2Be3-Net : Combining 2D and 3D convolutional neural networks for 3D PET scans predictions

Ronan Thomas[1⋈][0000-0002-9908-2362], Elsa Schalck[1][0000-0001-7872-3215], Damien Fourure[2][0000-0001-5085-0052], Antoine Bonnefoy[1] and Inaki Cervera-Marzal[1][0000-0002-9341-2252]

[1] Eura Nova, Marseille, France
[2] Eura Nova, Mont-Saint-Guibert, Belgium
ronan.thomas@euranova.eu
inaki.cerveramarzal@euranova.eu

Abstract. Radiomics - high-dimensional features extracted from clinical images - is the main approach used to develop predictive models based on 3D Positron Emission Tomography (PET) scans of patients suffering from cancer. Radiomics extraction relies on an accurate segmentation of the tumoral region, which is a time consuming task subject to inter-observer variability. On the other hand, data driven approaches such as deep convolutional neural networks (CNN) struggle to achieve great performances on PET images due to the absence of available large PET datasets combined to the size of 3D networks. In this paper, we assemble several public datasets to create a PET dataset large of 2800 scans and propose a deep learning architecture named "2Be3-Net" associating a 2D feature extractor to a 3D CNN predictor. First, we take advantage of a 2D pre-trained model to extract feature maps out of 2D PET slices. Then we apply a 3D CNN on top of the concatenation of the previously extracted feature maps to compute patient-wise predictions. Experiments suggest that 2Be3-Net has an improved ability to exploit spatial information compared to 2D or 3D-only CNN solutions. We also evaluate our network on the prediction of clinical outcomes of head-and-neck cancer. The proposed pipeline outperforms PET radiomics approaches on the prediction of loco-regional recurrences and overall survival. Innovative deep learning architectures combining a pre-trained network with a 3D CNN could therefore be a great alternative to traditional CNN and radiomics approaches while empowering small and medium sized datasets.

Keywords: Deep learning architecture · 2D and 3D convolutional network · PET

1 Introduction

18 F-fluorodeoxyglucose (FDG) in positron emission tomography (PET) enables to highlight areas with high glucose metabolism, which is characteristic of tumor cells. PET is often associated with computerized tomography (CT) in a PET-CT exam, a hybrid imaging modality that allows to correlate metabolic and anatomic information

to improve lesion localisation and characterisation. PET-CT is a useful tool for diagnosis, prognosis, staging or re-staging of patients affected by cancer, and has been widely used in many studies [1-3].

Two main approaches are generally considered to exploit these 3D images. A first approach consists in the extraction of radiomics [4], defined as high-dimensional imaging features extracted from a segmented region of interest (ROI): the tumor. Radiomics features allow to quantitatively describe a tumor and can be divided into 4 groups: tumor shape, intensity, texture, and statistical features extracted after applying filters or mathematical transformations to the image. Radiomics can lead to the discovery of new quantitative bio-markers. Standardized Uptake Value (SUV) is a common metric describing tumor uptake normalized to the injected dose of FDG and patient's body weight. SUV is used in clinical routine and is a precious indicator to differentiate benign from malignant tumors and provides important prognostic and diagnostic information [5-8]. Several studies [9-11] illustrate the interest of using radiomics for applications such as prognosis, non-invasive disease tracking, treatment response or clinical outcome prediction tasks. Despite their good performances, radiomics robustness and replicability is questioned [12]. One of the most limiting points is linked to the difficulty of producing standardised images prior to radiomic extraction. In addition, tumor segmentation, a requirement for radiomics extraction, remains a complicated task. This step, either done manually or by a semi-automatic algorithm, introduces biases and raises several issues related to the experience reproducibility, its consistency and therefore hinders their deployment in clinical routine.

The second main approach consists in the use of convolutional neural networks that recently demonstrated great performances on vision tasks such as image classification, semantic segmentation or object detection. A major contribution to this success relies on the massive amount of training data with detailed and accurate annotations. Natural images models often rely on a transfer from large datasets such as the ImageNet dataset [13]. However, due to data sensitivity, it remains extremely challenging to build large datasets in the medical imaging domain. As a consequence, no large PET dataset has been made available so far.

Different CNN methods were applied on PET images. Because of the 3D nature of PET images, 3D CNN are logic architectures fitting PET scans dimensionality. However, the use of 3D convolutions implies an increased number of parameters and therefore requires large sets to be trained, where a training example corresponds to a scan. The limited size of PET datasets increases 3D CNN tendency to overfitting and degrades their performances. Studies applying 3D CNN on PET images therefore rely on consequent data augmentation to improve their model robustness [14-15]. Other solutions reformulated the problem in 2D enabling the exploitation of pre-trained models or the use of lighter 2D CNN [16-18]. However, using 2D models also implies

losing rich 3D spatial information, which results in sub-optimal performances. Some publications [19-21] illustrated that pre-training some of the network layers can help to accelerate training, convergence speed and increase the accuracy of the target model. Zhou [22] trained a shared 3D encoder associated with 8 decoding branches to segment different organs. Then they used the encoder as a backbone architecture to the classification of pulmonary nodules. Clark [23] trained several auto encoders (AE), each specific to an image modality (MRI, CT, X-ray). Training was done through image restoration which allowed the AEs to learn on unlabeled data specific information in the image such as appearance, texture or context. They transferred the learned AE on several tasks as brain tumor segmentation or nodule classification and illustrated an improvement in performances. They also demonstrated that their 3D models outperform their 2D versions, confirming the importance of 3D spatial information.

As illustrated by these recent works, traditional CNN approaches either 2D or 3D have both advantages and inconveniences. We decided to take advantage of their assets by combining a 2D pre-trained feature extractor with a predictive model such as a 3D CNN. Such a combination hasn't been studied much yet and could result in a more stable network than a complete 3D CNN. Moreover, we justify the use of a 2D pre-trained network by the information learned on millions of images, where an equivalent training from scratch on a PET dataset would not have been possible. The contributions of this work are summarized as follows:

- We introduce an innovative deep neural network entitled "2Be3-Net" combining a 2D pre-trained model to a 3D CNN

- We illustrated through predicting patient gender that the proposed pipeline integrates an increased ability to exploit spatial information compared to traditional CNN

- We evaluated the proposed architecture on the prediction of several clinical outcomes of the head-neck cancer and illustrated that it achieves superior performances on two out of the three tasks compared to PET radiomics

2 Method

We propose an architecture entitled "2Be3-Net", described in Fig. 1, that enables exploitation of raw 3D PET scans by associating a 2D feature extractor to a 3D CNN predictor. K randomly (but ordered) slices are sampled out of the 3D PET image to form a batch of k 2D input images. A 2D feature extractor is applied on each slice independently, resulting in k groups of 2D features maps. A concatenation layer is used

to create one group of 3D features maps that are fed to a 3D classifier to get the final prediction.

The feature extractor is a pre-trained 2D model that extracts feature maps out of 2D PET slices. We choose to use a ResNet-50 [24] pre-trained on the ImageNet dataset [13]. As the network is trained with 3 channels RGB images, we transform each PET slice into a gray scale image with 3 channels. Deep neural networks are known to learn hierarchical features, going from textural features in the network first layers to semantic features in the last layers. We believe that the textural features learned by an ImageNet pretrained network are useful for PET scan images but that the gap between PET scans and natural images is too high for the semantic features to be useful. Knowing that, we decided to keep only the 5 first layers of the pretrained network and, because of our small amount of data, decided to freeze these layers instead of finetuning them.

We concatenate the feature maps extracted from each slice and apply a predictive model, whose objective is to correlate the spatial and metabolic available information to compute patient-wise predictions. The predictive model is a typical framework of a CNN that applies three 3D convolutional blocks to reduce feature map size, followed by fully connected (FC) layers to realize the prediction.

PET scans have variable resolutions and numbers of slices. We considered these constraints as opportunities to do data augmentation. We randomly select a fixed number of slices and crop them. Random slice selection ensures that the network can't rely on specific slices in the scan, which further increases its robustness, while cropping decreases the size of the feature maps outputted by the feature extractor.
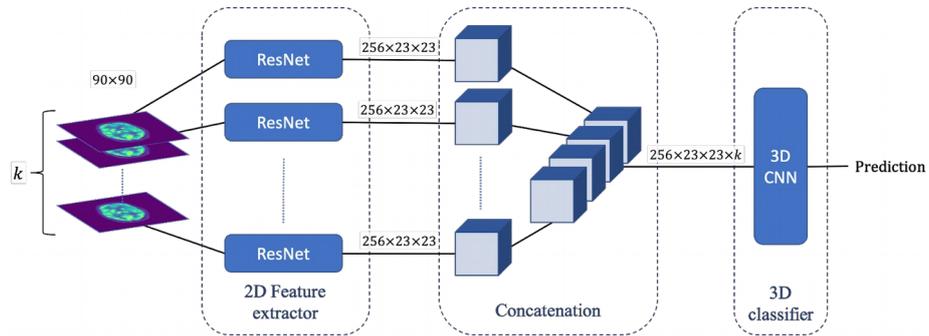
Fig 1. 2Be3-Net global pipeline

3 Experiments

3.1 Experimental Setup

Studies applying CNN models on medical images often train and evaluate their models on small datasets. In this work, we collected and assembled 10 public datasets available in The Cancer Imaging Archive [25-35]. These datasets contain images from different modalities, pathologies and centers, which implies a large variety of data characteristics, spatial resolution, range of pixel intensities and acquisition protocols. We selected PET with attenuated correction (AC) scans and normalized raw pixel values to SUV scale using the definition[1] provided by the Quantitative Imaging Biomarkers Alliance (QIBA). Scans containing SUV outliers were dropped before conversion to NIFTI format, resulting in a PET dataset containing 2834 scans. We split this dataset in two to create two sub-datasets for specific tasks: gender and clinical outcomes prediction (loco-regional recurrences (LR), distant metastases (DM) and overall survival (OS)). The clinical outcomes dataset is included in the gender dataset, but no overlapping exists between the validation set of the clinical outcomes dataset and the gender training set.

The number of input slices is set to k=66 as it corresponds to the number of slices of the smallest scan. We also chose a resolution of 90x90 pixels per slice. To match this resolution, we crop the slices at their center. Finally, we apply random flip and rotation (-10°, +10°) as data augmentation. Models were implemented in Python 3.7 and Pytorch 1.6 [36]. Experiments were conducted on a Ubuntu 18.04 system equipped with a Nvidia GeForce GTX 1070 with 8 Gb GPU memory and CUDA 11.0.

3.2 Experiment 1: Ability to Exploit Spatial Information

We compare 2Be3-Net capacity to exploit spatial information with a 2D and a 3D CNN. The 2D model is the 2D version of 2Be3-Net, where we replace the 3D blocks by their 2D alternative. The feature maps outputted by the ResNet-50 are provided as input to the 2D blocks and are flattened before applying FC layers. The full 3D CNN is composed of four 3D convolutional blocks similar to the ones of 2Be3-Net, followed by 2 FC layers. Batch size is set to 6, due to limited GPU memory.

In order to evaluate their capacity to exploit spatial information, we predict patient gender based on transverse slices. As transverse slices don't provide a whole body visualization, models should exploit the spatial information contained in the feature maps to compute predictions. The gender dataset used in this experiment contains 2459 scans in the training set and 377 in the validation set. The dataset is also imbal-

---

[1] https://qibawiki.rsna.org/index.php/Standardized_Uptake_Value_(SUV)

anced, as 77.2% of patients are men. We address this issue by using a weighted binary cross-entropy loss associated with Adam optimizer with an initial learning rate of 1e-5 and a weight decay of 1e-2. We evaluate models' performances with the area under the curve (AUC) of receiver characteristic operator (ROC) associated with sensitivity (SENS), specificity (SPEC) and accuracy (ACC). Experimental results are displayed in Table 1. The experiment shows that 2Be3-Net achieved a better result than the 2D version, showing the importance of taking into account the 3D spatial information while the full 3D CNN struggles in this task.

Table 1. Results of experiment 1

Metric	AUC	ACC	Sensitivity	Specificity
2Be3-Net	0.94	0.92	0.91	0.97
ResNet+Conv 2d	0.92	0.91	0.90	0.93
3D CNN	0.7	0.75	0.79	0.61

3.3 Experiment 2: Prediction of Clinical Outcomes

We compare the CNN models previously described to the radiomics approach developed by Vallières [37] on the prediction of clinical outcomes of head-and-neck cancer. To ease comparisons, we compare our results with 2 of their models. The PET radiomics model applies a logistic regression on different variables specific to each outcome. The best radiomics model uses random forests to combine PET and/or CT radiomics to clinical data. On the other side, we decline the 2Be3-Net in two versions. The first version entitled 2Be3-Net-[WS] takes as input 66 slices randomly selected in the whole scan. The second version, named 2Be3-Net-[H&N], takes as input 66 slices extracted from the head-neck region, following the intuition that this area is more inclined to contain relevant information for these tasks. As a pre-training, we use the weights of the models trained on gender prediction to initialize CNN models weights.

The clinical outcomes dataset used follows the dataset [37], at the difference of 9 patients (5 in the training set and 4 in the validation set), excluded due to image errors, initial data curation error (detected by TCIA) or missing information to calculate SUV. The resulting dataset contains 187 scans in the training set and 102 in the validation set. This dataset presents a pronounced class imbalance (LR: 14.6%, DM: 13.6% and OS: 18.4%). We apply the same strategy as experiment 1 to address this issue. Same data augmentation as before is applied to improve network robustness. Experiment results displayed in Table 2 show that both versions of 2Be3-Net outperforms the PET radiomics for LR and OS predictions, but achieves inferior performances in DM prediction.

Table 2. Clinical outcomes prediction results

	Metric	ResNet+ Conv2D	3D CNN	2Be3-Net [WS]	2Be3-Net [H&N]	PET radiomics [41]	Best radiomics model [41]
LR	ACC	0.58	0.83	0.6	0.73	0.67	0.67
	AUC	0.64	0.69	0.68	0.72	0.53	0.69
	SENS	0.71	0.5	0.79	0.71	0.38	0.63
	SPEC	0.56	0.89	0.57	0.73	0.7	0.68
DM	ACC	0.70	0.71	0.68	0.76	0.68	0.77
	AUC	0.67	0.71	0.78	0.71	0.8	0.86
	SENS	0.64	0.71	0.93	0.64	0.85	0.86
	SPEC	0.71	0.71	0.64	0.77	0.66	0.76
OS	ACC	0.75	0.68	0.71	0.71	0.64	0.62
	AUC	0.72	0.65	0.76	0.74	0.62	0.74
	SENS	0.68	0.59	0.86	0.79	0.58	0.79
	SPEC	0.76	0.7	0.66	0.69	0.66	0.57

4 Discussion

We first evaluated 2Be3-Net capacity to exploit spatial information through predicting patient gender based on transverse PET slices. The proposed pipeline achieves a 0.94 AUC, achieving the best score on this task. We attribute this performance to the 3D convolutional blocks that exploited the spatial information contained in the concatenated feature maps. In the ResNet+Conv 2D model, the 3D convolutional blocks were replaced by 2D convolutional blocks, which prevented the exploitation of feature maps spatial information. By contrast, we attribute the full 3D CNN poor results to it's deep architecture and it's training from scratch, where more training samples would have been required to improve its results.

We studied the gender predictions made by our CNN and found that 70% of mispredicted scans in the validation set have less than 180 slices. These scans only cover the body upper region and represent 25.2% of the total dataset size. As the whole body cannot be visualised on these scans, it is difficult for models to identify the gender and therefore are more prone to mispredictions.

We predicted different clinical outcomes of head-neck cancer (LR, DM, OS), and compared the deep learning models to a radiomics approach. CNN models took advantage of a pre-training on gender prediction, which improved their performances compared to training from scratch. In this experiment, all deep learning models achieved better results than PET radiomics on LR and OS predictions. We also note that both versions of 2Be3-Net achieved at least equivalent AUC and improved sensitivity compared to the best radiomics model. These results seem promising as the best

radiomics model combined information from PET-CT scans associated with clinical data while our models only had access to the PET scan. The improved sensitivity implies that the proposed 2Be3-Net correctly identified more examples of the minority class than the radiomics models. It is also important to note that specificity wasn't compromised while sensitivity increased. We also note that 2Be3-Net-[H&N] presents results more stable compared to 2Be3-Net-[WS] as the gap between sensitivity and specificity decreased. This difference can be attributed to the slice selection area, where 2Be3-Net-[H&N] used slices selected in the head-neck region, and was therefore able to focus on tumor related information. However, CNN models achieved inferior results on DM prediction compared to the radiomics models. The radiomics models were specific to each outcome, and reached their best score on DM prediction. Thus, we attribute radiomics superior results on DM prediction to the specific design of the radiomics DM model. On the other hand, both 2Be3-Net versions were designed to predict all clinical outcomes and achieved stable results on those.

In the light of these experimental results, alternative deep learning architectures seem promising alternatives to radiomics and CNN approaches. The proposed 2Be3-Net accepts as input 3D PET scans with SUV conversion as the only preprocessing step, and is able to predict clinical outcomes of head-neck cancer from a training done on a small size dataset.

5 Conclusion

This paper introduces 2Be3-Net, a new architecture allowing direct exploitation of 3D PET scans through the association of a 2D pre-trained network with a 3D CNN, which enables exploitation of spatial information between the feature maps extracted. We compared 2Be3-Net to a traditional 2D CNN, a 3D CNN and a radiomics approach on the prediction of clinical outcomes of head-neck cancer. Experiments illustrated that the proposed architecture is a good alternative to classic CNN models and radiomics approaches. Moreover, it accepts as input entire PET scans requiring few preprocessing steps.

We used a ResNet-50 pre-trained on natural images as a feature extractor. Future works could focus on using a model pre-trained on PET images. In this matter, auto-encoder architectures seem promising as they can be trained in an unsupervised manner, which solves the requirement of large annotated datasets and could further improve the relevance and quality of the feature maps extracted.

References

1. Krause, B. J. et al. FDG PET and PET/CT. Recent results in cancer research, 187, 351–369 (2013). doi: 10.1007/978-3-642-10853-2_12

2. Gallamini, A. et al. FDG-PET Scan: a new Paradigm for Follicular Lymphoma Management. Mediterranean journal of hematology and infectious diseases, 9(1), e2017029 (2017). doi: 10.4084/MJHID.2017.029

3. Fletcher, J. W. et al. Recommendations on the use of 18F-FDG PET in oncology. Journal of nuclear medicine : official publication, Society of Nuclear Medicine, 49(3), 480–508 (2008). doi: 10.2967/jnumed.107.047787

4. Mayerhoefer, M. E. et al. Introduction to Radiomics. Journal of nuclear medicine : official publication, Society of Nuclear Medicine, 61(4), 488–495 (2020). doi: 10.2967/jnumed.118.222893

5. Bastiaannet, E. et al. The value of FDG-PET in the detection, grading and response to therapy of soft tissue and bone sarcomas; a systematic review and meta-analysis. Cancer treatment reviews, 30(1), 83–101 (2004). doi: 10.1016/j.ctrv.2003.07.004

6. Slot, K.M. et al. Prediction of Meningioma WHO Grade Using PET Findings: A Systematic Review and Meta Analysis. Journal of Neuroimaging (2020). doi: 10.1111/jon.12795

7. Bailly, C. et al. Interest of FDG-PET in the Management of Mantle Cell Lymphoma. Frontiers in medicine, 6, 70 (2019). doi: 10.3389/fmed.2019.00070

8. Xie, P. et al. 18F-FDG PET or PET-CT to evaluate prognosis for head and neck cancer: a meta-analysis. J Cancer Res Clin Oncol 137, 1085–1093 (2011). doi: 10.1007/s00432-010-0972-y

9. Ha, S. et al. Radiomics in Oncological PET/CT: a Methodological Overview. Nucl Med Mol Imaging 53, 14–29 (2019). doi: 10.1007/s13139-019-00571-4

10. Aerts, H. J. et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nature communications, 5, 4006 (2014). doi: 10.1038/ncomms5006

11. Lucia, F. et al. Prediction of outcome using pretreatment 18F-FDG PET/CT and MRI radiomics in locally advanced cervical cancer treated with chemoradiotherapy. Eur J Nucl Med Mol Imaging 45, 768–786 (2018). doi: 10.1007/s00259-017-3898-7

12. Bogowicz, M. et al. CT radiomics and PET radiomics: ready for clinical implementation?. The quarterly journal of nuclear medicine and molecular imaging : official publication of the Italian Association of Nuclear Medicine (AIMN) [and] the International Association of Radiopharmacology (IAR), [and] Section of the Society of..., 63(4), 355–370 (2019). doi: 10.23736/S1824-4785.19.03192-3

13. Russakovsky, O. et al. ImageNet Large Scale Visual Recognition Challenge. Int J Comput Vis 115, 211–252 (2015). doi: 10.1007/s11263-015-0816-y

14. Islam, J. et al. Understanding 3D CNN Behavior for Alzheimer's Disease Diagnosis from Brain PET Scan, (2019). arXiv:1912.04563

15. Khvostikov, A. et al. 3D CNN-based classification using sMRI and MD-DTI images for Alzheimer disease studies (2018)

16. Han, X et al. MR based synthetic CT generation using a deep convolutional neural network method. Med. Phys., 44: 1408-1419 (2017). doi: 10.1002/mp.12155

17. Kawauchi, K. et al. A convolutional neural network-based system to prevent patient misidentification in FDG-PET examinations. Scientific reports, 9(1), 7192 (2019). doi: 10.1038/s41598-019-43656-y

18. Ding, Y. et al. A Deep Learning Model to Predict a Diagnosis of Alzheimer Disease by Using 18F-FDG PET of the Brain. Radiology, 290(2), 456–464 (2019). doi: 10.1148/radiol.2018180958

19. Ravishankar H. et al. Understanding the Mechanisms of Deep Transfer Learning for Medical Images. In: Carneiro G. et al. (eds) Deep Learning and Data Labeling for Medical Applications. DLMIA 2016, LABELS 2016. Lecture Notes in Computer Science, vol 10008. Springer, Cham (2016). doi: 10.1007/978-3-319-46976-8_20

20. N. Tajbakhsh et al. "Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?," in IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1299-1312 (2016). doi: 10.1109/TMI.2016.2535302

21. Erhan, D. et al. The Difficulty of Training Deep Architectures and the Effect of Unsupervised Pre-Training. Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics, in PMLR 5:153-160 (2009).

22. Zhou Z. et al. Models Genesis: Generic Autodidactic Models for 3D Medical Image Analysis. In: Shen D. et al. (eds) Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. MICCAI 2019. Lecture Notes in Computer Science, vol 11767. Springer, Cham (2019) doi: 10.1007/978-3-030-32251-9_42

23. Chen, S. et al. Med3D: Transfer Learning for 3D Medical Image Analysis (2019).

24. Kaiming, H. et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition (2016).

25. Clark, K. et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. Journal of digital imaging, 26(6), 1045–1057 (2013). doi: 10.1007/s10278-013-9622-7

26. Vallières, M. et al. Data from Head-Neck-PET-CT. The Cancer Imaging Archive. (2017). doi: 10.7937/K9/TCIA.2017.8oje5q00

27. Walter R.B. et al. Data From Head-Neck_Cetuximab. The Cancer Imaging Archive. (2015). doi: 10.7937/K9/TCIA.2015.7AKGJUPZ

28. Zuley, M. L., et al. Radiology Data from The Cancer Genome Atlas Head-Neck Squamous Cell Carcinoma [TCGA-HNSC] collection. The Cancer Imaging Archive. (2016). doi: 10.7937/K9/TCIA.2016.LXKQ47MS

29. Grossberg, A. et al. M.D. Anderson Cancer Center Head and Neck Quantitative Imaging Working Group. HNSCC [Dataset]. The Cancer Imaging Archive. (2020). doi: 10.7937/k9/tcia.2020.a8sh-7363

30. Beichel, R R. et al. Data From QIN-HEADNECK. The Cancer Imaging Archive. (2015). doi: 10.7937/K9/TCIA.2015.K0F5CGLI

31. Kinahan, P. et al. Data from the ACRIN 6668 Trial NSCLC-FDG-PET [Dataset]. The Cancer Imaging Archive. (2019). doi: 10.7937/tcia.2019.30ilqfcl

32. Bakr, S. et al. Data for NSCLC Radiogenomics Collection. The Cancer Imaging Archive. (2017). doi: 10.7937/K9/TCIA.2017.7hs46erv

33. Albertina, B. et al. Radiology Data from The Cancer Genome Atlas Lung Adenocarcinoma [TCGA-LUAD] collection. The Cancer Imaging Archive. (2016). doi: 10.7937/K9/TCIA.2016.JGNIHEP5

34. Kirk, S. et al. Radiology Data from The Cancer Genome Atlas Lung Squamous Cell Carcinoma [TCGA-LUSC] collection. The Cancer Imaging Archive. (2016). doi: 10.7937/K9/TCIA.2016.TYGKKFMQ

35. Muzi P. et al. Data From RIDER Lung PET-CT. The Cancer Imaging Archive. (2015). doi: 10.7937/K9/TCIA.2015.OFIP7TVM

36. Paszke, A. et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library, Advances in Neural Information Processing Systems 32, 8026-8037, Curran Associates, Inc., 1912.01703 (2019).

37. Vallières, M. et al. Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer. Scientific reports, 7(1), 10117 (2017). doi: 10.1038/s41598-017-10371-5