

# Spatiotemporal learning of dynamic positron emission tomography data improves diagnostic accuracy in breast cancer

Monday, 30 May 2022 09:30 (20 minutes)

## Abstract

Positron emission tomography (PET) can reveal metabolic activity in a voxel-wise manner. PET analysis is commonly performed statically by analyzing the standardized uptake value (SUV). A dynamic PET acquisition can provide a map of the spatio-temporal concentration of the tracer in vivo, hence conveying information about radiotracer delivery to tissue, its interaction with the target and washout. Tissue-specific biochemical properties are embedded in the shape of time activity curves (TACs), which are conventionally employed along with information about blood plasma activity concentration i.e., the arterial input function (AIF), and specific compartmental models to obtain a full quantitative analysis of PET data. The main drawback of this approach is the need for invasive procedures requiring arterial blood sample collection during the whole scan. In this paper, we address the challenge of improving the PET diagnostic accuracy through an alternative approach based on the analysis of time signal intensity patterns. Specifically, we demonstrate the diagnostic potential of tissue TACs provided by a dynamic PET acquisition using various deep learning models. Our framework is shown to outperform the discriminative potential of classical SUV analysis, hence paving the way for a more accurate PET-based lesion discrimination without additional acquisition time or invasive procedures.

## 1. Introduction

Positron emission tomography (PET) allows the quantification of the biochemical properties of the tissue under investigation through the injection and detection of a targeted radiotracer [1]. To date, PET tissue time activity curves (TACs), and dynamic PET data in general, are mainly used for kinetic modelling, which involves fitting tracer-specific compartmental models[4]–[7] and requires information about blood plasma activity concentration, i.e., the so-called arterial input function (AIF), which requires arterial cannulation and collection of blood samples during the whole PET acquisition (which can last up to 90 minutes) to be measured. In rare cases, the shape of the TACs is evaluated visually to qualitatively discern tissue types.

Currently, the standardized uptake value (SUV), or its normalized version (SUVR), are the most widely employed PET-derived measures both in clinical and research application. These static maps are equivalent of the late phase of a dynamic PET acquisition, and therefore discard most of the information possibly present in the time-evolution of tissue-specific TACs. Importantly, the quantitative accuracy of SUV estimates relies on the many assumptions[8] which fail in clinical PET, resulting in non-negligible errors in the estimation of the rate of tracer uptake. PET tracer distribution is a dynamic process influenced by a wide number of factors (e.g., tissue type, patient, time of scan) which are reflected in temporal PET dynamics and cannot be accurately accounted by static SUV imaging.

The aim of this paper is to explore the information content and added value of employing tissue TACs only (i.e. eliminating the need for arterial blood sampling). To this end, we combine several deep learning architectures to analyze clinical data obtained from a cohort of breast cancer patients who received dynamic 3'-deoxy-3'-18F-fluorothymidine (18F-FLT) PET scans.

## 2. Methods

### A. Dataset

We employed a publicly available clinical 3'-deoxy-3'-18F-fluorothymidine (18F-FLT) dynamic PET data set of 44 breast cancer patients, part of the "ACRIN-FLT-Breast (ACRIN 6688)" collection in the Cancer Imaging Archive (TCIA)[10]–[12].

### B. Data pre-processing

For each patient, consecutive regions of interest were manually contoured around the tumor by an experienced radiologist on the static PET image (obtained as the average of the last 5 timeframes of the dynamic PET data). The 18F-FLT radioactivity concentrations within the volumes of interests were normalized to injected radioactivity and body weight to obtain SUV values[8]. For each patient, the masks obtained from lesion segmentation were flipped on to the contralateral breast for the delineation of a reference healthy region. For each patient, a median of 574 (range, 63 –6954, according to lesion size) TACs were extracted from 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).

### C. Spatiotemporal models for dynamic PET data

We compare convolutional mono-dimensional models (CONV1D) to long short-term memory (LSTM) models, and also evaluate the performance of a combined model (CONV1D+LSTM). For comparison to standard PET analysis, voxelwise SUV data were extracted from both lesion and reference tissue and used as input for an RUS boost tree ensemble classifier [14] as well as a boosted tree technique and a support vector machine. For each model, hyperparameter optimization was performed in the optima library (with random search sampler; 200 trials) and involved the number of units (for fully connected and LSTM layers), the number of filters and the dimension of the stride (for convolutional layers), the activation function, the learning rate, the loss function, the metric, gamma, C and kernel for the support vector machine (SVM) classifier and, for the XGboost, the maximum depth of a tree, number of estimators, and the fraction of columns to be subsampled. In the following sections, optimized values are listed.

CONV1D: Two mono-dimensional convolutional layers using a rectified linear unit (relu) and linear activation function, respectively. The filter size was set to 16 and 32, the kernel size to 2 for the first convolutional layer and to 4 for the second one and the stride to 2 and 3. The output of the last convolutional layer was flattened into four fully connected layers with 64 neurons followed by the last softmax layer for classification.

LSTM: Two LSTM and two fully connected layers. The units of the LSTM layers, using a sigmoid activation function, were set to 16. The output of the last LSTM layer was flattened into two fully connected layer with 64 neurons followed by the last softmax layer for classification.

CONV1D+LSTM: For spatiotemporal feature extraction, the model included both convolutional and recurrent neural networks. This architecture combines the previous two.

TRANSFORMER: The transformer model consisted of stacked self-attention and pointwise, fully connected layers for both the encoder and decoder. It was adapted for timeseries classification by Vaswani et al. For full details, see [15].

### D. Implementation

All experiments were conducted using python version 3.8, the keras deep learning library using TensorFlow as backend. We employed a Linux machine and two Nvidia Pascal TITAN X graphics cards with 12 GB RAM each.

### E. Performance and evaluation

The sample was split into training (80%), validation (10%) and testing sets (10%). An early stopping method was used to select the optimum number of training epochs and the batch size (Keras callback function monitoring the loss function with a patience set to 10). Models were trained for 200 epochs on a batch size of 64 and evaluated on the independent test set [17].

## 3. Results

Table 1 summarizes the results obtained with the validation of our models in terms of accuracy, precision and recall. When classifying 1D time series, the best performance was obtained by the CONV1D model with a 87% accuracy in comparison to 81% accuracy obtained with the LSTM and a 77% accuracy obtained with a combination of the two (CONV1D+LSTM). Transformer models classified lesion TACs with 69% accuracy. In comparison to SVM and XGboost used on the most commonly employed SUV imaging technique, which discriminated tumour tissue with 84% and 50% accuracy, respectively, our CONV1D model provided more stable results (confusion matrix –Table 1).

## 4. Discussion and Conclusion

We employed mono-dimensional filters to learn temporal patterns from time sequences data. The performance of our models was finally compared to the gold-standard SUV method. This proof-of-concept study demonstrated that the diagnostic accuracy of a static PET can be easily improved with a non-invasive deep learning approach which exploits the biochemical and metabolic information embedded in the tissue time activity curves obtained with a dynamic PET acquisition. Our 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, more interestingly, for tracer kinetic assessment without any pharmacokinetic model or measurement of the AIF.

## References

- [1] Gupta, Chest 1998; [2] Thorwarth, BMC Cancer 2005; [3] Sinibaldi, J Tissue Eng Regen Med 2018; [4] Sharma, Eur J Nucl Med Mol Imaging 2020; [5] Sharma, J Nucl Med 2020; [6] Dubash, Theranostics 2020; [7] Li, Pharmaceutics 2021; [8] Westerterp, Eur J Nucl Med Mol Imaging 2007; [9] Karakatsanis, Phys. Med. Biol 2013; [10] Kinahan, <http://doi.org/10.7937/K9/TCIA.2017.ol20zmxg>; [11] Kostakoglu, J Nucl Med 2015; [12] Clark, J Digit Imaging 2013; [13] Conti, Seminars in Cancer Biology 2021; [14] Seiffert, Trans. Syst., Man, Cybern 2010; [15] Vaswani, <http://arxiv.org/abs/1706.03762>; [16] Zhang, <http://arxiv.org/abs/1708.06578>; [17] Šimundić, EJIFCC 2009

**Primary authors:** INGLESE, Marianna (Department of Biomedicine and Prevention, University of Rome Tor Vergata); DUGGENTO, Andrea (University of Rome "Tor Vergata"); TOSCHI, Nicola (University of Rome "Tor Vergata", Harvard Medical School); BOCCATO, Tommaso (Department of Biomedicine and Prevention - University of Rome Tor Vergata)

**Presenter:** INGLESE, Marianna (Department of Biomedicine and Prevention, University of Rome Tor Vergata)

**Session Classification:** Metabolic Imaging with PET, MR and PET/MR

**Track Classification:** Metabolic Imaging with PET, MR and PET/MR