

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/339708598>

# Radiomics-net : analyse Deep-radiomics des images TEP FDG pour prédire la réponse au traitement du cancer

Article in *Médecine Nucléaire* · March 2020

DOI: 10.1016/j.mednuc.2020.01.127

CITATIONS

0

READS

99

4 authors, including:



Su Ruan

Université de Rouen

345 PUBLICATIONS 0 CITATIONS

SEE PROFILE



Isabelle Gardin

Centre Henri Becquerel

208 PUBLICATIONS 0 CITATIONS

SEE PROFILE



Romain Modzelewski

Centre Henri Becquerel

167 PUBLICATIONS 0 CITATIONS

SEE PROFILE

# 3D RPET-NET: DEVELOPMENT OF A 3D PET IMAGING CONVOLUTIONAL NEURAL NETWORK FOR RADIOMIC ANALYSIS

A. Amyar<sup>1,2</sup>, S. Ruan<sup>1</sup>, I. Gardin<sup>1,2</sup>, C. Chatelain<sup>1</sup>, R. Herault<sup>1</sup>, P. Decazes<sup>1,2</sup>, R. Modzelewski<sup>1,2</sup>

<sup>1</sup> LITIS - EA4108 - University of Rouen and INSA of Rouen, Rouen, France

<sup>2</sup> Nuclear Medicine Department, Henri Becquerel Center, Rouen, France

## ABSTRACT

Radiomics is now widely used to improve prediction of treatment response and patient's prognosis in oncology. In this work, we propose an end-to-end prediction model based on a 3D convolutional neural network (CNN), called 3D RPET-NET, which extracts 3D image features through four layers. Our model was evaluated for predicting the response to radio-chemotherapy in 97 patients with esophageal cancer from positron emission tomography (PET) images. The accuracy of the model was compared to five other methods proposed in the literature for PET images, based on 2D CNN and random forest algorithms. The role of the volume of interest on the accuracy of 3D RPET-NET was also evaluated using isotropic margin of 1, 2, 3 and 4 cm around the tumor volume. After segmentation of the lesion using a fixed threshold value of 40% of the maximum standard uptake value, the best accuracy of 3D RPET-NET reached 72% and outperformed the other methods tested. We also showed that using an isotropic margin of 2 cm around the tumor volume improved the performances of 3D RPET-NET to reach an accuracy of 75%.

**Index Terms**— Positron Emission Tomography, Machine Learning, Deep learning, Esophageal cancer

## 1. INTRODUCTION

The concept of radiomics is defined as the extraction of dozens of quantitative features from the image that could be incorporated in predictive models for patient management [1]. Many reports suggest that radiomic features extracted from baseline images can contribute to improve patient's prognosis and prediction of the treatment response in oncology [2]. Images can come from Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) [3], but also from Positron Emission Tomography (PET) using FluoroDeoxyGlucose (FDG), which is a radioactive glucose analog. The visualization of glucose metabolism of tumor cells provides additional information than anatomical imaging (CT or MRI). These so-called radiomic features are assumed to highlight some informative tissue characteristics, such as heterogeneity in glucose metabolic activity, necro-

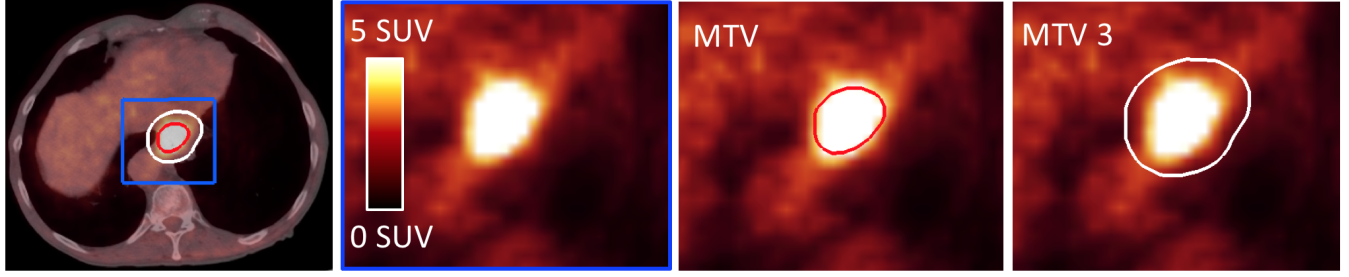
sis, etc. Numerous image features have been proposed in the literature [4] based on the shape and size of the lesion, 1st order statistics, textural features, filter and model-based features, potentially leading to several hundreds of image characteristics.

Several authors have used machine learning (ML) methods to build models for predicting treatment response or patient's survival based on radiomic features, such as random forests (RFs) and support vector machines (SVMs) associated or not to a feature selection strategy [5] [6] [7]. The main drawback of these approaches is that they require an initial extraction of radiomic features, while those with predictive value are not known.

Recently, a new paradigm in PET radiomic analysis has been proposed based on convolutional neural networks (CNNs) for predicting response to therapy [8]. It has been shown that deep learning architectures [9] can outperform traditional ML methods in classification tasks [10]. CNNs [11] [12] have proven to be very powerful tools in computer vision [13] for classifying images from different domains [14] [15] [16], and in the field of medical imaging [17]. In [18], Greenspan *et al* reported that deep convolutional neural networks have improved the performances of several applications in medical imaging. One of the advantages of CNN methods is that they do not require an initial calculation of features.

CNNs were not fully studied in radiomics, especially in PET imaging. Some papers have investigated baseline PET analysis based on 2 Dimensional (2D) CNN architectures [8] [19], but to our knowledge there is no work using 3D-CNN. These two first applications dealt with the prediction of the response to neoadjuvant chemotherapy in esophageal cancer [8] and the classification of mediastinal lymph node metastasis of Non-Small Cell Lung Cancer (NSCLC) [19].

In [8], Ypsilantis *et al*. proposed to learn a hierarchical representation directly from PET images in 107 patients with esophageal cancer using two CNN architectures. The first one, called 1S-CNN, corresponds to an architecture where the input is one slice. The process is repeated on each slice where the tumor is present. The spatial dependency between slices is not exploited in this architecture. For this reason, a second architecture was proposed where the input of the CNN



**Fig. 1.** From left to right: Fused PET/CT slice, zoomed on the esophageal tumor seen on FDG-PET only, MTV (40% SUVmax thresholding) in red and MTV3 (MTV + 3 cm isotropic margin) in white.

is composed by 3 adjacent slices, called 3S-CNN. For each exam containing  $m$  slices, each set of three spatially adjacent slices is taken as input leading to a total of  $m-2$  possible combinations. This 3S-CNN better exploits the spatial relationship between slices, but is limited to 3 slices. For both architectures, a post processing is required to predict the response based on a majority vote process. This study has shown the superiority of these two deep learning methods compared to other ML methods, such as RF, SVM, gradient boosting, and logistic regression.

In [19], Wang *et al.* used a centered axial slice and two others, separated by 4 mm in two image modalities (PET and CT) to obtain a limited number of six slices for each tumor to make a prediction. They compared the performances of their CNN and four other methods including RF, SVM, adaptive boosting, and artificial neural network. The methods were evaluated to discriminate against benign and malignant lymph nodes (1397) in 168 patients. The study showed that there were no significant difference between CNN and the best classical ML method for classifying mediastinal lymph node metastasis of NSCLC from PET/CT images. Nevertheless, Wang *et al.* concluded that CNNs are more convenient to use because the method does not require an initial feature extraction.

Our goal was to develop a new 3D-CNN architecture, called 3D RPET-NET, for predicting the response to treatment by learning from FDG-PET images of the tumor. Considering our small dataset, a two-layer 3D-CNN was proposed. Our study used a database of baseline FDG PET images of 97 patients treated by Radio-ChemoTherapy (RCT) for esophageal cancer. The optimal hyperparameters of 3D RPET-NET were found and the influence of the learning volume was investigated. The performances of the model were compared to 1S-CNN and 3S-CNN [8], as well as to three RF methods [5].

## 2. METHOD

### 2.1. Data

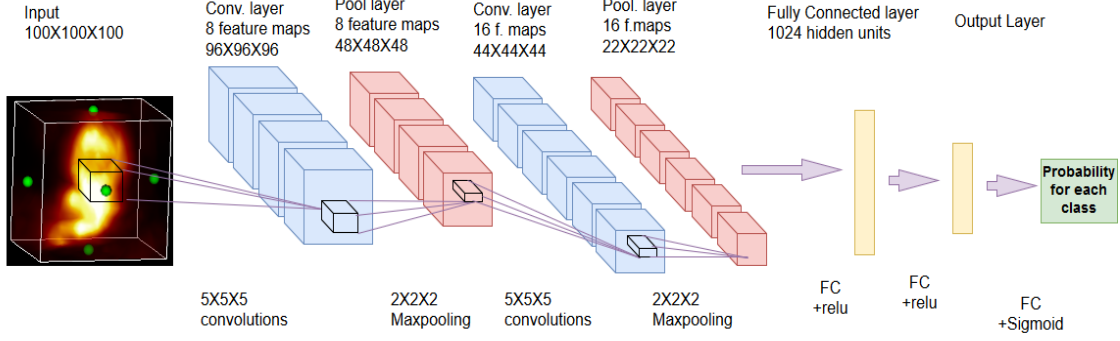
In this study, 97 patients with one lesion histologically proven to be locally advanced esophageal cancer and eligible for

RCT were included. All procedures performed in this study were conducted according to the principles expressed in the Declaration of Helsinki. The study was approved as a retrospective study by the Henri Becquerel Center Institutional Review Board (number 1506B). All patient information was de-identified and anonymized prior to analysis.

Patients underwent FDG PET with a CT before treatment (baseline PET), at the initial stage. They were treated by RCT, corresponding to an uninterrupted radiation therapy in the form of external radiation delivered by a 2-field technique of 2 Gy per fraction per day, 5 sessions per week, for a total of 50 Gy, as well as chemotherapy including platinum and 5-fluorouracil.

The FDG-PET/CT data were acquired on a Biograph<sup>®</sup> Sensation 16 Hi-Res device (Siemens Medical Solutions, IL, USA). Patients were required to fast for at least 6 hours before imaging. A total of 5 MBq/kg of FDG was injected after 20 min of rest. Sixty minutes later ( $\pm 10$  min), 6 to 8 bed positions per patient were acquired using a whole-body protocol (3 min per bed position). The PET images were reconstructed using Fourier rebinning and attenuation-weighted ordered subset expectation maximization algorithms. The images were corrected for random coincidences, scatter, and attenuation. Finally, the FDG-PET images were smoothed with a Gaussian filter (full width at half maximum = 5 mm). The reconstructed image voxel size was  $4.06 \times 4.06 \times 2.0$  mm<sup>3</sup>.

For the determination of treatment response, the response assessment included clinical examination, CT, FDG-PET, and esophagoscopy with biopsies performed 1 month after the end of treatment. Patients were classified as showing a clinically complete response (CR, 56 patients) to RCT if no residual tumor was detected on the endoscopy (negative biopsies) and if no locoregional or distant disease was identified on CT or PET evaluation. Patients were classified as showing a non-complete response (NCR, 41 patients) if a residual tumor or locoregional or distant disease was detected or if death occurred.



**Fig. 2.** 3D RPET-NET architecture composed by two 3D convolutional layers followed by 3D pooling layers and two dense layers.

## 2.2. Image preprocessing

Tumor images were spatially normalized by resampling all the dataset to an isotropic resolution of  $2 \times 2 \times 2 \text{ mm}^3$  using the k-nearest neighbor interpolation algorithm.

The Metabolic Tumor Volume (MTV) was segmented by a physician who manually defined a cuboid volume around the lesion and used a fixed threshold value of 40% of the maximum Standard Uptake Value (SUVmax) in the cuboid. In order to study the influence of the volume of interest on the performances of 3D RPET-NET, several isotropic margins of 1, 2, 3 and 4 cm around MTV were also applied, leading to define MTV1 to MTV4, respectively. In Fig. 1, are given an example of a PET/CT slice with two volumes of interest (MTV and MTV3).

Tumor gray level intensity were normalized to absolute SUV level between [0 30] and translated between [0 1] to be used in CNN architectures. The volumes of interest were included into a 3D empty cuboid of standard width, length and height of  $100^3$  voxels to learn tumoral radiomic features (see input part on Fig. 2) [8].

## 2.3. 3D RPET-NET architecture

We have developed a new CNN architecture based on two 3D convolutional layers and two fully connected layers, as shown in Fig. 2. Each convolutional layer, denoted  $C^{(m)}$ , consists of  $F^{(m)}$  feature maps, where  $m$  is the layer number (1 or 2). For the first layer,  $C^{(1)}$ , each feature map is obtained by convolving the volume of interest with a weight matrix  $W_i^{(1)}$  to which a bias term  $b_i^{(1)}$  is added, where  $i$  is the feature map number. Then, the output goes through a non linear function  $f(x)$ , where  $x$  is the input to a neuron, such as:

$$c_i^{(1)} = f(b_i^{(1)} + W_i^{(1)} * x) \quad \text{with } i = 1, \dots, F^{(1)}. \quad (1)$$

Each element of a feature map,  $c_i^{(1)}$ , is obtained by convolving the input  $x$  with a 3D kernel. A large receptive field tends to preserve better the relationship between slices and the local 3D tumor information, than a small one ( $5 \times 5 \times 5$ ) vs.

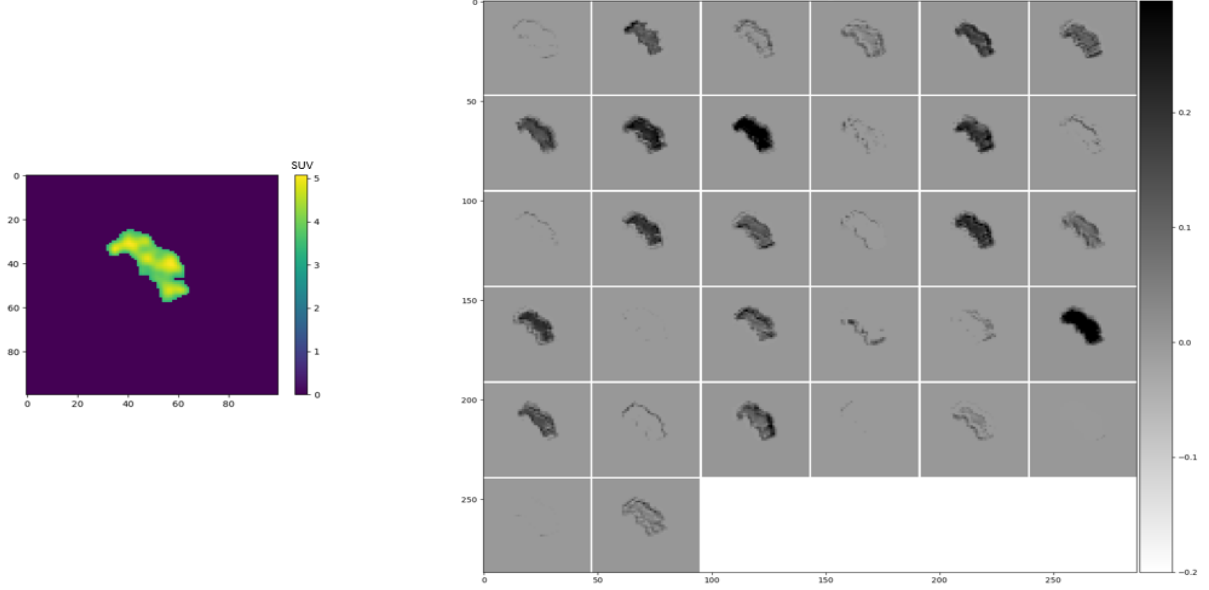
( $3 \times 3 \times 3$ ). The  $F^{(1)}$  weight matrices (one matrix per feature map) are learned by looking at different position of the input, leading to the extraction of the description of features. Thus, the weight parameters are shared for all tumor input sites, so that the layer has an equivariance property and is invariant to the input tumor transformations (such as translation and rotation). It also results in a sparse weight, which means that the kernel can detect small, but meaningful features, as shown in Fig. 3. For instance, it can be seen that some kernels are learning the tumor shape (e.g., [(1,1),(1,3),(2,4),(2,6)...etc]), while others tend to focus on features within the tumor (e.g., [(1,2),(1,5),(2,2),(2,3)...etc]).

Then, the output of this first convolutional layer is followed by a 3D pooling layer, to reduce the dimensionality of feature maps. The Max-pooling operator is used as a stage detector to report the maximum value within each cuboid of size ( $2 \times 2 \times 2$ ) for all feature maps. The purpose of this operation is to down-sample the feature maps by a factor of 2 along each direction (width, height, length), and to better generalize learning by selecting approximately invariant features. This invariance to local translation is very important in radiomics because tumors do not have a particular direction. The resulting feature maps are denoted  $P^{(m)}$ .

In order to extract high-level features from the low-level ones obtained in the initial layer, a second additional convolutional layer is added, followed by a pooling layer. This convolutional layer learns from the pooled feature maps of the first layer (see Fig. 2). Each feature map in the second layer is obtained as follow:

$$c_i^{(2)} = f(b_i^{(2)} + \sum_{s=1}^{P^{(1)}} W_i^{(2)} * P_s^{(1)}) \quad \text{with } i = 1, \dots, F^{(2)}. \quad (2)$$

The resulting set of the Max-pooled feature maps,  $P^{(2)}$ , encloses the entire spatial local information, as well as the hierarchical representation of the input 3D tumor. Then, each  $P^{(2)}$  feature map is flattened out, and all the elements are collected into a single vector  $V$  of dimension  $K$ , providing the input for a fully connected hidden layer, called  $h$ , consisting



**Fig. 3.** Visualization of a 2D slice of a tumor and the resulting 32 feature maps in the second convolutional layer of the 1S-CNN architecture.

of  $H$  units. The activation of the  $i^{th}$  unit of the  $h$  hidden layer is given by:

$$h_i = f(b_i + W h_i * V) \quad \text{with } i = 1, \dots, H. \quad (3)$$

All weights are collected into a matrix  $Wh$ .

At the end, the probability that the tumor is assigned to class 1 (responder) is given by the sigmoid function:

$$\text{Sig}(x) = \frac{1}{1 + e^{-x}} \quad (4)$$

The parameters of the CNN consist of all the convolutional weights  $W$ , and the weight matrix  $Wh$ , denoted by  $\theta$ . They are learned by minimizing the binary cross-entropy function:

$$L(\theta) = -\frac{1}{n} \sum_{i=1}^n [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (5)$$

which is a special case of the multinomial cross-entropy loss function for  $m = 2$ :

$$L(\theta) = -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m y_{ij} \log(\hat{y}_{ij}) \quad (6)$$

where  $n$  is the number of patients, and  $y$  is the tumor label (binary, 1 if the patient responded to treatment, 0 otherwise) and  $\hat{y}_{ij} \in (0, 1)$ :  $\sum_j \hat{y}_{ij} = 1 \forall i, j$  is the prediction response of a patient.

In our experiments, the adaptive gradient algorithm optimizer (AdaDelta) was used with mini batches. At each update of weights using the AdaDelta algorithm, only one mini batch of training data was used, which is changed for each gradient

calculation. Our CNN also incorporated L2 normalization of the weights and a dropout regularization of 50% to prevent the model from overfitting.

To find the best 3D RPET-NET, called 3D RPET-NETBest, the best hyperparameters were found. To this end, the influence of  $F$ , the number of 3D feature maps (8 to 64), the number of neurons (128 to 4096) was evaluated, as well as different receptive field sizes ( $3 \times 3 \times 3$ ,  $5 \times 5 \times 5$ ) and different sizes of mini-batch (2 to 16). Several expressions of  $f(x)$ , the activation function, were also evaluated (relu, elu, selu and tanh).

## 2.4. Implementation

The implementation of 3D RPET-NET was done using the keras library which is built on top of theano and tensorflow. We took advantage of graphical processing units (GPUs) to accelerate the algorithm. The CNNs training was performed on an NVIDIA Tesla 80 with 12 GO memory.

## 3. EXPERIMENTATION

Three experiments were performed to evaluate 3D RPET-NET.

**Experiment 1:** The first experiment consisted in tuning the optimal hyperparameters to find 3D RPET-NETBest on MTV.

**Experiment 2:** The second experiment consisted in comparing our architecture with 2 other CNN methods proposed in the literature: 1S-CNN and 3S-CNN [8]. The same tuning

process of 3D RPET-NET was done to find the best 1S-CNN and 3S-CNN hyperparameters.

The results were also compared to 3 RFs based methods: one without any feature selection strategy, called RF, and 2 other RF methods proposed in the literature based on a feature selection strategy, called GARF (for Genetic Algorithm based on Random Forest) and FIC (for Features Important Coefficient) methods. For the details of the methods refer to [20]. Briefly, 45 image features were extracted from PET images corresponding to first-order statistics (18), one feature of the lesion form, and textural features (26). Five hundred decision trees were built leading to the creation of the random forest classifiers.

**Experiment 3:** The third experiment consisted in assessing the influence of the volume of interest on the performances of 3D RPET-NETBest using MTV, MTV1, MTV2, MTV3 and MTV4.

For each experiment, 57 patients were used for the training, 20 for the validation and 20 for the test. For the test phase, a 5 fold cross-validation was performed.

The performances of methods were evaluated after each cross-validation process including sensitivity (Sens), specificity (Spec), accuracy (Acc), and area under the Receiver Operating Characteristic (ROC) curve (AUC), such as:

$$\text{Sens} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (7)$$

where TP is the number of true positives, representing correctly classified tumors and FN is the number of false negatives, representing misclassified tumors, for patients whose responded positively.

$$\text{Spec} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (8)$$

where TN is the number of true negatives, representing correctly classified tumors and FP is the number of false positives, representing misclassified tumors, for patients whose responded negatively.

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{TN} + \text{FP}} \quad (9)$$

## 4. RESULTS

In Table 1, are given the main results found during the 3 experiments (accuracy, sensibility, specificity, AUC of ROC curves).

**Experiment 1:** The best accuracy  $\text{Acc}=0.72$  and  $\text{AUC}=0.70$  were achieved as shown in Fig. 2, by two 3D convolutions layers and two 3D pooling layers, followed by two fully connected layers with the following hyperparameters for the first 3D convolutional layer: 8 3D feature maps with a filter size of  $5 \times 5 \times 5$  and a relu activation function. This operation is followed by 3D Max-pooling of size  $2 \times 2 \times 2$ . The second

3D convolutional layer corresponds to 16 3D feature maps of  $5 \times 5 \times 5$  convolutions, followed again by a  $2 \times 2 \times 2$  3D pooling layer. Then, the last two layers are composed of fully connected layers of respectively 1024 hidden neurons and finally 2 neurons for both classes.

From Experiment 1, are also given the results of two other models showing interesting performances, with no significant statistical difference with 3D RPET-NETBest. 3D RPET-NETBest and 3D RPET-NET1 differ by the activation function (relu vs. elu). 3D RPET-NETBest and 3D RPET-NET2 differ by the activation function (relu vs. elu) and the kernel size ( $(5 \times 5 \times 5)$  vs.  $(3 \times 3 \times 3)$ ).

**Experiment 2:** In Table 1 are also given the best results found with 1S-CNN, 3S-CNN, RF, GARF and FIC. ROC curves of Experiment 2 are presented in Figs. 4.a.

The best results were found with 3D PET-NETBest. 1S-CNN, seems to have lower performances ( $\text{Acc}=0.67 \pm 0.06$ ,  $\text{AUC}=0.67 \pm 0.06$ ), but the 1S-CNN ROC curve was not statistically significantly different from 3D RPET-NETBest ( $p=0.53$ ) and 3S-CNN ( $p=0.48$ ) ROC curves. From the RF classifiers, the best results were found with the GARF algorithm. GARF ROC curve was not statistically significantly different from 1S-CNN ( $p=0.10$ ) and 3S-CNN ( $p=0.058$ ) ROC curves, while the 3D RPET-NETBest ROC curve gave better results compared to GARF ROC curve ( $p=0.028$ ).

**Experiment 3:** Results of Experiment 3 are given in Table 1 and ROC curves are presented in Figs. 4.b. When studying the influence of the volume of interest, the best performances of 3D RPET-NETBest were obtained with MTV2 ( $\text{Acc}=0.75$  and  $\text{AUC}=0.74$ ). The performances of the 3D RPET-NETBest tend to increase from no margin to a margin of 2 cm, and then decrease with higher margins (MTV3 and MTV4). Only 3D RPET-NETBest performances on MTV2 were statistically better than those on MTV4 ( $p=0.04$ ).

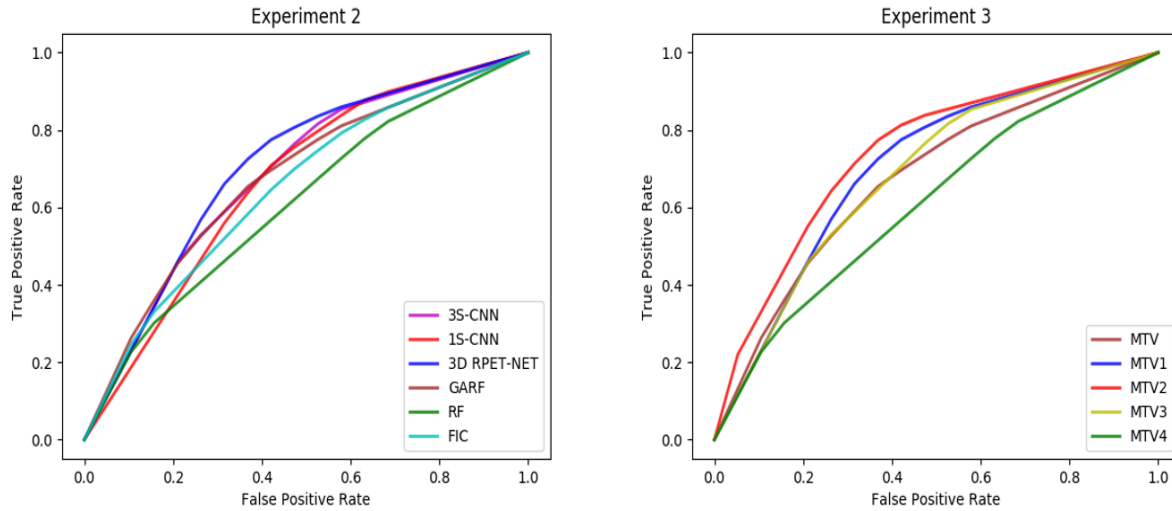
## 5. DISCUSSION

We have developed an end-to-end 3D convolutional neural network (3D PET-NET) based on TEP images. We have also evaluated 5 other methods from the literature [8] [5]. For each CNN, the search for the best architecture was achieved using a validation procedure, by tuning hyperparameters, such as the number of feature maps and the size of filters. For comparison, a radiomic analysis using 3 random forest classifiers: without feature selection (RF), with a selection strategy based on a genetic algorithm (GARF) and based on the forests importance coefficient (FIC).

Unlike Ypsilantis et al. in [8] who claimed that the use of a 3D ROI as direct input of the CNN is infeasible due to the fact that every tumor has a different shape and size, we show that englobing the tumor into a 3D cuboid of standard width, length and height allows to benefit from the spacial relationship between slices using a large 3D receptive field. Our assumption is that a neural network architecture able to

	Method	VOI	Acc	Sens	Spec	AUC
Experiment 1						
	<b>3D RPET-NETBest</b>	<b>MTV</b>	<b>0.72±0.08</b>	<b>0.79±0.17</b>	<b>0.62±0.21</b>	<b>0.70±0.04</b>
	3D RPET-NET1	MTV	0.71±0.10	0.65±0.26	0.75±0.23	0.69 ±0.08
	3D RPET-NET2	MTV	0.70±0.04	0.77±0.17	0.59±0.26	0.68±0.06
Experiment 2						
	<b>3D RPET-NET</b>	<b>MTV</b>	<b>0.72±0.08</b>	<b>0.79±0.17</b>	<b>0.62±0.21</b>	<b>0.70±0.04</b>
	1S-CNN	MTV	0.69±0.06	0.79±0.15	0.57±0.24	0.65±0.08
	3S-CNN	MTV	0.67±0.08	0.73±0.19	0.60±0.20	0.67±0.08
	GARF	MTV	0.68±0.08	0.80±0.11	0.46±0.09	0.62±0.04
	FIC	MTV	0.65±0.07	0.78±0.21	0.46±0.38	0.61 ±0.16
	RF	MTV	0.65±0.04	0.65±0.18	0.53 ±0.18	0.59±0.04
Experiment 3						
	3D RPET-NET	MTV	0.72±0.08	0.79±0.17	0.62±0.21	0.70±0.04
	3D RPET-NET	MTV1	0.73±0.04	0.76±0.07	0.69±0.1	0.72±0.04
	<b>3D RPET-NET</b>	<b>MTV2</b>	<b>0.75±0.03</b>	<b>0.76±0.45</b>	<b>0.74±0.15</b>	<b>0.74±0.02</b>
	3D RPET-NET	MTV3	0.72±0.09	0.71±0.09	0.74±0.14	0.72±0.09
	3D RPET-NET	MTV4	0.63±0.09	0.77±0.10	0.46±0.21	0.61±0.11

**Table 1.** Classification results: Each result corresponds to the average of five independent experiments using different training and test datasets and the hyperparameters used for.



**Fig. 4.** a. ROC curves comparing the 6 classifiers (RF, GARF, FIC, 1S-CNN, 3S-CNN and 3D RPET-NET) with the best parameters on MTV. b. ROC curves of 3D RPET-NETBest for different VOIs.



capture patterns of FDG uptake that occur within the whole lesion may detect imaging features that are more relevant for predicting treatment response than each slice individually or 3 adjacent slices. Under this assumption, we propose an architecture that initially fuses the spatial information across intra slices.

3D RPET-NETBest is composed of only 2 convolutional layers. A higher number of convolutional layers were tested, without conclusive results. The small number of patients in our database (without artificial data augmentation) is a limiting factor for the development of a deeper network, but also in radiomic analysis in general. Indeed, the current trend is in favour of the use of a network with an increasing number of convolutive layers (very deep neural network). This is only possible on very important image databases (e. g. ImageNet) that are not currently available in medical imaging.

It is possible to artificially increase the number of data. Since, learning takes place on a tumor inside a black box, this solution leads to an overfitting. Another solution is to provide complementary information by adding clinical data.

To ensure a fair comparison between the different methods, the database was splitted into 3 groups of 57 patients for the training, 20 for the validation and 20 for the test before any operation. Every CNN and RF classifiers used the same folds so as to obtain an exact comparison between methodology.

We have shown that isotropic dilation of MTV tends to increase the performances of RPET-NET 3D. When the margin around the MTV is too large ( $>2$  cm) the network performances decrease. When the MTV is increased by a margin which is too large, the volume of interest can include parts of metabolically active organs that are likely to interfere with the CNN analysis. Adding a peritumoral volume to the radiomic analysis has already been tested in MRI [21] [22] but never in PET imaging. These initial results must be confirmed on other types of cancer. The influence of the initial volume of interest and the segmentation method have to be continued.

## 6. CONCLUSION

3D RPET-NET shows very promising results in the prediction of treatment response in esophageal cancer, that outperformed 2D CNN architectures, as well as the RF classifiers tested. In addition, we have shown that the best volume to be used for PET radiomic prediction is the metabolic tumor volume with an isotopic margin of 2 cm. This peritumoral region seems to contain information that is potentially relevant for learning tumor features.

These results need to be confirmed on a larger database. The integration of clinical data in the model is an interesting and challenging perspective for such architectures that could improved the performances of the classifier.

## 7. REFERENCES

- [1] P Lambin et al., "Radiomics: extracting more information from medical images using advanced feature analysis," *European journal of cancer*, vol. 48, no. 4, pp. 441–446, 2012.
- [2] M Avanzo et al., "Beyond imaging: The promise of radiomics," *Physica Medica: European Journal of Medical Physics*, vol. 38, pp. 122–139, 2017.
- [3] T Nishioka et al., "Image fusion between 18fdg-pet and mri/ct for radiotherapy planning of oropharyngeal and nasopharyngeal carcinomas," *International Journal of Radiation Oncology Biology Physics*, vol. 53, no. 4, pp. 1051–1057, 2002.
- [4] M Sollini et al., "Pet radiomics in nslc: state of the art and a proposal for harmonization of methodology," *Scientific reports*, vol. 7, no. 1, pp. 358, 2017.
- [5] P Desbordes et al., "Predictive value of initial fdg-pet features for treatment response and survival in esophageal cancer patients treated with chemo-radiation therapy using a random forest classifier," *PloS one*, vol. 12, no. 3, pp. e0173208, 2017.
- [6] C Parmar et al., "Machine learning methods for quantitative radiomic biomarkers," *Scientific reports*, vol. 5, pp. 13087, 2015.
- [7] S Leger et al., "A comparative study of machine learning methods for time-to-event survival data for radiomics risk modelling," *Scientific reports*, vol. 7, no. 1, pp. 13206, 2017.
- [8] P Ypsilantis et al., "Predicting Response to Neoadjuvant Chemotherapy with PET Imaging Using Convolutional Neural Networks," *PLoS One*, vol. 10, no. 9, pp. e0137036, sep 2015.
- [9] G Hinton et al., "Reducing the dimensionality of data with neural networks," *science*, vol. 313, no. 5786, pp. 504–507, 2006.
- [10] A Krizhevsky et al., "Imagenet classification with deep convolutional neural networks," 2012, pp. 1097–1105.
- [11] K Fukushima, "Neocognitron: A hierarchical neural network capable of visual pattern recognition," *Neural networks*, vol. 1, no. 2, pp. 119–130, 1988.
- [12] Y LeCun et al., "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2323, 1998.
- [13] A Karpathy et al., "Large-scale video classification with convolutional neural networks," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2014, pp. 1725–1732.



- [14] S Chen et al., “Target classification using the deep convolutional networks for sar images,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 8, pp. 4806–4817, 2016.
- [15] P Sermanet et al., “Convolutional neural networks applied to house numbers digit classification,” in *Pattern Recognition (ICPR), 2012 21st International Conference on*. IEEE, 2012, pp. 3288–3291.
- [16] D CireşAn et al., “Multi-column deep neural network for traffic sign classification,” *Neural networks*, vol. 32, pp. 333–338, 2012.
- [17] P Rajpurkar et al., “Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning,” *arXiv preprint arXiv:1711.05225*, 2017.
- [18] H Greenspan et al., “Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1153–1159, 2016.
- [19] H Wang et al., “Comparison of machine learning methods for classifying mediastinal lymph node metastasis of non-small cell lung cancer from 18F-FDG PET/CT images,” *EJNMMI Res.*, vol. 7, no. 1, pp. 11, 2017.
- [20] P Desbordes et al., “Feature selection for outcome prediction in oesophageal cancer using genetic algorithm and random forest classifier,” *Computerized Medical Imaging and Graphics*, vol. 60, pp. 42–49, 2017.
- [21] P Prasanna et al., “Radiomic features from the peritumoral brain parenchyma on treatment-naive multiparametric MR imaging predict long versus short-term survival in glioblastoma multiforme: Preliminary findings,” *European radiology*, vol. 27, no. 10, pp. 4188–4197, oct 2017.
- [22] N Bramanet al., “Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI,” *Breast cancer research : BCR*, vol. 19, no. 1, pp. 57, 2017.