Spatiotemporal Learning of Dynamic Positron Emission Tomography Data Improves Diagnostic Accuracy in Breast Cancer

Marianna Inglese, Matteo Ferrante, Andrea Duggento, Tommaso Boccato, and Nicola Toschi

***Abstract*—Positron emission tomography (PET) is a noninvasive imaging technology able to assess the metabolic or functional state of healthy and/or pathological tissues. In clinical practice, PET data are usually acquired statically and normalized for the evaluation of the standardized uptake value (SUV). In contrast, dynamic PET acquisitions provide information about radiotracer delivery to tissue, its interaction with the target, and its physiological washout. The shape of the time activity curves (TACs) embeds tissue-specific biochemical properties. Conventionally, TACs are employed along with information about blood plasma activity concentration, i.e., the arterial input function, and tracer-specific compartmental models to obtain a full quantitative analysis of PET data. This method's primary disadvantage is the requirement for invasive arterial blood sample collection throughout the whole PET scan. In this study, we employ a variety of deep learning models to illustrate the diagnostic potential of dynamic PET acquisitions of varying lengths for discriminating breast cancer lesions in the absence of arterial blood sampling compared to static PET only. Our findings demonstrate that the use of TACs, even in the absence of arterial blood sampling and even when using only a share of all timeframes available, outperforms the discriminative ability of conventional SUV analysis.**

***Index Terms*—4-D convolutions, arterial input function, deep learning, dynamic positron emission tomography (PET), long short-term memory (LSTM).**

I. Introduction

Positron emission tomography (PET) allows the quantification of the biochemical properties of tissue through the injection and detection of a targeted radiotracer [1].

Over the past 40 years, there have been many studies on dynamic PET-based parametric imaging [2]. In fact, by examining the time-dependent evolution of the observed PET signal (i.e., dynamic PET), it is possible to retrieve a number of biological characteristics of interest and to generate multiparametric images that quantify the underlying biological mechanisms [2]. A PET image is, in fact, 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 how it is influenced by tissue- and tracer-specific washout effects. These tissue-specific biochemical characteristics can be inferred from the shape of the tissue time activity curves (TACs) [3], [4], which, to date, are mainly employed in conjunction with information about blood plasma activity concentration and specific compartmental models for the full quantification of dynamic PET data [5], [6], [7], [8]. In rare cases, the shape of the TACs is evaluated visually to qualitatively discern tissue types [9]. Along with the voxelwise extraction of TACs, this quantitative analysis often requires (especially for the kinetic assessment of a novel tracer), a painful and intrusive process, including arterial cannulation and blood sample collection throughout the whole PET acquisition (which can last up to 90 min). Therefore, in clinical practice, PET data are acquired following a static acquisition protocol. In particular, the standardized uptake value (SUV), or its normalized version (SUVR), is the most widely employed PET-derived measure in both clinical and research applications [10]. These static maps are equivalent to the late phase of dynamic PET acquisition and therefore discard the majority of information possibly present in the time evolution of tissue-specific TACs. A crucial assumption when calculating SUV estimates is that the amount of nonmetabolized tracer in the region of interest (ROI) is negligible compared to the amount of metabolized tracer there and that the time integral of plasma tracer concentration is proportional to the amount of

Manuscript received 7 March 2023; accepted 5 April 2023. Date of publication 26 April 2023; date of current version 3 July 2023. This work was supported in part by the Italian Ministry of University and Research (MUR) through Project #NEXTGENERATIONEU (NGEU) and in part by the European Innovation Council through Project CROSSBRAIN under Grant 101070908 and through Project BRAINSTORM under Grant 101099355. The work of Nicola Toschi was supported in part by the National Recovery and Resilience Plan (NRRP) through Project MNESYS—A Multiscale Integrated Approach to the Study of the Nervous System in Health and Disease (DN. 1553 11.10.2022) under Grant PE0000006; in part by the MUR-PNRR M4C2I1.3 PE6 Project (Heal Italia) under Grant PE00000019; and in part by the National Centre for HPC, Big Data and Quantum Computing, within the spoke "Multiscale Modeling and Engineering Applications." *(Corresponding author: Marianna Inglese.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by IRB.

Marianna Inglese is with the Department of Biomedicine and Prevention, University of Rome Tor Vergata, 00133 Rome, Italy, and also with the Department of Surgery and Cancer, Imperial College London, W12 0NN London, U.K. (e-mail: marianna.inglese@uniroma2.it).

Matteo Ferrante, Andrea Duggento, and Tommaso Boccato are with the Department of Biomedicine and Prevention, University of Rome Tor Vergata, 00133 Rome, Italy (e-mail: matteo.ferrante@uniroma2.it; duggento@med.uniroma2.it; tommaso.boccato@uniroma2.it).

Nicola Toschi is with the Department of Biomedicine and Prevention, University of Rome Tor Vergata, 00133 Rome, Italy, and also with the Department of Radiology, Athinoula A. Martinos Center for Biomedical Imaging and Harvard Medical School, Boston, MA 02129 USA (e-mail: toschi@med.uniroma2.it).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TRPMS.2023.3268361.

Digital Object Identifier 10.1109/TRPMS.2023.3268361

tracer injected, normalized by body weight [11]. This assumption frequently fails in clinical PET, leading to non-negligible errors in the calculation of the rate of tracer uptake [12]. The distribution of PET tracers is a dynamic process that is affected by several factors (such as tissue type, patient, scan time) which cannot be adequately predicted by static PET imaging (e.g., SUV analysis). It follows that dynamic PET may be able to lessen the large time dependence seen in SUV quantification of normal tissue and tumor uptake values, hence allowing greater flexibility as well as reliability in clinical practice [13]. Of note, dynamic PET data used for tumor detection [14], [15] have already demonstrated the superiority of an automated dynamic observer over a static observer, supporting the conclusion that lesion detection can be enhanced by using time-resolved information [16].

The main objective of this article 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′-^{18}F-fluorothymidine (^{18}F-FLT) PET scans for a relatively straightforward task (classification between lesion and reference tissue) in order to highlight any potential static versus dynamic effect.

II. Materials and Methods

A. Dataset

We employed a publicly available clinical ^{18}F-FLT PET dataset consisting of 44 breast cancer patients, part of the "ACRIN-FLT-Breast (ACRIN 6688)" collection in the cancer imaging archive (TCIA) [17], [18], [19]. Eligibility criteria included histologically confirmed breast cancer diagnosis, primary breast cancer measuring greater than or equal to 2.0 cm, being a candidate for systemic neoadjuvant chemotherapy (NAC) and surgical resection of the residual primary tumor after NAC, and no evidence of stage IV disease [18]. Dynamic PET images were acquired after a bolus injection of 167 MBq (mean; range, 110–204 MBq). The dynamic scan comprises 45 timeframes (16 × 5, 7 × 10, 5 × 30, 5 × 60, 5 × 180, and 6 × 300 s) for a 60-min acquisition duration (mean, 70 min; range, 50–101 min). All patients were scanned on calibrated and ACRIN-accredited PET/CT scanners, which included a review of image quality and testing of SUVs using a uniform phantom [18]. For the purpose of this study, only baseline scans were employed.

B. Data Processing

For each patient, volumes of interest were manually drawn by an experienced radiologist using the combined summed PET (obtained as the average of the last five timeframes of the dynamic PET data) and CT volume. The ^{18}F-FLT radioactivity concentrations within the volumes of interest were normalized to injected radioactivity and patient body weight to obtain SUV values [11]. For each patient, an additional ROI was obtained from the centroid of the healthy contralateral breast where the same mask obtained from lesion segmentation was flipped and used for the delineation of a reference healthy region. PET data were preprocessed into various shapes to test our deep learning architectures.

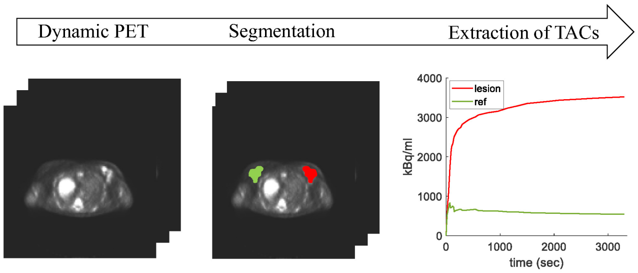

Fig. 1. TACs extraction. 3′-deoxy-3′-^{18}F-fluorothymidine (^{18}F-FLT) PET data was dynamically acquired on a cohort of 44 breast cancer patients belonging to the "ACRIN-FLT-Breast (ACRIN 6688)" collection in TCIA. Volumes of interest were drawn in the lesion (red) and reference healthy tissue (green) for the extraction of TACs They represent the concentration of the tracer in the tissue over time (average shown in the figure). For each patient, a median of 574 (range, 63–6954) TACs were extracted from each mask.

1) *Time-Series (1-D Data):* For each patient, a median of 574 (range, 63–6954, according to lesion size) TACs were extracted in a voxelwise manner using the reference and lesion masks. TACs were linearly resampled onto a uniform time axis (one sample every 10 s for a total of 331 samples) (Fig. 1).
2) *Static Images (3-D Data):* For each patient, a 30 × 30 × 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.
3) *Dynamic Images (4-D Data):* The box outlined in 2) was extended to the 45 timeframes of the dynamic PET acquisition.

C. Spatiotemporal Models for Dynamic PET Data

Dynamic PET data were employed to perform a binary classification task: tumor versus healthy tissue. For 1-D 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 these data, we compared convolutional monodimensional filters (CONV1D), long short-term memory (LSTM) models, and a combination of the two (CONV1D + LSTM). In addition, we performed a binary classification between boxes that contained cancerous lesions and contralateral control regions using (separately) static and dynamic images. We compared models that employed 3-D convolutional layers (CONV3D) for the classification of static (3-D) PET images to more sophisticated architectures where we extracted spatiotemporal features from dynamic (4-D) PET data using a combination of 3-D 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). We also tested a transformer model adapted for time-series classification [20], which, unlike the previously mentioned architectures, relies on an attention mechanism. For comparison to standard clinical procedures, the performances of our models were compared to the commonly employed SUV measure. For 1-D data classification, voxelwise SUV values were extracted from both

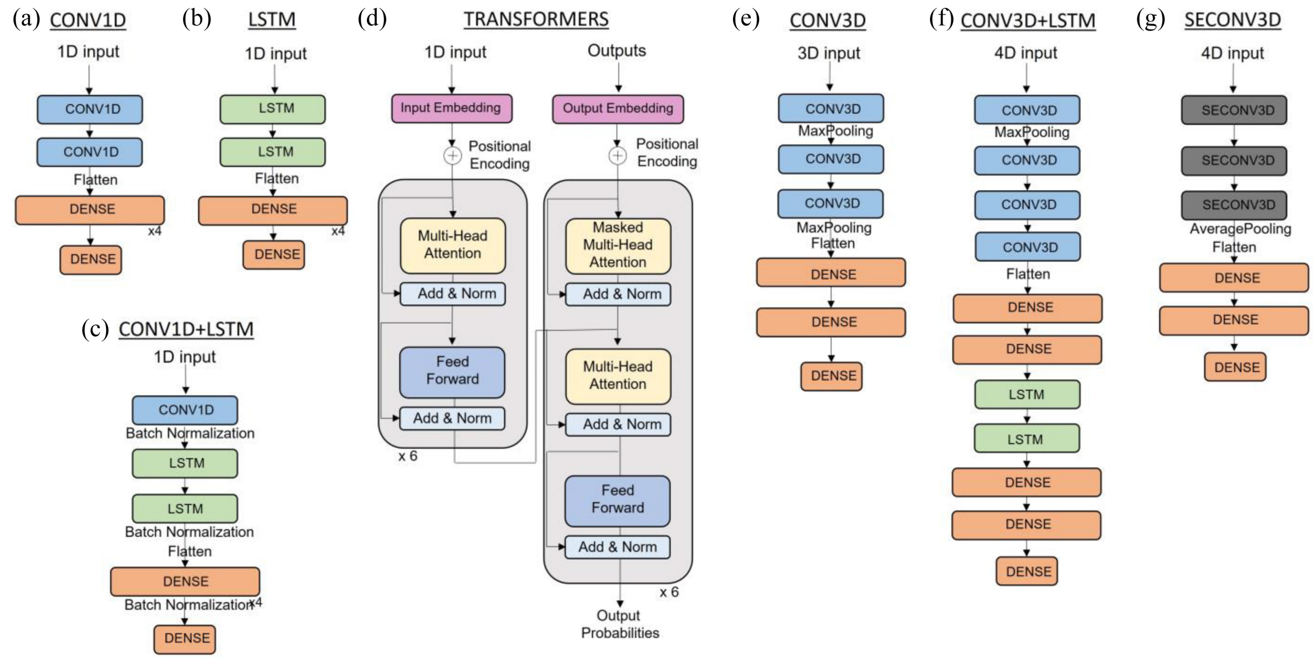

Fig. 2. Representative model architectures. (a) CONV1D comprises two mono-dimensional convolutional layers, four fully connected layers followed by the first softmax layer for classification. (b) LSTM comprises two long-short term memory (LSTM) and four fully connected layers followed by the last softmax layer for classification. (c) CONV1D+LSTM combines (a) and (b) for spatiotemporal feature extraction. (d) TRANSFORMERS comprises stacked self-attention and pointwise, fully connected layers for both the encoder and decoder (adapted for time series classification by Vaswani et al. [20]). (e) CONV3D comprises three 3-D convolutional layers, two fully connected layers, and the last softmax classification layer. (f) CONV3D+LSTM combines a cascade of convolutional and recurrent neural networks for the extraction of spatial and temporal features. The architecture comprises four 3-D convolutional layers followed by two fully connected layers, two LSTM layers followed by two fully connected layers, and the last softmax classification layer. (g) SECONV3D comprises three depthwise separable convolutional layers, two fully connected layers followed by the last softmax layer for classification. The wording "dense" (Keras nomenclature) refers to a module used in convolutional neural networks that connects all layers (with matching feature-map sizes) directly with each other.

lesion and reference tissue and used as input for XGBoost and support vector machine (SVM) classifiers and for a linear discriminant analysis (singular value decomposition (SVD) model) [13]. For image classification, static SUV images (3-D data) were compared to both static and dynamic PET data using the CONV3D model. For each model, hyperparameter optimization was performed with Optuna (with random search sampler and 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, gamma, and kernel for the SVM classifier and, for the XGBoost model, the maximum depth of a tree, the minimum sum of instance weight needed in a child, and the subsample ratio of columns for each tree (Table I). For the SECONV3D model, hyperparameter optimization was performed with a random search sampler and 100 trials. In the following sections, optimized values are listed [14]. Details of the models we employed are as follows.

CONV1D: Two mono-dimensional convolutional layers using a rectified linear unit (*ReLU*) and linear activation function. 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 convolutional layer, and the stride to 2 and 3. The output of the last convolutional layer was flattened into four fully connected layers with 64 neurons using sigmoid, *ReLU*, linear, and sigmoid activation functions, followed by the last softmax layer for classification [Fig. 2(a)].

LSTM: Two LSTM and four fully connected layers. The units of the LSTM layers, using a *ReLU* activation function, were set to 16 and 4 for the first and second layers, respectively. The output of the last LSTM layer was flattened into four fully connected layers with 64 neurons using a *ReLU* activation function, followed by the last softmax layer for classification [Fig. 2(b)].

CONV1D+LSTM: For spatiotemporal feature extraction, the model included both convolutional and recurrent neural networks (RNNs). This architecture combines the previous two [Fig. 2(c)].

TRANSFORMERS: The transformer model was adapted for time-series classification by Vaswani et al. [20]. This architecture consists of stacked self-attention and pointwise, fully connected layers for both the encoder and decoder. The encoder is composed of a stack of six identical layers. Each layer has two sublayers. The first is a multihead self-attention mechanism, and the second is a simple, positionwise fully connected feed-forward network. A residual connection was used around each of the two sublayers, followed by layer normalization. The decoder is also composed of a stack of six identical layers. In addition to the two sublayers in each encoder layer, the decoder contains a third sublayer, which performs multihead attention over the output of the encoder stack. As in the encoder, residual connections are used around each of the sublayers, followed by layer normalization [Fig. 2(d)].

CONV3D: Three 3-D convolutional layers using linear, sigmoid, and linear activation functions. The filter size was

TABLE I
HYPERPARAMETER VALUES EMPLOYED IN OPTIMIZATION

Parameter	**Search space**
Number of filters Conv1D layer	[4, 16, 64, 128, 512]
Number of filters Conv3D layer	[16, 64, 128, 512]
Number of units LSTM layer	[16, 64, 128]
Stride	[1, 2]
Number of units Dense layer	[64, 100, 200, 512]
Number of filters SECONV3D layer	[128,256,512]
Number of units SECONV3D	[512,256,128]
Batch Size SECONV3D	[4,8,12]
XGBoost maximum depth	[3, 5, 6, 10, 15, 20]
XGBoost Subsample ratio of columns for each tree	[0.3, 0.4, 0.5, 0.7]
XGBoost scale positive weight	[1, n1/n2, sqrt(n1/n2)]
XGBoost minimum child weight	[1, 3, 5, 7]
SVM kernel	rbf, poly, sigmoid
SVM gamma	1, 0.1, 0.01, 0.001
SVM C	0.1, 1, 10, 100
Activation function	*ReLU*, linear, sigmoid, gelu, leaky_*ReLU*
Adam learning rate	1e-1, 1e-2,1e-3,1e-4, 1e-5,
Loss function	mae, mse, sparse categorical crossentropy
Metric	accuracy, sparse categorical accuracy

set to 16, and the kernel size and stride were set to 2. The output of the last convolutional layer was flattened into two fully connected layers with 64 neuron (using the *ReLU* activation function) and the last softmax classification layer [Fig. 2(e)].

CONV3D+LSTM: A cascade of convolutional and RNN architectures that was originally built for electroencephalographic signals [21]. Spatial features are extracted from boxed dynamic PET images and then fed into the RNN for the extraction of temporal features. One fully connected layer receives the output of the last time step of the RNN layers and feeds the softmax layer for final classification. In detail, there are four time-distributed 3-D convolutional layers with filter sizes set to 64 (except for the first one with a filter size set to 32 and followed by a maxpooling layer with the pool size set to 2) with the same kernel size set to 3 and stride to 2 (except for the first one where it was set to 1) for spatial feature extraction. In each convolutional operation, zero-padding techniques were used to create feature maps with the same size as the raw input PET data. The output of the time-distributed 3-D convolution block was flattened to feed two fully connected layers with 64 neurons (and a sigmoid and *ReLU* activation function). They are followed by the RNN block made of two LSTM layers (with 64 units each) and two final fully connected layers, a *ReLU* activation function and the last softmax for classification [Fig. 2(f)].

SECONV3D: Three 3-D depthwise separable convolutional layers with a *gelu* activation function. The kernel size of the first two blocks and the stride were set to 2. An average pooling layer (size = 2) was applied before passing the output to a multilayer perceptron with two hidden layers of 128 and 64 neurons using the *ReLU* activation function. Data were encoded with the time dimension as the channel dimension, allowing the evaluation of the time interaction through the convolutions [Fig. 2(g)].

D. Implementation

All experiments were conducted using Python version 3.8, the Keras deep learning library, using TensorFlow as the backend. We employed a Linux machine and two Nvidia Pascal TITAN V graphics cards with 12-GB RAM each. SECONV3D was implemented in PyTorch and trained on the same machine.

E. Performance Evaluation

In all cases, 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. 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 patience set to 10). In the case of CONV3D, CONV3D+LSTM, and SECONV3D, given the dataset size, a fivefold cross validation was performed to quantify performance, reported in terms of area under the receiver operating characteristic (ROC) curve (AUC), accuracy, precision and recall [22].

F. Predictive Value of Partial Dynamic PET Data

Our main assumption is that dynamic PET can provide additional information compared to static imaging. In a clinical context, this is often counterbalanced by the limited in-scanner time available to each patient. In this context, we extended our analysis to multiple training and optimization of our models with increasingly shorter segments (starting from the first timepoint) of dynamic PET data, hence testing the additional hypothesis that even an incomplete dynamic PET acquisition can be of value, especially given that most of the dynamic range is contained in the first part of the acquisition. We, therefore, explored the added value of dynamic PET as a function of the number of time points and hence of acquisition time.

III. RESULTS

Table II summarizes the results obtained when classifying 1-D time series. For each model, the ROC curves and AUC values are also shown in Fig. 3(a). The best performance was obtained by the CONV1D model with a 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 models, based on temporal features only, showed better performance than traditional models (SVM and XGBoost) across all metrics, whereas LSTM and CONV1D+LSTM showed mixed or worse performance compared to baselines (SVM and XGBoost delivered 76% and 68% accuracy (AUCs

TABLE II
CLASSIFICATION OF 1-D DATA. NUMBER OF TRAINABLE PARAMETERS AND MODEL PERFORMANCE

Classification model	Input Data	Training Samples	Validating Samples	Testing Samples	Trainable Parameters	AUC	Accuracy	Precision		Recall	
		n_sample x t_dim	*n_sample x t_dim*	*n_sample x t_dim*				Ref	Lesion	Ref	Lesion
CONV1D	Dynamic PET data (time activity curves)	68951 x 331	8619 x 331	8619 x 331	129490	0.97	0.92	0.91	0.92	0.93	0.92
LSTM	Dynamic PET data (time activity curves)	68951 x 331	8619 x 331	8619 x 331	354882	0.86	0.78	0.70	0.96	0.94	0.61
CONV1D +LSTM	Dynamic PET data (time activity curves)	68951 x 331	8619 x 331	8619 x 331	94546	0.90	0.80	0.79	0.82	0.83	0.79
TRANSFORMERS	Dynamic PET data (time activity curves)	68951 x 331	8619 x 331	8619 x 331	48202	0.70	0.65	0.59	1.0	0.99	0.33
XGBoost	Dynamic SUV data (voxelwise)	22822781 x 1	2852889 x 1	2852889 x 1	-	0.67	0.68	0.87	0.62	0.39	0.94
SVM	Dynamic SUV data (voxelwise)	22822781 x 1	2852889 x 1	2852889 x 1	-	0.76	0.76	0.68	0.93	0.95	0.57
Linear Discriminant (SVD)	Static SUV data (voxelwise)	77570 x 1	-	8619 x 1	-	0.75	0.75	0.67	0.94	0.96	0.56
XGBoost	Static SUV data (voxelwise)	68951 x 1	8619 x 1	8619 x 1	-	0.85	0.85	0.81	0.93	0.91	0.78
SVM	Static SUV data (voxelwise)	68951 x 1	8619 x 1	8619 x 1	-	0.80	0.82	0.80	0.84	0.85	0.79

TABLE III
CLASSIFICATION OF 3-D AND 4-D DATA. NUMBER OF TRAINABLE PARAMETERS AND MODEL PERFORMANCE

Classification model	Input Data	Training Samples	Validating Samples	Testing Samples	Trainable Parameters	AUC	Accuracy	Precision		Recall	
		n_sample x (x_dim x y_dim x z_dim x t_dim)	*n_sample x (x_dim x y_dim x z_dim x t_dim)*	*n_sample x (x_dim x y_dim x z_dim x t_dim)*				Ref	Lesion	Ref	Lesion
CONV3D	Static PET image	64 x (30 x 30 x 10 x 1)	8 x (30 x 30 x 10 x 1)	8 x (30 x 30 x 10 x 1)	37234	0.59 (±0.09)	0.63 (±0.99)	0.56 (±0.13)	0.61 (±0.21)	0.67 (±0.17)	0.48 (±0.15)
CONV3D +LSTM	Dynamic PET image	64 x (30 x 30 x 10 x 45)	8 x (30 x 30 x 10 x 45)	8 x (30 x 30 x 10 x 45)	364610	0.81 (±0.08)	0.75 (±0.09)	0.69 (±0.09)	0.91 (±0.09)	0.96 (±0.03)	0.55 (±0.17)
SECONV3D	Dynamic PET image	64 x (30 x 30 x 10 x 45)	8 x (30 x 30 x 10 x 45)	8 x (30 x 30 x 10 x 45)	558340	0.84 (±0.08)	0.73 (±0.07)	0.84 (±0.14)	0.73 (±0.14)	0.66 (±0.21)	0.81 (±0.18)
CONV3D	Static SUV image	64 x (30 x 30 x 10 x 1)	8 x (30 x 30 x 10 x 1)	8 x (30 x 30 x 10 x 1)	35074	0.60 (±0.15)	0.60 (±0.08)	0.68 (±0.19)	0.63 (±0.21)	0.37 (±0.20)	0.51 (±0.12)
XGBoost	Static SUVmax data	64	8	8	-	0.73 (±0.03)	0.74 (±0.05)	0.78 (±0.07)	0.68 (±0.15)	0.67 (±0.08)	0.78 (±0.15)
SVM	Static SUVmax data	64	8	8	-	0.61 (±0.05)	0.62 (±0.06)	0.59 (±0.09)	0.65 (±0.18)	0.79 (±0.06)	0.42 (±0.11)

= 0.76, 0.67), respectively, when applied to dynamic SUV data [(voxelwise)—Table II)]. For comparison with the clinical gold standard, a linear discriminant analysis (SVD) as well as SVM and XGBoost models were trained on static SUV values (Static SUV values (voxelwise), see table) and discriminated tumor tissue with 75%, 85%, and 82% accuracy, respectively (Table II). Similarly, Table III summarizes the results obtained when classifying lesions using 3-D and 4-D data. For each model, the ROC curves and AUC are shown in Fig. 3(b). 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). Overall, our dynamic approaches outperformed both the SUV_CONV3D model (CONV3D model applied to 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.

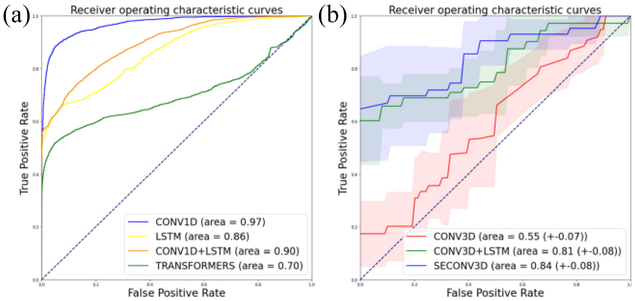

Fig. 3. ROC curves corresponding to model performances in discriminating (a) 1-D and (b) 3/4-D data of lesion and reference tissue. In (b), average ROC curves with 1 standard deviation (shaded area) across folds are shown.

Finally, the performances of our best models for both voxel (CONV1D) and image classification (SECONV3D) were tested on an increasingly shorter dynamic PET dataset (Fig. 4). In these experiments, the dynamic scan time was progressively reduced from 45 time frames to 1 time frame.

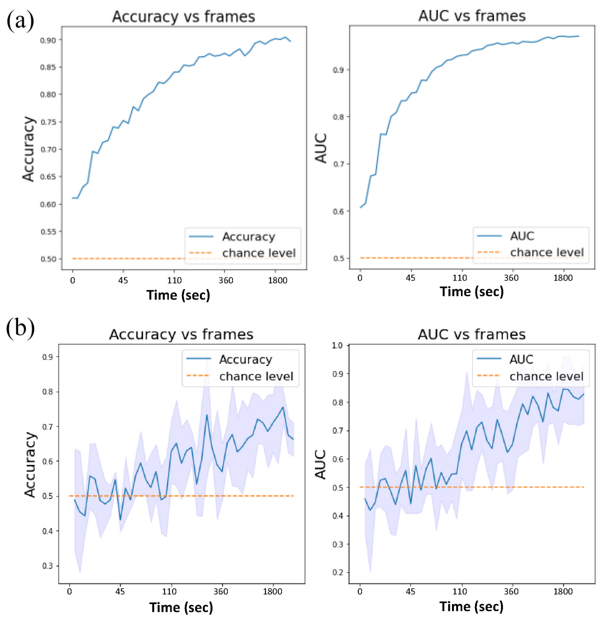

Fig. 4. Accuracy and AUC values obtained when testing the (a) CONV1D and (b) SECONV3D models for voxel and image classification, respectively, on a dynamic PET dataset with a variable duration. In (b), mean values with 1 standard deviation (shaded area) across folds are shown.

The accuracy of the SECONV3D model trained on the full dataset (45 time frames/55 min scan, 73% ± 0.07) decreased by only 11% when trained on 31 time frames/6 min scan [62% (±0.05)]. Similarly, the accuracy of the CONV1D model trained on the full dataset (92%) decreased by only 5% when trained on 31 time frames (87%).

IV. Discussion and Conclusion

The ability to define diverse biological processes at multiple levels is a unique advantage offered by PET technology [23]. The first level of analysis only requires static PET imaging, the gold standard in clinical practice. The radiotracer is administered intravenously and, after a predetermined delay from the injection (e.g., 60 min for the FDG tracer), a few minutes of PET acquisition (2–5 min per bed position) are initiated [24]. The acquired static image, which is an average snapshot of the entire acquisition time for each PET bed position, is usually normalized to generate SUV maps [10]. The use of SUV is now commonplace in clinical PET/CT oncology imaging and plays a crucial part in assessing patient response to cancer therapy [10], [25]. In fact, the SUV removes the variability of the raw static PET image introduced by differences in patient size and the amount of the injected radiotracer. However, using SUV only has a number of drawbacks, e.g., it cannot be used in multicenter studies due to intercenter variability, it is heavily influenced by physical and technical acquisition parameters, and it provides inaccurate semiautomated lesion segmentations based on intensity thresholds [26]. The limited information provided by a static PET acquisition can be augmented with a dynamic acquisition that starts when the radiotracer is injected and consists of a continuous acquisition of a PET bed position for a few minutes to 1 h (or more) depending on the mathematical model to be used for image postprocessing [27]. Currently, dynamic PET is mainly employed in research applications and allows the quantification of radiotracer kinetics, hence capturing information not available with conventional static acquisition protocols [5], [6], [7], [8]. Dynamic PET returns an in vivo map of the spatiotemporal tracer concentration, which incorporates information about its interaction with the target and washout effects. This information is embedded in the shape of the tissue TACs, which reflect tissue-specific biochemical properties that are lost when a static PET protocol is acquired [28]. The quantification of tracer uptake based on compartmental modeling approaches, as applied to dynamic PET images, improves both tumor characterization and treatment response monitoring.

In this study, we demonstrated the superior diagnostic potential of dynamic over static PET imaging using deep learning models for a binary classification task. We employed monodimensional filters (e.g., CONV1D and LSTM) to learn temporal patterns from time sequence (1-D) data for the voxelwise classification of tumor versus reference tissue. This was done without pharmacokinetic modeling and without invasively measuring the AIF. In addition, we employed more sophisticated architectures, able to process both spatial and temporal information from static and dynamic PET images (3-D and 4-D data) for a lesion-level (as opposed to voxel-level) classification task (i.e., tumor versus reference tissue).

The performances of our models were compared to the gold-standard SUV analysis. For TAC classification, the highest accuracy was obtained by the CONV1D model (92% accuracy and AUC = 0.97), which, with two convolutional layers followed by four fully connected layers, outperformed the voxelwise classification of SUV values, which reached 85% accuracy with the XGBoost model, 82% accuracy with the SVM model, and 75% accuracy with a linear discriminant analysis (SVD). Interestingly, the performance of SVM and XGBoost with dynamic SUV data is a few percent points worse than that with static SUV data. Provided that no statistically significant tests were run, this may be due to the fact that the number of features included in dynamical data may be too large for a simple machine learning model to represent/separate. For the image classification task, comparable results were obtained by the CONV3D+LSTM and SECONV3D models, which processed the information provided by the time evolution of the PET signal (4-D data). In particular, combining both temporal and spatial feature extraction, the CONV3D+LSTM model reached 75% accuracy. Despite the complexity of the CONV3D model, when we applied it to both raw static PET data and SUV maps (3-D images), we obtained lower performances than those shown in clinical studies [18]. In this regard, it is important to note that the literature reports clinical FLT studies where statistical tests are performed on SUV mean (or max, peak) values averaged over the whole lesion (while the input of our CONV3D model is the whole 3-D image). Perhaps more importantly, it is customary in clinical studies to use the whole dataset for training and evaluation, as opposed to evaluating models on unseen test data potentially inflating reported results due to overfitting. Finally, it should be noted that higher model complexity does not necessarily guarantee better performance unless very large amounts of data are available.

Our results show that by encoding the time information provided by dynamic PET data in the channel dimension of the 3-D convolutional filters in the SECONV3D architecture, the model reached 73% accuracy. Overall, the performance obtained when using gold standard SUV measures in classifying tumor versus reference tissue, as well as the performance of the raw static PET data in the same classification task, was lower than the one obtained using dynamic PET data in the shape of both tissue TACs (1-D) and 4-D images. Of note, the SECONV3D model, which used depthwise separable convolutions, performed reasonably well despite challenges related to both the use of a small dataset and the number of trainable parameters of a conventional 4-D convolution layer. Furthermore, we were not able to perform any pharmacokinetic analysis or to provide any comparison with kinetic parameters, as the database we employed did not include information about the percentage of the metabolite FLT-glucuronide present in the blood after the injection of [^{18}F]FLT tracer [29], [30]. Therefore, the parent plasma (metabolite-corrected) input function necessary for pharmacokinetic fitting was not available.

Finally, we investigated the robustness of our best models when applied to a shorter dynamic PET dataset. This was done to investigate the suitability of dynamic PET in a clinical context, where scanner time is heavily constrained and arterial blood sampling is often hampered by additional logistic challenges. Our results show that by reducing dynamic scan times from 45 timeframes (55-min scan) as low as 30 timeframes (6-min scan), the accuracy and AUC values of both models (CONV1D and SECONV3D for 1-D and 4-D data classification, respectively) did not suffer notable decreases compared to when using the full dynamic dataset. Further work could address a more fine-grained classification of time-dependent information with nonstandard techniques, such as, e.g., parasitic modeling [31], deployed in a cloud-based environment [32].

This proof-of-concept study demonstrated that the diagnostic accuracy of static PET can be easily improved with an automatic and noninvasive deep learning approach that exploits the biochemical and metabolic information embedded in the tissue TACs obtained with dynamic PET acquisition. Importantly, this goal appears feasible especially in light of the fact that some classifiers are able to deliver good performance while employing only a small portion (3 min) of the dynamic data. 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.

Acknowledgment

All authors declare that they have no known conflicts of interest in terms of competing financial interests or personal relationships that could have an influence or are relevant to the work reported in this paper. The Titan V GPUs employed in this research were generously donated to NT by NVIDIA.

References

[1] N. Gupta, H. Gill, G. Graeber, H. Bishop, J. Hurst, and T. Stephens, "Dynamic positron emission tomography with F-18 fluorodeoxyglucose imaging in differentiation of benign from malignant lung/mediastinal lesions," *Chest*, vol. 114, no. 4, pp. 1105–1111, Oct. 1998, doi: 10.1378/chest.114.4.1105.

[2] G. Wang, A. Rahmim, and R. N. Gunn, "PET parametric imaging: Past, present, and future," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 4, no. 6, pp. 663–675, Nov. 2020, doi: 10.1109/TRPMS.2020.3025086.

[3] D. Thorwarth, S.-M. Eschmann, J. Scheiderbauer, F. Paulsen, and M. Alber, "Kinetic analysis of dynamic 18F-fluoromisonidazole PET correlates with radiation treatment outcome in head-and-neck cancer," *BMC Cancer*, vol. 5, p. 152, Dec. 2005, doi: 10.1186/1471-2407-5-152.

[4] R. Sinibaldi et al., 'Multimodal-3D imaging based on μMRI and μCT techniques bridges the gap with histology in visualization of the bone regeneration process," *J. Tissue Eng. Regen. Med.*, vol. 12, no. 3, pp. 750–761, Mar. 2018, doi: 10.1002/term.2494.

[5] R. Sharma et al., "[18F]fluciclatide PET as a biomarker of response to combination therapy of pazopanib and paclitaxel in platinum-resistant/refractory ovarian cancer," *Eur. J. Nucl. Med. Mol. Imag.*, vol. 47, no. 5, pp. 1239–1251, 2020, doi: 10.1007/s00259-019-04532-z.

[6] R. Sharma et al., "Monitoring response to transarterial chemoembolization in hepatocellular carcinoma using 18F-fluorothymidine PET," *J. Nucl. Med.*, vol. 61, no. 12, pp. 1743–1748, 2020, doi: 10.2967/jnumed.119.240598.

[7] S. Dubash et al., "Spatial heterogeneity of radiolabeled choline positron emission tomography in tumors of patients with non-small cell lung cancer: First-in-patient evaluation of [18F]fluoromethyl-(1, 2-2H4)-choline," *Theranostics*, vol. 10, no. 19, pp. 8677–8690, 2020, doi: 10.7150/thno.47298.

[8] Y. Li et al., "Consideration of metabolite efflux in radiolabelled choline kinetics," *Pharmaceutics*, vol. 13, no. 8, p. 1246, 2021, doi: 10.3390/pharmaceutics13081246.

[9] M. Veronese, G. Rizzo, A. Bertoldo, and F. E. Turkheimer, "Spectral analysis of dynamic PET studies: A review of 20 years of method developments and applications," *Comput. Math. Methods Med.*, vol. 2016, Dec. 2016, Art. no. e7187541, doi: 10.1155/2016/7187541.

[10] P. E. Kinahan and J. W. Fletcher, "PET/CT standardized uptake values (SUVs) in clinical practice and assessing response to therapy," *Semin. Ultrasound CT MR*, vol. 31, no. 6, pp. 496–505, Dec. 2010, doi: 10.1053/j.sult.2010.10.001.

[11] M. Westerterp et al., "Quantification of FDG PET studies using standardised uptake values in multi-centre trials: Effects of image reconstruction, resolution and ROI definition parameters," *Eur. J. Nucl. Med. Mol. Imag.*, vol. 34, no. 3, pp. 392–404, Mar. 2007, doi: 10.1007/s00259-006-0224-1.

[12] J. W. Keyes, "SUV: Standard uptake or silly useless value?" *J. Nucl. Med.*, vol. 36, no. 10, pp. 1836–1839, 1995.

[13] N. A. Karakatsanis, M. A. Lodge, A. K. Tahari, Y. Zhou, R. L. Wahl, and A. Rahmim, "Dynamic whole-body PET parametric imaging: I. Concept, acquisition protocol optimization and clinical application," *Phys. Med. Biol.*, vol. 58, no. 20, pp. 7391–7418, Sep. 2013, doi: 10.1088/0031-9155/58/20/7391.

[14] K.-P. Wong, D. Feng, S. R. Meikle, and M. J. Fulham, "Segmentation of dynamic PET images using cluster analysis," *IEEE Trans. Nucl. Sci.*, vol. 49, no. 1, pp. 200–207, Feb. 2002, doi: 10.1109/TNS.2002.998752.

[15] J. G. Brankov, N. P. Galatsanos, Y. Yang, and M. N. Wernick, "Segmentation of dynamic PET or fMRI images based on a similarity metric," *IEEE Trans. Nucl. Sci.*, vol. 50, no. 5, pp. 1410–1414, Oct. 2003, doi: 10.1109/TNS.2003.817963.

[16] Z. Li, Q. Li, X. Yu, P. S. Conti, and R. M. Leahy, "Lesion detection in dynamic FDG-PET using matched subspace detection," *IEEE Trans. Med. Imag.*, vol. 28, no. 2, pp. 230–240, Feb. 2009, doi: 10.1109/TMI.2008.929105.

[17] P. Kinahan, M. Muzi, B. Bialecki, and C. Laura. "Data from ACRIN-FLT-Breast." The Cancer Imaging Archive. 2017. [Online]. Available: http://doi.org/10.7937/K9/TCIA.2017.ol20zmxg

[18] L. Kostakoglu et al., "A phase II study of 3$'$-deoxy-3$'$-18F-fluorothymidine PET in the assessment of early response of breast cancer to neoadjuvant chemotherapy: Results from ACRIN 6688," *J. Nucl. Med.*, vol. 56, no. 11, pp. 1681–1689, Nov. 2015, doi: 10.2967/jnumed.115.160663.

[19] K. Clark et al., ‘The cancer imaging archive (TCIA): Maintaining and operating a public information repository,” *J. Digit. Imag.*, vol. 26, no. 6, pp. 1045–1057, Dec. 2013, doi: 10.1007/s10278-013-9622-7.

[20] A. Vaswani et al., “Attention is all you need,” Dec. 2017, *arXiv:1706.03762*.

[21] D. Zhang, L. Yao, X. Zhang, S. Wang, W. Chen, and R. Boots, “Cascade and parallel convolutional recurrent neural networks on EEG-based intention recognition for brain computer interface,” Jun. 2021, *arXiv:1708.06578*.

[22] A.-M. Šimundić, “Measures of diagnostic accuracy: Basic definitions,” *EJIFCC*, vol. 19, no. 4, pp. 203–211, Jan. 2009.

[23] A. K. Shukla and U. Kumar, “Positron emission tomography: An overview,” *J. Med. Phys.*, vol. 31, no. 1, pp. 13–21, 2006, doi: 10.4103/0971-6203.25665.

[24] A. Bertoldo, G. Rizzo, and M. Veronese, “Deriving physiological information from PET images: From SUV to compartmental modelling,” *Clin. Transl. Imag.*, vol. 2, no. 3, pp. 239–251, Jun. 2014, doi: 10.1007/s40336-014-0067-x.

[25] W. A. Weber, “PET for response assessment in oncology: Radiotherapy and chemotherapy,” *BJR*, vol. 78, no. S28, pp. 42–49, Nov. 2005, doi: 10.1259/bjr/59640473.

[26] V. Kumar et al., “Variance of standardized uptake values for FDG-PET/CT greater in clinical practice than under ideal study settings,” *Clin. Nucl. Med.*, vol. 38, no. 3, pp. 175–182, Mar. 2013, doi: 10.1097/RLU.0b013e318279ffdf.

[27] A. Dimitrakopoulou-Strauss, L. Pan, and L. G. Strauss, “Quantitative approaches of dynamic FDG-PET and PET/CT studies (dPET/CT) for the evaluation of oncological patients,” *Cancer Imag.*, vol. 12, no. 1, pp. 283–289, Sep. 2012, doi: 10.1102/1470-7330.2012.0033.

[28] A. Dimitrakopoulou-Strauss, L. Pan, and C. Sachpekidis, “Kinetic modeling and parametric imaging with dynamic PET for oncological applications: General considerations, current clinical applications, and future perspectives,” *Eur. J. Nucl. Med. Mol. Imag.*, vol. 48, no. 1, pp. 21–39, Jan. 2021, doi: 10.1007/s00259-020-04843-6.

[29] M. Muzi et al., “Kinetic analysis of 3′-deoxy-3′-fluorothymidine PET studies: Validation studies in patients with lung cancer,” *J. Nucl. Med.*, vol. 46, no. 2, pp. 274–282, Feb. 2005.

[30] M. Muzi, D. A. Mankoff, J. R. Grierson, J. M. Wells, H. Vesselle, and K. A. Krohn, “Kinetic modeling of 3′-deoxy-3′-fluorothymidine in somatic tumors: Mathematical studies,” *J. Nucl. Med.*, vol. 46, no. 2, pp. 371–380, Feb. 2005.

[31] J. Loncarski, V. G. Monopoli, G. L. Cascella, and F. Cupertino, “SiC-MOSFET and Si-IGBT-based DC-DC interleaved converters for EV chargers: Approach for efficiency comparison with minimum switching losses based on complete parasitic modeling,” *Energies*, vol. 13, no. 17, p. 4585, Jan. 2020, doi: 10.3390/en13174585.

[32] E. Brescia, D. Costantino, F. Marzo, P. R. Massenio, G. L. Cascella, and D. Naso, “Automated multistep parameter identification of SPMSMs in large-scale applications using cloud computing resources,” *Sensors*, vol. 21, no. 14, p. 4699, 2021, doi: 10.3390/s21144699.

Open Access funding provided by ‘Università degli Studi di Roma “Tor Vergata”’ within the CRUI CARE Agreement