Convolutional Neural Networks Detect Local Infiltration of Lung Cancer Primary Lesions on Baseline FDG-PET/CT

Margarita Kirienko[1], **Martina Sollini**[1], **Giorgia Silvestri**[2], **Serena Mognetti**[2],
Emanuele Voulaz[3], **Lidija Antunovic**[4], **Alexia Rossi**[1,5], **Luca Antiga**[2] and **Arturo Chiti**[1,4]

[1] Department of Biomedical Sciences, Humanitas University, Pieve Emanuele (Milan), Italy,
[2] Orobix Srl, Bergamo, Italy,
[3] Thoracic Surgery, Humanitas Clinical and Research Hospital, Rozzano (Milan), Italy,
[4] Nuclear Medicine, Humanitas Clinical and Research Hospital, Rozzano (Milan), Italy,
[5] Radiology, Humanitas Clinical and Research Hospital, Rozzano (Milan), Italy,
margarita.kirienko@st.hunimed.eu, giorgia.silvestri@orobix.com

Abstract

The aim of our work is to develop an algorithm for the classification of lung cancer as T1–T2 or T3–T4 on staging fluorodeoxyglucose positron emission tomography (FDG-PET)/CT images. We retrospectively selected a cohort of 632 patients from patients submitted to FDG-PET/CT, over a 6-year period, for the purpose of staging a suspected lung lesion, within 60 days before biopsy or surgical procedure. Post-acquisition processing was performed to generate an adequate dataset for the convolution neural network (CNN) analyses. The input of CNNs in this study was a bounding box on both PET and CT images, cropped around the lesion centre identified by two nuclear medicine physicians. The algorithm developed and tested in the present work achieved an accuracy of 83.9%, 76.2% and 69.8% in the training, validation and test set, respectively, for the identification of T1–T2 and T3–T4 lung cancer. We obtained proof of concept that CNNs can be used as a tool to assist in the staging of patients affected by NSCLC. Further research in this field is already being addressed by our group.

1 Introduction

In recent years, advanced analysis of medical imaging using radiomics and machine and deep learning, including approaches using convolutional neural networks (CNNs), has been explored. These approaches offer great promise for future applications for both diagnostic and predictive purposes. CNNs are non-explicitly programmed algorithms that identify relevant features on the images that allow them to classify an input object. They have been applied in various tasks such as detection (e.g. breast lesions on mammographic scans), segmentation (e.g. liver and liver lesions on computed tomography CT) and diagnosis (e.g. lung lesions on screening low-dose CT) [1]. In the field of lung imaging, CNNs have been tested in nodule segmentation from CT images. Average dice scores of 82% and 80% for the training and test datasets, respectively, have been reported [2]. Moreover, a model for assessment of cancer probability in patients with pulmonary nodules has been proposed. The area under the curve (AUC) was found to be 0.90 and 0.87 for the training and test sets, respectively [3]. Stage assessment has not yet been described. The present study, as a first step towards complete TNM parameter assessment, aimed to develop an algorithm for the classification of lung cancer as T1–T2 or T3–T4 on staging fluorodeoxyglucose positron emission tomography (FDG-PET)/CT images.
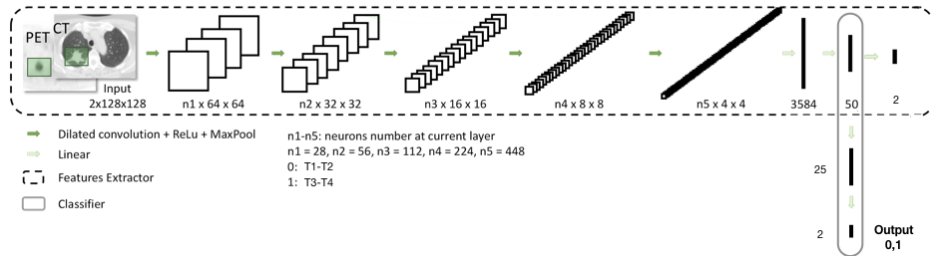
Figure 1: CNN architecture

2 Methods & Experiments

In this retrospective single-centre investigation, we screened all patients who underwent FDG-PET/CT for the purpose of staging a suspected lung lesion, within 60 days before biopsy or surgical procedure. The inclusion criteria were: (a) age >18 years and (b) histological diagnosis of primary lung cancer. The exclusion criteria were: (a) inconclusive histology due to inadequate biopsy sample, (b) diagnosis of non-malignancy. From the institutional database, a cohort of 632 patients (T1-T2 = 432, T3-T4 = 200) was selected applying the above-mentioned criteria. FDG-PET/CT was performed according to standard institutional procedures, previously detailed [4].

The original CT and PET image size is $512 \times 512 \times N_{slices}$ and $128 \times 128 \times N_{slices}$ respectively, where N_{slices} is the number of slices in which the lesion appears. The CT images were clipped between -1000 and 400 Hounsfield units, PET images were resampled in the CT space and both images were re-scaled to lie between 0 and 1. After that, this study's dataset consists of 3D bounding boxes on both PET and CT images, performing crops around the lesion's centre, identified by two nuclear medicine physicians (M.S. and M.K.) with dimension $128 \times 128 \times N_{slices}$.

The algorithm is composed of two networks: a *feature extractor* and a *classifier*. The *feature extractor* is a CNN that takes an CT-PET image patch of 128×128 pixels as input and performs classification (T1–T2 with $label = 0$ and T3–T4 with $label = 1$) according to the appearance of the image patch. The aim of this network is to extract the most relevant features from a single patch. The *classifier* takes as input the mean of the second-last layer of features extracted from all slices of a single patient and aims to perform a classification (T1–T2 vs T3–T4) for it. The *softmax function* is applied to the last layer of both networks, in order obtain the probability of being T1–T2 and T3–T4. The class having the highest probability is assigned to each patient. The architecture for both of the models is represents in Figure 1. We trained both models with the Adam algorithm [5]. We used an initial learning rate of $1e^{-6}$ for the *feature extractor* and $1e^{-3}$ for the *classifier* and applied the *cross entropy loss function* for both cases. For what concerns data augmentation during the model training, image patches were rotated in 2D space around the lesion centre on the z axis by a randomly selected angle in the range $[-10°, 10°]$. All code is written in Python, using the PyTorch deep learning library (http://pytorch.org/).

3 Results & Conclusions

All results are summarized in Table 1 and Figure 2 shows examples of patients classified by CNNs as T1–T2 and T3–T4. The algorithm developed and tested in the present work achieved an accuracy of 83.9%, 76.2% and 69.8% in the training, validation and test sets, respectively, for the identification of T1–T2 and T3–T4 lung cancer.

Some limitations of this study have to be acknowledged. Firstly, the limited number of patients precluded design of a network for a classification task with four outputs (T1, T2, T3 and T4). Moreover, better performance is expected in larger datasets. In future work, we plan to include a larger number of cases in a multicentre framework. Secondly, the reference to define the T parameter was based on the pathological assessment in surgically treated patients (62% of cases), while in patients not suitable for surgery, clinical assessment was used as the standard. A homogeneous clinical assessment will be considered for future studies. In conclusion, the key result in the present preliminary investigation is the feasibility and promising performance of CNNs in assessing the

Table 1: Results

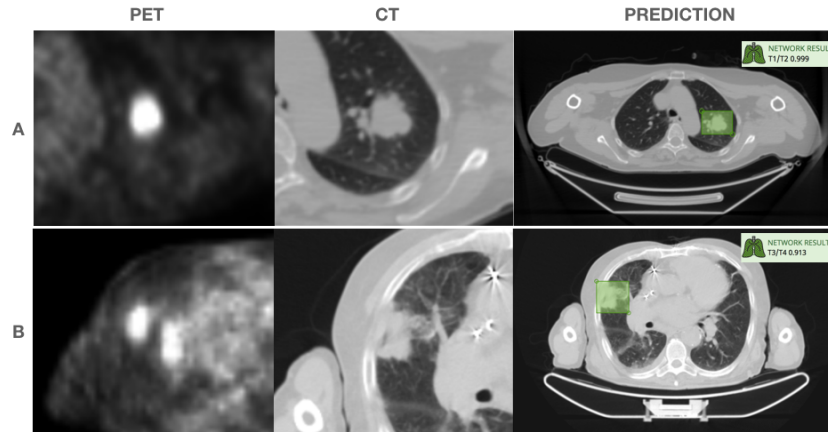|	Accuracy (%)	Recall (%)	Specificity (%)
Train (n = 405)	83.9	78.9	86.3
Validation (n = 101)	76.2	62.5	82.6
Test (n = 126)	69.8	62.5	73.3

Figure 2: (A) Axial PET-CT images of a patient affected by lung adenocarcinoma (T1-T2) and its prediction as belonging to the class T1-T2 with 99% probability. (B) Axial PET-CT images of a patient affected by lung adenocarcinoma (T3-T4) and its prediction as belonging to the class T3-T4 with 91% probability.

T parameter in lung cancer. Further investigations are needed to develop robust algorithms for complete TNM assessment. Compared with radiomics, CNNs have the advantage of eliminating the need for tumour segmentation, feature calculation and selection, which are even more critical issues in small lesions. Still, the possibility of a complementary role of radiomics and artificial intelligence techniques should be addressed. Moreover, improvement in risk stratification is foreseen with the incorporation of patients' clinical features in the neural network algorithm. The development of an automated system based on neural networks that is able to perform a complete stage and risk assessment will ensure more accurate and reproducible use of imaging data and allow deeper integration of this information in the therapeutic plan for each individual patient.

References

[1] Gabriel Chartrand, Phillip M Cheng, Eugene Vorontsov, Michal Drozdzal, Simon Turcotte, Christopher J Pal, Samuel Kadoury, and An Tang. Deep learning: A primer for radiologists. *RadioGraphics*, 37(7):2113–2131, 2017.

[2] Shuo Wang, Mu Zhou, Zaiyi Liu, Zhenyu Liu, Dongsheng Gu, Yali Zang, Di Dong, Olivier Gevaert, and Jie Tian. Central focused convolutional neural networks: Developing a data-driven model for lung nodule segmentation. *Medical image analysis*, 40:172–183, 2017.

[3] Fangzhou Liao, Ming Liang, Zhe Li, Xiaolin Hu, and Sen Song. Evaluate the malignancy of pulmonary nodules using the 3d deep leaky noisy-or network. *arXiv preprint arXiv:1711.08324*, 2017.

[4] Margarita Kirienko, Luca Cozzi, Lidija Antunovic, Lisa Lozza, Antonella Fogliata, Emanuele Voulaz, Alexia Rossi, Arturo Chiti, and Martina Sollini. Prediction of disease-free survival by the pet/ct radiomic signature in non-small cell lung cancer patients undergoing surgery. *European journal of nuclear medicine and molecular imaging*, 45(2):207–217, 2018.

[5] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014.