IL NUOVO CIMENTO **46 C** (2023) 134
DOI 10.1393/ncc/i2023-23134-6

COMMUNICATIONS: SIF Congress 2022

Study of the spatiotemporal learning of dynamic positron emission tomography data for the improvement of diagnostic accuracy in breast cancer(*)

M. INGLESE, M. FERRANTE, T. BOCCATO, A. DUGGENTO and N. TOSCHI
Department of Biomedicine and Prevention, Universitá di Roma Tor Vergata - Roma, Italy

received 10 March 2023

Summary. — This article is based on the talk delivered on 12 September 2022, in the Medical Physics section of the 108th National Congress of the Italian Physical Society (Milan). This work addresses the challenge of improving PET diagnostic accuracy through an alternative approach based on the analysis of time signal intensity patterns extracted from dynamic PET tissue time activity curves with a deep learning model. Our framework outperforms the discriminative potential of the classical SUV analysis, paving the way for more accurate dynamic PET-based lesion discrimination without additional acquisition time or invasive procedures. The full study has been published in *IEEE Trans. Radiat. Plasma Med. Sci.*, **7** (2023) 630, with the title *Spatiotemporal learning of dynamic positron emission tomography data improves diagnostic accuracy in breast cancer.*

1. – Introduction

Positron emission tomography (PET) allows the quantification of the biochemical properties of tissue under investigation through the injection and detection of a targeted radiotracer [1]. A PET image is an *in vivo* map of the spatiotemporal tracer concentration that includes details on the delivery of the tracer to tissue, how it interacts with the target, and the washout effects, which can be inferred from the shape of the tissue time activity curves (TACs) extracted voxelwise from the dynamic PET image [2,3]. However, this type of quantitative analysis often requires arterial cannulation and blood sample collection throughout the whole PET acquisition [4-6]. In clinical practice, PET data is acquired following a static acquisition protocol and the standardized uptake value (SUV) is the most widely used PET-derived (semi-)quantitative information in clinical and research applications [7]. This standardization assumes the amount of the non-metabolized tracer in the region of interest (ROI) to be negligible and that the time integral of plasma tracer concentration is proportional to the amount of tracer injected [8]. However, these presumptions frequently lead to non-negligible errors in the full quantification of tracer

(*) This article is a short overview of INGLESE M., FERRANTE M., DUGGENTO A., BOCCATO T. and TOSCHI N., *IEEE Trans. Radiat. Plasma Med. Sci.*, **7** (2023) 630, https://doi.org/10.1109/TRPMS.2023.3268361.

Creative Commons Attribution 4.0 License (https://creativecommons.org/licenses/by/4.0)

kinetics [8]. Dynamic PET may be able to lessen the large time dependence seen in SUV quantification of normal tissue and tumor uptake values, allowing greater flexibility and reliability in clinical practice [9]. The main objective of this paper is to compare the information content, and therefore the discrimination performance, embedded in the time domain of dynamic PET acquisitions compared to a traditional, static dataset. To this end, we combined several machine- and deep learning architectures on clinical data obtained from a cohort of breast cancer patients who received dynamic 3-deoxy-3-18F-fluorothymidine (18F-FLT) PET scans for a relatively straightforward task (classification between lesion and reference tissue) in order to highlight any potential static-vs dynamic effect.

2. – Materials and methods

2˙1. *The Dataset*. – We employed a publicly available clinical 18F-FLT PET dataset of 44 breast cancer patients, part of the "ACRIN-FLT-Breast (ACRIN 6688)" collection in the Cancer Imaging Archive (TCIA) [10-12]. For the purpose of this study, only baseline dynamic 18F-FLT PET scans were employed.

2˙2. *Data processing*. – For each patient, consecutive regions of interest were manually contoured around the tumor and the contralateral healthy breast by an experienced radiologist on the static PET image. The 18F-FLT radioactivity concentrations within the volumes of interest were normalized to injected radioactivity and patient body weight to obtain SUV values [10]. PET data were preprocessed into various shapes to test our deep learning architectures: a) Time series (1D data). For each patient, a median of 574 (range, 63 – 6954, according to lesion size) TACs were extracted in a voxel-wise manner using the above-mentioned reference and lesion masks. TACs were linearly resampled onto a uniform time axis (one sample every 10 seconds for a total of 331 samples) (fig. 1). b) Static images (3D data). For each patient, a $30 \times 30 \times 10$ box was positioned around the tumor and, as before, flipped onto the contralateral healthy breast on the static PET image to obtain a control image. c) Dynamic images (4D data). The box outlined in b) was extended to the 45 timeframes of the dynamic PET acquisition.

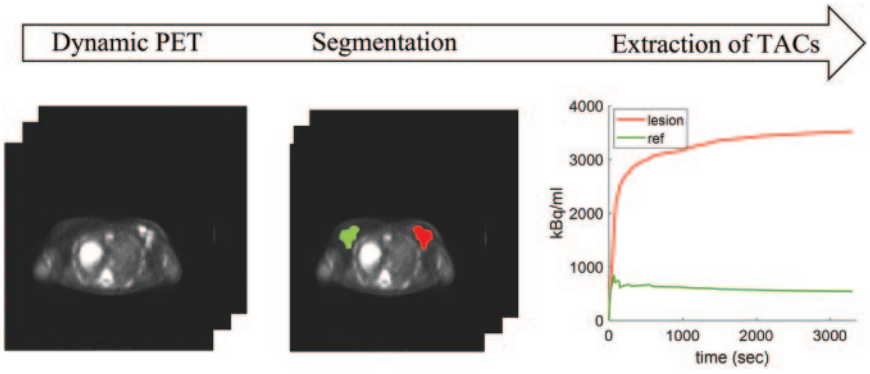

Fig. 1. – Pipeline. Volumes of interest were drawn in the lesion (red) and reference healthy tissue (green) for the extraction of time activity curves. They represent the concentration of the tracer in the tissue over time (average shown in the figure).

2˙3. *Spatiotemporal deep models for dynamic PET data processing.* – Dynamic PET data were employed to perform a binary classification task: tumor vs healthy tissue. For 1D data, we performed a binary classification between tumor and healthy reference tissue at the voxel level (*i.e.*, a massively univariate segmentation task). Given the temporal structure of this data, we compared convolutional monodimensional filters (CONV1D), long short-term memory (LSTM) models, and a combination of the two (CONV1D+LSTM). We also tested a transformer model adapted to time series classification [13], which, unlike the previously mentioned architectures, relies on an attention mechanism. In addition, we performed a binary classification between boxes which contained cancerous lesions and contralateral control regions using (separately) static and dynamic images. We compared models that employed three-dimensional convolutional layers (CONV3D) for the classification of static (3D) PET images to more sophisticated architectures, where we extracted spatiotemporal features from dynamic (4D) PET data using a combination of three-dimensional convolutional filters and LSTM in the CONV3D+LSTM model, and a set of depthwise separable convolutional layers where dynamic PET time evolution was encoded in the channel dimension of the filters (SECONV3D model). For comparison to standard clinical procedures, the performances of our models were compared to the commonly employed SUV measure. For 1D data classification, voxelwise SUV values were extracted from both lesion and reference tissue and used as input for XGBoost and support vector machine (SVM) classifiers. For image classification, static SUV images (3D data) were compared to both static and dynamic PET data using the CONV3D model. The sample was split into training (80%), validation (10%) and testing sets (10%) and normalized by the mean and standard deviation value evaluated on the training set. For each model, hyperparameter optimization was performed with Optuna with a random search sampler and 200 trials.

3. – Results

When classifying 1D time series, the best performance was obtained by the CONV1D model with 92% accuracy (AUC = 0.97) in comparison to 78% accuracy obtained with the LSTM (AUC = 0.86) and 80% accuracy obtained with a combination of the two (CONV1D+LSTM, AUC = 0.90). The transformer-based architecture discriminated lesion-derived TACs with 65% accuracy (AUC = 0.70). CONV1D model, based on temporal features only, showed better performance than gold standard models (SVM and XGBoost) across all metrics, whereas LSTM and CONV1D+LSTMs showed mixed or worse performance as compared to baselines (SVM and XGBoost delivered 76% and 68% accuracy (AUCs = 0.76, 0.67), respectively). For each model, the receiver operating characteristic curves and AUC values are also shown in fig. 2(A). When classifying 3D and 4D images, the CONV3D model reached a 63% (± 0.99) accuracy (0.59 ± 0.09 AUC). This performance was notably improved when combining both temporal and spatial feature extraction in the CONV3D+LSTM model (75% (± 0.09) accuracy, 0.81 (± 0.08) AUC) and when encoding the dynamic PET data time information on the channel dimension of the SECONV3D model (73% (± 0.07) accuracy, 0.84 (± 0.08) AUC). For each model, the receiver operating characteristic curves and AUC values are also shown in fig. 2(B). Overall, our dynamic approaches outperformed both the SUV_CONV3D model (CONV3D model applied onto static SUV images), which classified lesion and reference tissue with 60% (± 0.08) accuracy (AUC = 0.60 (± 0.15)), and SVM and XGBoost models trained on maximum SUV values (SUVmax) which delivered 62% (± 0.06) and 74% (± 0.05) accuracies, respectively.

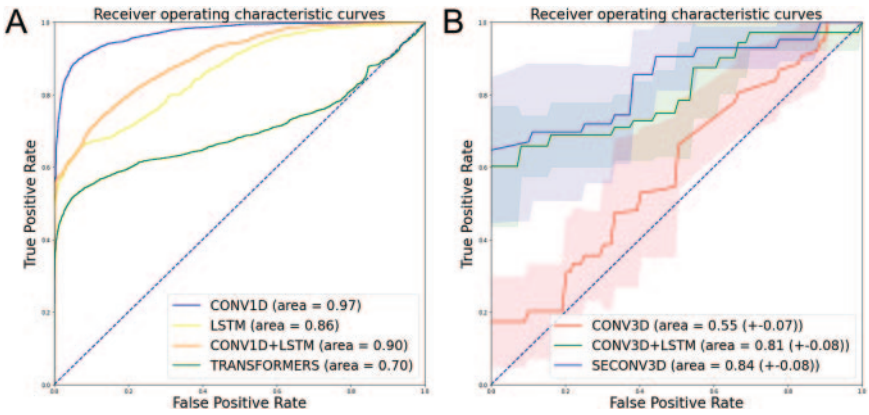

Fig. 2. – Receiver operating characteristic (ROC) curves corresponding to model performances in discriminating 1D (A) and 3/4D data (B) of lesion and reference tissue. In (B) average ROC curves with 1 standard deviation (shaded area) across folds are shown.

4. – Discussion and conclusions

This study used deep architectures to process both spatial and temporal information from static and dynamic PET images for a lesion-level classification task. When classifying 1D time series, the highest accuracy was obtained by the CONV1D model which outperformed the voxelwise classification of SUV values. Similarly, the CONV3D+LSTM and SECONV3D models, which processed the information provided by the time evolution of the PET signal (4D data), reached 75% accuracy and 73% accuracy, respectively, higher than gold standard SUV measures. This proof-of-concept study demonstrated that the diagnostic accuracy of static PET can be improved with an automatic and non-invasive deep learning approach that exploits the biochemical and metabolic information embedded in the tissue time activity curves obtained with dynamic PET acquisition. The results pave the way for more specific and sophisticated applications, where deep-learned time signal intensity pattern analysis can be used for tumor segmentation or tracer kinetic assessment without any pharmacokinetic model or measurement of the AIF.

REFERENCES

[1] Gupta N., *Chest*, **114** (1998) 1105.
[2] Wang G. *et al.*, *IEEE Trans. Radiat. Plasma Med. Sci.*, **4** (2020) 663.
[3] Thorwarth D. *et al.*, *BMC Cancer*, **5** (2005) 152.
[4] Sharma R., *Eur. J. Nucl. Med. Mol. Imaging*, **47** (2020) 1239.
[5] Sharma R., *J. Nucl. Med.*, **61** (2020) 1743.
[6] Dubash S., *Theranostics*, **10** (2018) 8677.
[7] Li Y., *Pharmaceutics*, **13** (2021) 8.
[8] Veronese M., *Comput. Math. Methods Med.*, **2016** (2016) 19.
[9] Kinahan P. E. *et al.*, *Semin. Ultrasound CT MR*, **31** (2010) 496.
[10] Westerterp M., *Eur. J. Nucl. Med. Mol. Imaging*, **34** (2007) 496.
[11] Keyes J. W., *J. Nucl. Med.*, **36** (1995) 1836.
[12] Karakatsanis N. A., *Phys. Med. Biol.*, **58** (2013) 7391.
[13] Vaswani A., arXiv:1706.03762 (2022).