December 2019 and July 2020. All patients were older than 18 years and showed a maximum of two FDG-PET positive pulmonary lesions with a maximum diameter of 5 cm.

Patients from these three cohorts were split into the training and test cohorts. A representation of the cohorts division employed for the development of CT and PET models is shown in Fig. 1. The training set involved 27 patients from PET-Plan trial and 29 from STRIPE for PET model and 41 patients from PET-Plan trial and 59 from STRIPE for CT model, all of them recruited in Medical

Centre Freiburg. The performance of the models was independently tested on a multicentre cohort, which includes prospective and retrospective patients: 36 patients from PET-Plan and 8 from retrospective cohort for PET model and 19 patients from PET-Plan trial and 8 from retrospective cohort for CT model.

PET/CT acquisition

After a fasting period of at least 4 h (glucose level less than 150 mg/dl), all patients underwent a diagnostic whole-body [^{18}F]FDG PET/CT scan at approximately 60 min following weight-adjusted [^{18}F]FDG intravenous injection.

All patients imaged at University Medical Centre Freiburg underwent a chest-limited 4D PET/CT scan for one bed position (15 min acquisition time) at approximately 102 min after FDG injection. Patient respiration was monitored with a belt (Mayo Clinic Respiratory feedback system) and the respiratory curve was synchronised with the acquisition time of scanner. The data was reconstructed in 10 time-based [34] respiratory phases with a fully 3D list mode LOR TOF algorithm, involving a relaxed List Mode Ordered Subset (BLOB-OS-TF), resulting in 10 PET and 10 CT sets. 4DPET was reconstructed in a 144×144 matrix and $4 \times 4 \times 4 \text{ mm}^3$ voxel size and the 4DCT with 512×512 matrix and $1.17 \times 1.17 \times 2 \text{ mm}^3$ voxel size. Attenuation correction of 4DPET was based on the corresponding respiratory-phase 4DCT. Correction for scatter and randoms was applied and the standard uptake value (SUV) was calculated with body weight.

At the University Medical Centre Freiburg scans were performed on two different PET/CT systems from Philips: GEMINI TF TOF 64 (TF64) and GEMINI TF 16 Big Bore (BB). The imaging systems employed in the other nine institutions were: two GEMINI TF TOF 16 from Philips, two SIEMENS Biograph 64, two SIEMENS Biograph 40 mCT, one SIEMENS Biograph 128, one Alegro Body from Philips and one Guardian Body from Philips. The scanners fulfilled the requirements indicated in the European Association of Nuclear Medicine (EANM) imaging guidelines and obtained EANM Research Ltd. (EARL) accreditation. PET image voxel size was $4 \times 4 \times 4 \text{ mm}^3$ for eight patients, $4 \times 4 \times 5 \text{ mm}^3$ for seven patients and $2.65 \times 2.65 \times 3.37 \text{ mm}^3$ for three patients. In CT, voxel sizes varied more significantly across institutions, with a thickness ranging from 1 to 5 mm and transaxial pixel ranging from $0.63 \times 0.63 \text{ mm}^2$ to $1.37 \times 1.37 \text{ mm}^2$.

Segmentation approaches employed for training and validation

Tumor manual segmentation

Manual segmentation of the primary tumor and lymph nodes based on PET and CT was performed according

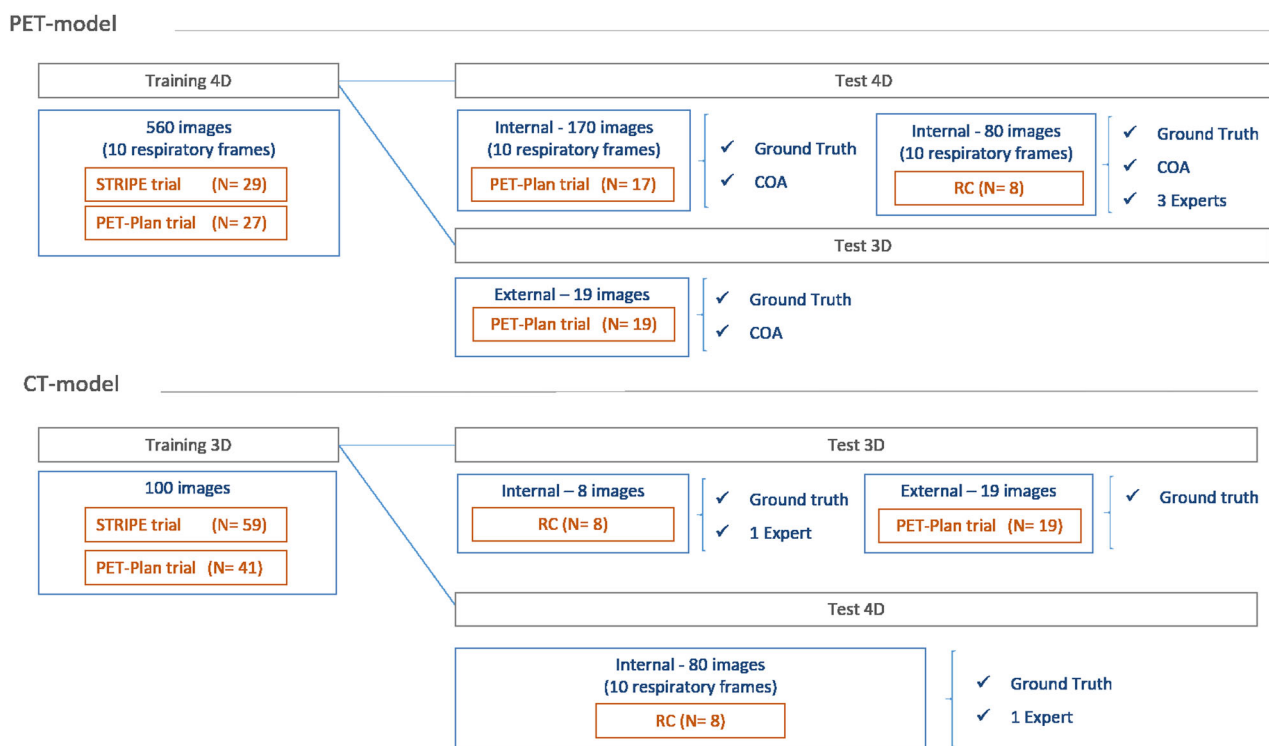


Fig. 1 Representation of the training and test sets with (4D) and without (3D) respiratory motion compensated images. External refers to the sets with patients scanned at institutions different from the one in which training patients were scanned

to the clinical study protocol and current guidelines for contouring with standardised window-level [35]: on CT the pre-set window should range from 600 to 1600 HU for tumours surrounded by lung tissue and from 20 to 400 for lymph nodes and primary tumours invading the mediastinum or chest wall. For PET a window with standardised-uptake-values ranging from 2 to 10 was recommended.

Ground truth was defined by the consensus of two radiation oncologists. One additional expert contoured the 3D- and 4D-CT sets for the eight patients for the internal validation of CT. Three additional experts contoured each of the time-bins for the eight patients for the retrospective cohort employed in the internal evaluation of 4D-PET algorithm. Besides technical instructions for the use of the delineation software, the readers received no further assistance and were blinded to all other delineations.

PET semiautomatic contrast-oriented algorithm

The semiautomatic approach employed for PET segmentation was based on the application of an adaptive threshold taking into account the contrast between tumor concentration (mean value for a 70% isocontour of maximum intensity within the lesion) and background (automatically

derived from the whole image) [36]. The semi-automatic character of this approach relied on the fact that the user was required to indicate a tumor voxel (seed) and maximum intensity was automatically identified in its neighborhood. In order to avoid disconnected regions, a threshold was applied with a six-neighborhood 3D region-growing algorithm from the seed. This approach required phantom-based calibration.

nn-UNet

The nnU-Net [37] was employed for the development of (i) a PET segmentation model, trained with 560 PET respiratory motion compensated images and their corresponding ground truth segmentations and (ii) a CT segmentation model, trained with 100 CT non-motion compensated images and their corresponding ground truth segmentations. nnU-Net is a deep learning-based segmentation method that automatically configures itself, including preprocessing, network architecture, training and post-processing. nnU-Net has a U-Net architecture with plain convolutions, z-score normalisation, leaky ReLU and two blocks (encoder-decoder) per resolution stage. From the three configurations available, the 3D U-Net with full image resolution was employed in this study. Based on the results from a 5-fold cross-validation, the best model was

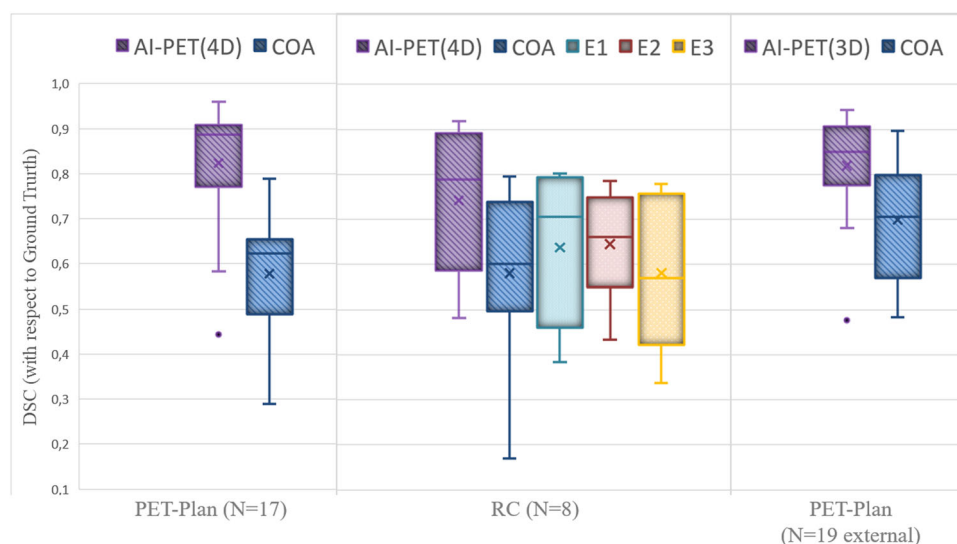


Fig. 2 DSC for the different PET segmentations approaches (artificial- intelligence-algorithm AI, contrast-oriented-algorithm COA and experts E) with respect to the ground truth defined by consensus of two radiation oncologists. Performance for the segmentation model on PET images with (4D) and without (3D) respiratory motion compensation are presented. The average value is represented by the cross and the median by the line

ensemble. The nnU-Net [37] and the resulting models ([GitHub_PETCTLungSegmentationModels_Link](#)) can be free download.

Metrics for accuracy evaluation

We evaluated the agreement between segmented volumes by computing Dice Similarity Coefficient (DSC) [38], which represents the size of the union of two volumes and the formula is $2(A \cap B)/(A + B)$, where A is the ground truth and B the segmented volume. We additionally computed the Positive Predictive Value (PPV) in order to account for the cases in which one set of segmented volumes tends to be included within the other set [36]. The formula for PPV is $(A \cap B)/B = \text{True-Positive}/(\text{True-Positive} + \text{False-Positive})$. The Hausdorff distance [39], which represents the maximum distance between contours surfaces is of interest in order to identify distant regions incorrectly classified as tumour.

Results

PET segmentation model

After a 5-fold cross-validation training with 560 4D-PET images, the resulting optimal PET segmentation model was tested on three test sets, being the ground truth defined by the same radiation oncologists as in the training data. The accuracy of the segmentation provided by the nnU-Net PET model, with respect to the GT, was compared with the accuracy for the semiautomatic contrast-oriented-algorithm (COA) and the accuracy for the segmentation performed by other experts. Results for DSC are shown in Fig. 2.

First, the model was tested with the 170 4D-PET images of patients involved in the prospective PET-Plan trial, the DSC(4D-PET) was 0.83 ± 0.13 , improving 43% with respect to the accuracy for the COA-algorithm, $\text{DSC}(4\text{D-COA}) = 0.58 \pm 0.15$. Second, the model was tested with the 80 4D-PET images of patients involved in the retrospective cohort, the DSC(4D-PET) was 0.74 ± 0.06 , improving 28% with respect to the accuracy for the COA-algorithm with $\text{DSC}(4\text{D-COA}) = 0.58 \pm 0.08$ and improving 16%, 16% and 28% with respect to the accuracy for the three other experts, with $\text{DSC}(4\text{D-Expert1}) = 0.64 \pm 0.06$, $\text{DSC}(4\text{D-Expert2}) = 0.64 \pm 0.09$ and $\text{DSC}(4\text{D-Expert3}) = 0.58 \pm 0.06$. In Fig. 2 it is shown that the model accuracy for the patient of this cohort with the best performance, $\text{DSC}(4\text{D-PET}) = 0.92$, was better than for COA (0.79) and for the three experts (0.80, 0.78 and 0.78) and the DSC for the patient with worst performance by the AI-model (0.48) was also better than for COA (0.38) and experts (0.38, 0.43 and 0.34). In addition, in order to reject distant false positives, the Housdorff distance was computed. Values were comparable across the different approaches 2.1 ± 0.7 for AI-model, 2.5 ± 0.5 for COA and 2.4 ± 0.4 , 2.4 ± 0.5 and 2.6 ± 0.4 for the three experts. An example of the different PET tumor segmentation approaches is shown in Fig. 3.

Finally, the model was tested with the 19 3D-PET images of patients prospectively recruited in 9 different centres involved in the PET-Plan trial. DSC(3D-PET) was 0.83 ± 0.11 , improving 19% with respect to the accuracy for the COA-algorithm, with $\text{DSC}(4\text{D-COA}) = 0.70 \pm 0.12$. It should be remarked that for patients who received whole-body PET scans, the AI-model also segmented the

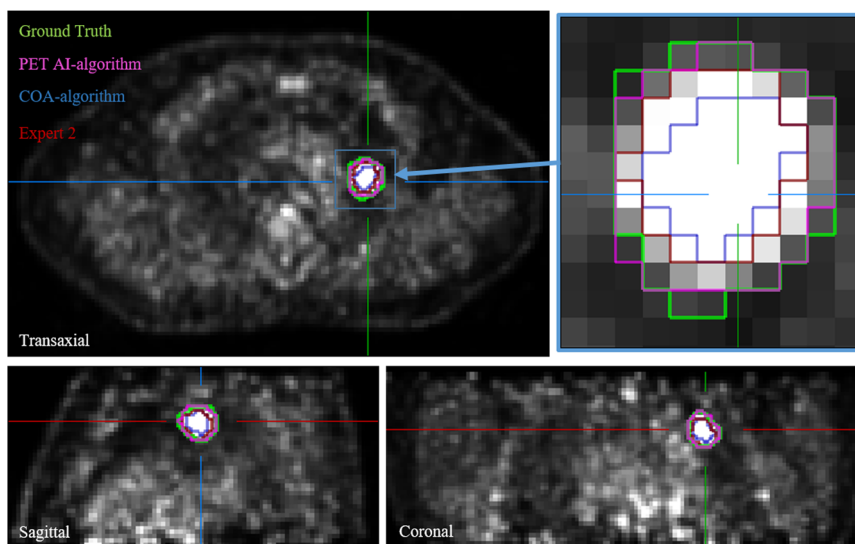


Fig. 3 Different PET segmentations approaches (ground truth, artificial intelligence-algorithm AI, contrast-oriented-algorithm COA and expert with the best DSC with respect to the ground truth) for a patient employed in the validation of the performance of the algorithm on PET images with respiratory motion compensation

bladder and other tracer-enhanced pelvic structures. This limitation was avoided by a simple lung region restriction manually performed by interpolation of two regions of interest, Fig. 4. The time invested for this lung mask was comparable to the time required to place the seeds for the COA algorithm (< 1 min).

CT segmentation model

After a 5-fold cross-validation training with 100 3D-CT images, the resulted optimal CT segmentation model was tested on three test sets. In Fig. 5, values for the DSC are shown.

First, the model was tested with the 8 3D-CT images of patients retrospectively recruited. $DSC(3D-CT)$ was 0.63 ± 0.34 , improving 15% with respect to the experts $DSC(3D-Expert) = 0.55 \pm 0.11$. The nnU-Net model showed an accuracy higher than 0.69, for all patients apart from two. The two DSC values for these two patients (0.12 and 0) are responsible of the difference between the average value (0.63 box cross in Fig. 5) and the median value (0.81, box line in Fig. 5). Second, the model was tested with the 19 3D-CT images of patients prospectively recruited in 9 different centres involved in the PET-Plan trial. $DSC(3D-CT)$ was 0.59 ± 0.24 . Overall, a trend of good detection but an underestimation of the size of the lesion was observed. This trend was confirmed by an average value of positive-predictive-value of 0.77 ± 0.23 . In Fig. 6 examples of contours for two patients are shown, being DSC and PPV values also provided.

Finally, the model was tested with the 80 4D-CT images of patients involved in the prospective PET-Plan cohort, the $DSC(4D-CT)$ was 0.61 ± 0.28 , improving 4% with respect to the accuracy for the expert, with $DSC(3D-Expert) = 0.59 \pm 0.24$. $DSC = 0$ for the expert data corresponding to a patient for which one of the experts considered image quality too poor in order to delineate the lesion. For this patient, an algorithm was able to detect the lesion, but segmentation accuracy was very poor ($DSC = 0.33$).

Discussion

In this study, we present a completely automatic algorithm to segment lung cancer tumors on PET images, with and without respiratory motion compensation. The performance of the algorithm has been evaluated with patients of ten different centres and by comparison with respect to a previously evaluated semiautomatic algorithm and with respect to the variability observed across different experts. In order to facilitate its clinical implementation, an automatic method for tumor segmentation on CT images, with and without motion compensation, based on the same nnU-Net has been additionally developed. The convolutional-network employed by training [38] and the resulting models ([GitHub_PETCTLungSegmentationModels_Link](#)) are open source.

One of the clinical applications for the resulting models would be their use for RT planning and monitoring in lung cancer. According to the ESTRO-ACROP guidelines for target volume (TV) definition in the treatment of locally advanced non-small cell lung cancer,

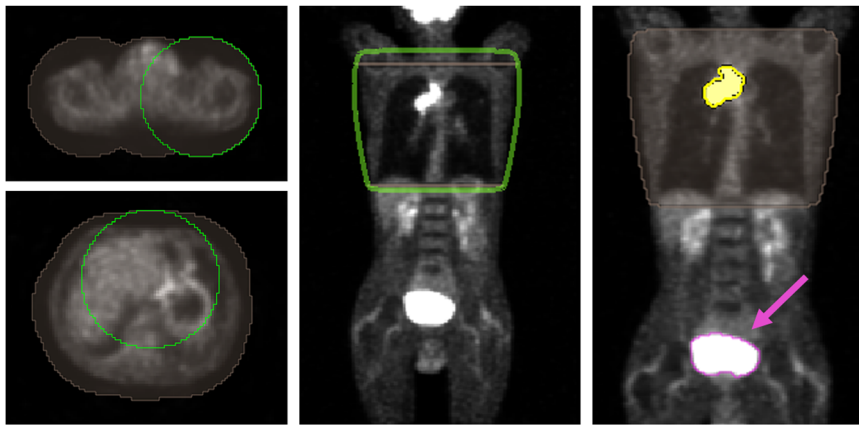


Fig. 4 Workflow for the lung mask creation: **left** regions of interest (brown) are manually delineated surrounding the body by a 50 voxel circle (green) in two slices located up and down the lung region; **middle** automatic interpolation is performed between the slices and lung region mask resulted (green) and **right** boolean intersection (yellow) between the AI segmentation and the lung region mask is obtained and consequently, the bladder region is rejected (segmentation pointed by the pink arrow)

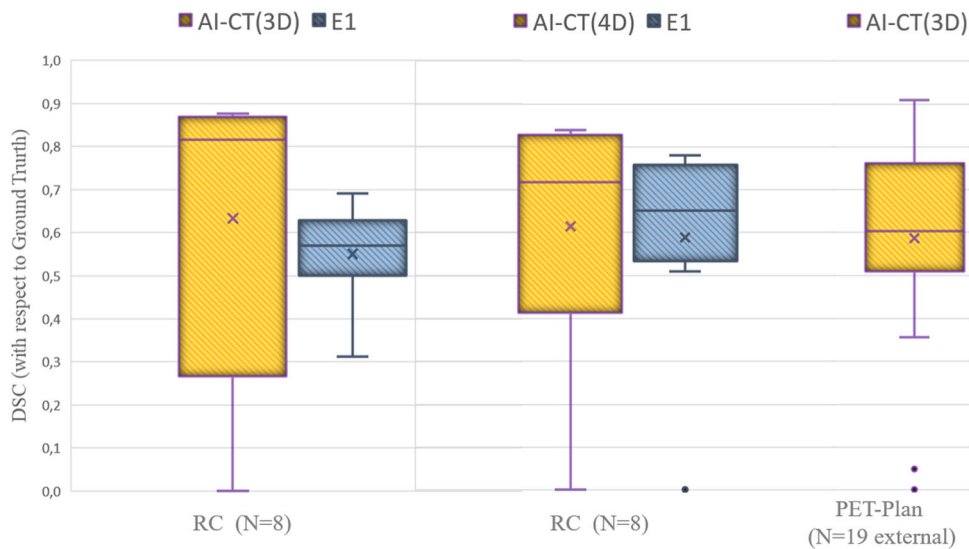


Fig. 5 DSC of the different CT segmentation approaches (artificial- intelligence-algorithm AI and expert) with respect to the ground truth defined by consensus of two radiation oncologists. Performance for the segmentation model on CT images with (4D) and without (3D) respiratory motion compensation are presented. The average value is represented by the cross and the median by the line

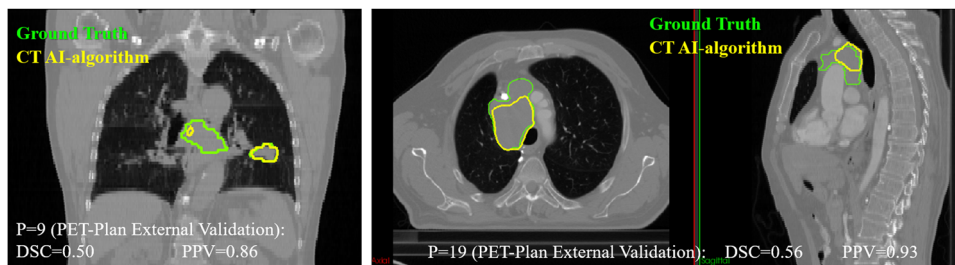


Fig. 6 Different CT segmentation approaches for two PET-Plan patients involved in the external validation

both a contrast-enhanced diagnostic CT scan and whole-body diagnostic FDG-PET-CT are considered mandatory in preparation for TV delineation for curative RT or chemo-RT [38]. Typically, the definition of the target volume is performed on the CT, adding the information derived from PET. Assessment of respiratory motion on a respiratory-correlated 4D CT scan is also recommended. 4D-gated PET/CT in the treatment position is optional but may improve the sensitivity of nodal identification and provide additional valuable information to help differentiate between tumor extent and adjacent tissues [39]. By applying our resulting models, the information from both images could be automatically integrated. Furthermore, with the advantage of the enhanced accuracy and robustness provided by 4D-protocols, these segmentations could be employed to improve the TV definition, optimising therefore treatment planning and monitoring.

Most of the previous convolutional-neural-network (CNN) approaches for lung tumor segmentation have been based on CT images [40–42] or on MR images [43] with 330, 1210, 19 and 9 patients, respectively. The benefit of PET imaging for lung cancer patients has been already demonstrated and consequently, segmentation models based on PET/CT images have been developed [44, 45], with 84 and 32 patients, respectively. It is known that the performance of AI segmentation models depends on the availability of a large amount of data for its training step. In PET, it is usually challenging to gather such large cohorts of patients compared to other imaging modalities [29]. In our study, the size of the PET cohort (100 patients) is larger than the size of the cohorts commonly employed for PET segmentation model development and validation [46]. It is also known that segmentation CNN-based approaches are used to perform well during training but demonstrate reduced performance during the validation and test steps. Therefore, cross-validation and testing of the model performance have been strongly recommended [29]. Our initial training cohort of 56 patients has been 5-fold cross-validated in order to take more benefit from the limited size of the sample. The performance of the resulting model after cross-validation, has been tested with 25 patients prospectively recruited at the same centre. Because the level of performance of AI models, as well as their reproducibility and robustness, is sensitive to imaging protocols and devices, we additionally evaluated the performance of our model for 19 patients prospectively recruited at 9 other centres. Segmentation accuracy in terms of DSC values was comparable for both, internal and external cohorts, allowing concluding an acceptable generalizability of the model. In terms of the clinical acceptability, it is important to include a comparison of the algorithm performance with respect to the manual contours and the variability across experts, which is

the standard method commonly employed in our clinical practice for tumor segmentation. In contrast to most of the previous publications, our study not only includes the comparison with respect to the established ground truth (consensus of two experts), but the analysis also involved the manual contours of three additional experts. From the results in this comparison, it could be concluded that the accuracy of the proposed model is promising in terms of robustness. In terms of the viability of its clinical implementation, a CT-based segmentation model has been developed with the same nnU-Net in order to facilitate the integration of PET segmentation model in the clinical workflow. Although the average values of DSC for the resulting CT model were not high, the performance of the algorithm was proved to be better than the variability observed among experts. Finally, previous literature has demonstrated that the compensation of respiratory motion improves accuracy and reproducibility for the biomarker distribution quantification [20, 23, 24]. However, 4D-PET imaging implementation in the clinical workflow is hindered by the time and expertise, required from the clinical site, for a proper analysis and integration of the functional information. The segmentation algorithm resulted from our study has been proven to accurately segment tumors with both imaging protocols, with (4D) and without (3D) motion compensation. To evaluate the impact of motion compensation on the performance of our model, images with and without motion compensation for the same patients should be compared. Unfortunately, only for 8 patients and only for CT images, both imaging protocols (3D and 4D) were available. For these patients, the better performance obtained for the model with non-compensated images ($DSC(3D-CT) = 0.63 \pm 0.34$, improving 15% improvement relative to experts) than with motion-compensated images ($DSC(4D-CT) = 0.61 \pm 0.28$, improving 4% relative to experts), could be justified by the fact that the model has been trained with non-compensated images. It is not the case for the manual contours, where the accuracy with motion-compensated images, $DSC(4D-Expert) = 0.59 \pm 0.24$, was better than without motion compensation, $DSC(3D-Expert) = 0.55 \pm 0.11$. However, the accuracy for the nnU-Net model is better than the variability observed across experts in both 3D and 4D images, supporting the applicability of the model independently of the imaging protocol. Therefore, this segmentation model could facilitate the clinical implementation of 4D-PET images for monitoring and for the development of prognosis models, where the quantification of PET images plays a crucial role in order to identify clinically relevant changes and where, consequently, the accuracy and reproducibility provided by these 4D-protocols are of special interest.

In order to quantify the performance of our models, 3D computation of DSC was employed, because it is the most

common and standardised method for the comparison of segmentations in medical imaging [38]. However, a high DSC does not necessarily indicate good agreement. An example could be a prediction of a tumor similar in size and location to the ground truth, but with a distant small region incorrectly classified also as a tumor. This scenario is likely to be found in lung tumor segmentation on 4D-PET thorax images, because of the high uptake of the heart and the noise implied by respiratory frames. For these reasons, Hausdorff distance [39] was additionally computed, and the results allowed us to confirm the good performance of our model. In addition, a positive-predictive-value [36] permitted to quantify the qualitatively observed trend in CT model performance, i.e., a high positive-predictive-value demonstrated that the model is able to detect the lesion and its location, but underestimates its size.

The main advantages of nnU-Net are that a preprocessing of the input data is not required, the non-fixed parameters for the training process of the model are automatically configured based on the input image data and the resulting model is generated and optimised without user interaction. In addition, the good performance of the nnU-Net method has been proved in different data sets and the code can be free to download [37]. For these reasons, the whole process can be easily reproduced. Medical data sharing is one of the most important concerns in AI development. It facilitates the comparison of the performance for different algorithms, can be used for the improvement of the generalizability of the algorithms, and makes feasible for developers to test their new CNNs. The main limitation in our study was that the clinical data employed is not open access. There are different CT and MR image repositories for lung cancer. The Lung Image Database Consortium-Image Database Resource Initiative (LIDC) is the world's largest publicly available database and contains 1018 CT scans of 1010 patients. The Automatic Nodule Detection 2009 (ANODE09) is another publicly available database consisting of 55 CT scans, annotated by two radiologists. Also, the TIME [47]; ELCAP [48] and LISS [49], are available in the public domain for lung nodule research. Future work should focus on PET image repositories that will also facilitate the development of PET AI tools for its use in clinical practice.

Conclusion

We presented two open-source nnU-Net models for lung tumor segmentation on PET and CT images, with and without respiratory motion compensation. Performance evaluation demonstrated that the automatic segmentation models have the potential to achieve accuracy comparable to manual segmentation and thus hold promise for clinical application. They could therefore be employed to facilitate the integration of FDG-PET/CT in lung cancer and to take advantage of the better quality provided by 4D quantification.

List of abbreviations

3D	Without motion compensation
4D	With motion compensation - respiratory gated
AI	Artificial Intelligence
CNN	Convolutional-neural-network
COA	Contrast-oriented-algorithm
CT	Computed tomography
EANM	European Association of Nuclear Medicine
FDG	[18F]fluoro-2-deoxy-D-glucose
(N)SCLC	Non-small cell lung cancer
PET	Positron emission tomography
RC	Retrospective cohort
RT	Radiation therapy
TV	Target volume

Author contributions

All authors contributed to the study's conception and design. Material preparation, data collection and analysis were performed by Montserrat Carles, Gianluca Radicioni and Dejan Kuhl. The first draft of the manuscript was written by Montserrat Carles and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding

Montserrat Carles was funded by the Conselleria de Sanitat Universal i Salut Pública of the Comunitat Valenciana through the project CDEI-01/20-B. Her 4 months stay at Medical Centre Freiburg in 2022 was funded by the call BEST-22 from Conselleria de Innovació, Universitats, Ciència i Societat Digital of the Comunitat Valenciana. In addition, this study has been funded by Instituto de Salud Carlos III (ISCIII) through the project PI22/01604 and co-funded by the European Union. The PET Plan trial was supported from the German Cancer Aid grant NR 108237. The STRIPE trial was supported from the German Cancer Aid grant NR Nr 108472. The funding sources had no involvement in the writing of the manuscript or in the decision to submit the article for publication. Open Access funding enabled and organised by Projekt DEAL.

Data availability

The resulted segmentation models can be downloaded ([GitHub_PETCTLungSegmentationModels_Link](#)).

Compliance with ethical standards

Guarantor

The scientific guarantor of this publication is Luis Martí Bonmatí.

Conflict of interest

The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

Statistics and biometry

One of the authors has significant statistical expertise.

Informed consent

Written informed consent was obtained from all subjects (patients) in this study.

Ethical approval

Institutional Review Board approval was obtained. This study was performed in line with the principles of the Declaration of Helsinki. The institutional ethics committee approved the study protocols (EK-Nr 21-1228-S1-retro, EK Nr 108237, EK 113/12, EK-Nr Nr. 108472).

Study subjects or cohorts overlap

Some study subjects or cohorts have been previously reported in previous publications but with clinical purposes instead of segmentation purposes.

Methodology

- Prospective
- Randomised controlled trial
- Multicentre study

Author details

¹La Fe Health Research Institute, Biomedical Imaging Research Group (GIBI230-PREBI) and Imaging La Fe node at Distributed Network for Biomedical Imaging (ReDIB) Unique Scientific and Technical Infra-structures (ICTS), Valencia, Spain. ²Department of Radiation Oncology, Division of Medical Physics, University Medical Center Freiburg, Faculty of Medicine, Freiburg, Germany. ³German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Partner Site Freiburg, German Cancer Research Center (DKFZ), Heidelberg, Germany. ⁴Department of Nuclear Medicine, Faculty of Medicine, University Medical Center Freiburg, Freiburg, Germany. ⁵Department of Radiation Oncology, Faculty of Medicine, University Medical Center Freiburg, Freiburg, Germany. ⁶Department of Radiation Oncology, Kliniken Maria Hilf GmbH Moenchengladbach, Moenchengladbach, Germany.

Received: 7 January 2024 Revised: 22 February 2024 Accepted: 28 March 2024

Published online: 25 April 2024

References

- Smith TA (1998) FDG uptake, tumour characteristics and response to therapy: a review. *Nuclear Med Commun* 19:97–105. <https://doi.org/10.1097/00006231-199802000-00002>
- Coleman RE (2002) Value of FDG-PET scanning in management of lung cancer. *Lancet* 359:1361–1362. [https://doi.org/10.1016/s0140-6736\(02\)08388-5](https://doi.org/10.1016/s0140-6736(02)08388-5)
- UyBico SJ, Wu CC, Suh RD, Le NH, Brown K, Krishnam MS (2010) Lung cancer staging essentials: the new tnM staging system and potential imaging pitfalls. *Radiographics* 30:1163–1181
- Pommier P, Touboul E, Chabaud S et al (2010) Impact of (18)F-FDG PET on treatment strategy and 3D radiotherapy planning in non-small cell lung cancer: a prospective multicenter study. *AJR Am J Roentgenol* 195:350–355
- Postmus PE, Kerr KM, Oudkerk M et al (2017) Early and locally advanced non-small-cell lung cancer (nscl): ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 28:iv1–iv21
- Manafi-Farid R, Askari E, Shiri I et al (2022) [(18)F]FDG-PET/CT radiomics and artificial intelligence in lung cancer: technical aspects and potential clinical applications. *Semin Nucl Med* 52:759–780. <https://doi.org/10.1053/j.semnuclmed.2022.04.004>
- Carles M, Fechter T, Radicioni G et al (2021) FDG-PET radiomics for response monitoring in non-small-cell lung cancer treated with radiation therapy. *Cancers (Basel)* 13:814. <https://doi.org/10.3390/cancers13040814>
- Kong F, Hu C, Machta M et al (2021) OA02.04 randomized phase II trial (RTOG1106) on midtreatment PET/CT guided adaptive radiotherapy in locally advanced non-small cell lung cancer. *J Thoracic Oncol* 16:S104–S105
- Vinod SK, Hau E (2020) Radiotherapy treatment for lung cancer: current status and future directions. *Respirology* 25:61–71
- Nestle U, Schimek-Jasch T, Kremp S et al (2020) Imaging-based target volume reduction in chemoradiotherapy for locally advanced non-small-cell lung cancer (PET-plan): a multicentre, open-label, randomised, controlled trial. *Lancet Oncol* 21:581–592
- Cooke SA, de Ruyscher D, Reymen B et al (2023) [(18)F]FDG-PET guided vs whole tumour radiotherapy dose escalation in patients with locally advanced non-small cell lung cancer (PET-Boost): results from a randomised clinical trial. *Radiother Oncol* 181:109492. <https://doi.org/10.1016/j.radonc.2023.109492>
- Gkika E, Grosu AL, Nestle U (2023) The use of 18 F-FDG PET/CT for radiotherapy treatment planning in non-small cell lung cancer: a mini-review. *Precision Cancer Med* 6:1–7.
- Bettinardi V, Picchio M, Di Muzio N, Gianolli L, Gilardi MC, Messa C (2010) Detection and compensation of organ/lesion motion using 4D-PET/CT respiratory gated acquisition techniques. *Radiother Oncol* 96:311–316. <https://doi.org/10.1016/j.radonc.2010.07.014>
- Büther F, Vehren T, Schäfers KP, Schäfers M (2016) Impact of data-driven respiratory gating in clinical PET. *Radiology* 281:229–238
- Huang TC, Chou KT, Wang YC, Zhang G (2014) Motion freeze for respiration motion correction in PET/CT: a preliminary investigation with lung cancer patient data. *BioMed Res Int* <https://doi.org/10.1155/2014/167491>
- Chang G, Chang T, Pan T, Clark JW, Mawlawi OR (2010) Joint correction of respiratory motion artifact and partial volume effect in lung/thoracic PET/CT imaging. *Med Phys* 37:6221–6232
- Dawood M, Büther F, Lang N, Schober O, Schäfers KP (2007) Respiratory gating in positron emission tomography: a quantitative comparison of different gating schemes. *Med Phys* 34:3067–3076
- Nehmeh SA, Erdi YE, Rosenzweig KE et al (2003) Reduction of respiratory motion artifacts in PET imaging of lung cancer by respiratory correlated dynamic PET: methodology and comparison with respiratory gated PET. *J Nucl Med* 44:1644–1648
- Park SJ, Ionascu D, Killoran JJ et al (2008) Evaluation of the combined effects of target size, respiratory motion and background activity on 3D and 4D PET/CT images. *Phys Med Biol* 53:3661–3679
- Chirindel A, Adebahr S, Schuster D et al (2015) Impact of 4D-18FDG-PET/CT imaging on target volume delineation in SBRT patients with central versus peripheral lung tumors. Multi-reader comparative study. *Radiother Oncol* 115:335–341
- Heath E, Unkelbach J, Oelfke U (2009) Incorporating uncertainties in respiratory motion into 4D treatment plan optimization. *Med Phys* 36:3059–3071. <https://doi.org/10.1118/1.3148582>
- Bittermann G, Scheifele C, Prokic V et al (2013) Description of a method: computer generated virtual model for accurate localisation of tumour margins, standardised resection, and planning of radiation treatment in head & neck cancer surgery. *J Craniomaxillofac Surg* 41:279–281. <https://doi.org/10.1016/j.jcms.2012.10.011>
- Carles M, Torres-Espallardo I, Alberich-Bayarri A et al (2016) Evaluation of PET texture features with heterogeneous phantoms: complementarity and effect of motion and segmentation method. *Phys Med Biol* 62:652–9251
- Carles M, Bach T, Torres-Espallardo I, Baltas D, Nestle U, Martí-Bonmatí L (2018) Significance of the impact of motion compensation on the variability of pet image features. *Phys Med Biol* 63:065013
- Zaidi H, El Naqa I (2010) PET-guided delineation of radiation therapy treatment volumes: a survey of image segmentation techniques. *Eur J Nucl Med Mol Imaging* 37:2165–2187. <https://doi.org/10.1007/s00259-010-1423-3>
- Morra J, Tu Z, Toga A, Thomson P (2010) Machine learning for brain image segmentation. *Biomedical image analysis and machine learning technologies: applications and techniques*, edited by Gonzalez FA and Romero E, IGI Global, 102–126. <https://doi.org/10.4018/978-1-60566-956-4.ch005>
- Kao D, Cheze Le Rest C, Jaouen V, Hatt M (2019) Artificial intelligence, machine (deep) learning and radio(geno)tics: definitions and nuclear medicine imaging applications. *Eur J Nucl Med Mol Imaging* 46:2630–2637. <https://doi.org/10.1007/s00259-019-04373-w>
- Litjens G, Kooi T, Bejnordi BE et al (2017) A survey on deep learning in medical image analysis. *Med Image Anal* 42:60–88
- Halder A, Dey D, Sadhu AK (2020) Lung nodule detection from feature engineering to deep learning in Thoracic CT images: a comprehensive review. *J Digit Imaging* 33:655–677. <https://doi.org/10.1007/s10278-020-00320-6>
- Liu X, Li KW, Yang R, Geng LS (2021) Review of deep learning based automatic segmentation for lung cancer radiotherapy. *Front Oncol* 11:717039. <https://doi.org/10.3389/fonc.2021.717039>
- Wang S, Mahon R, Weiss E et al (2022) Automated lung cancer segmentation using a PET and CT dual-modality deep learning neural network. *Int J Radiat Oncol Biol Phys* 115:529–539. <https://doi.org/10.1016/j.ijrobp.2022.07.2312>
- Xiang D, Zhang B, Lu Y, Deng S (2023) Modality-specific segmentation network for lung tumor segmentation in PET-CT images. *IEEE J Biomed Health Inform* 27:1237–1248. <https://doi.org/10.1109/JBHI.2022.3186275>

33. Nestle U, Adebahr S, Kaier K et al (2020) Quality of life after pulmonary stereotactic fractionated radiotherapy (SBRT): results of the phase II STRIFE trial. *Radiother Oncol* 148:82–88
34. Martinez-Movilla A, Mix M, Torres-Espallardo I et al (2022) Comparison of protocols with respiratory-gated (4D) motion compensation in PET/CT: open-source package for quantification of phantom image quality. *EJNMMI Phys* 9:80. <https://doi.org/10.1186/s40658-022-00509-4>
35. Nestle U, De Ruyscher D, Ricardi U et al (2018) ESTRO ACROP guidelines for target volume definition in the treatment of locally advanced non-small cell lung cancer. *Radiother Oncol* 127:1–5. <https://doi.org/10.1016/j.radonc.2018.02.023>
36. Carles M, Fechter T, Nemmer U et al (2015) Feasibility of a semi-automated contrast-oriented algorithm for tumor segmentation in retrospectively gated PET images: phantom and clinical validation. *Phys Med Biol* <https://doi.org/10.1088/0031-9155/60/24/9227>
37. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH (2021) nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods* 18:203–211. <https://doi.org/10.1038/s41592-020-01008-z>
38. Zou KH, Warfield SK, Bharatha A et al (2004) Statistical validation of image segmentation quality based on a spatial overlap index. *Acad Radiol* 11:178–189
39. Huttenlocher DP, Klanderman GA, Rucklidge WJ (1993) Comparing images using the Hausdorff distance. *IEEE Trans Pattern Anal Mach Intell* 15:850–863. <https://doi.org/10.1109/34.232073>
40. Zhang F, Wang Q, Li H (2020) Automatic segmentation of the gross target volume in non-small cell lung cancer using a modified version of ResNet. *Technol Cancer Res Treat* 19:1533033820947484. <https://doi.org/10.1177/1533033820947484>
41. Jiang J, Hu YC, Liu CJ et al (2019) Multiple resolution residually connected feature streams for automatic lung tumor segmentation from CT images. *IEEE Trans Med Imaging* 38:134–144. <https://doi.org/10.1109/TMI.2018.2857800>
42. Bi N, Wang J, Zhang T et al (2019) Deep learning improved clinical target volume contouring quality and efficiency for postoperative radiation therapy in non-small cell lung cancer. *Front Oncol* 9:1192. <https://doi.org/10.3389/fonc.2019.01192>
43. Wang C, Tyagi N, Rimmer A et al (2019) Segmenting lung tumors on longitudinal imaging studies via a patient-specific adaptive convolutional neural network. *Radiother Oncol* 131:101–107. <https://doi.org/10.1016/j.radonc.2018.10.037>
44. Zhao X, Li L, Lu W, Tan S (2018) Tumor co-segmentation in PET/CT using multi-modality fully convolutional neural network. *Phys Med Biol* 64:015011. <https://doi.org/10.1088/1361-6560/aaf44b>
45. Zhong Z, Kim Y, Zhou L et al (2018) 3D fully convolutional networks for co-segmentation of tumors on PET-CT images. *Proc IEEE Int Symp Biomed Imaging* 2018:228–231. <https://doi.org/10.1109/isbi.2018.8363561>
46. Kao YS, Yang J (2022) Deep learning-based auto-segmentation of lung tumor PET/CT scans: a systematic review. *Clin Transl Imaging* 10:217–223. <https://doi.org/10.1007/s40336-022-00482-z>
47. TIME. [Online] <http://cmp.felk.cvut.cz/projects/LungTIME/>, 2009. Accessed 14 Nov 2017
48. ELCAP. [Online] <http://www.via.cornell.edu/databases/lungdb.html/>, 2013. Accessed 25 Nov 2017
49. LISS. [Online] <http://isc.cs.bit.edu.cn/MLMR/LISS.html/>, 2015. Accessed 26 Nov 2017

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.