

The Impact of Artificial Intelligence CNN Based Denoising on FDG PET Radiomics

Cyril Jaudet, Kathleen Weyts, Alexis Lechervy, Alain Batalla, Stéphane Bardet, Aurélien Corroyer-Dulmont

▶ **To cite this version:**

Cyril Jaudet, Kathleen Weyts, Alexis Lechervy, Alain Batalla, Stéphane Bardet, et al.. The Impact of Artificial Intelligence CNN Based Denoising on FDG PET Radiomics. Frontiers in Oncology, 2021, 11, pp.692973. 10.3389/fonc.2021.692973 . hal-03335315

HAL Id: hal-03335315

https://normandie-univ.hal.science/hal-03335315

Submitted on 6 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

frontiers in Oncology

ORIGINAL RESEARCH
published: 24 August 2021
doi: 10.3389/fonc.2021.692973

The Impact of Artificial Intelligence CNN Based Denoising on FDG PET Radiomics

Cyril Jaudet[1], Kathleen Weyts[2], Alexis Lechervy[3], Alain Batalla[1], Stéphane Bardet[2] and Aurélien Corroyer-Dulmont[1,4]**

[1] Medical Physics Department, CLCC François Baclesse, Caen, France, [2] Nuclear Medicine Department, CLCC François Baclesse, Caen, France, [3] UMR GREYC, Normandie Univ, UNICAEN, ENSICAEN, CNRS, Caen, France, [4] Normandie Univ, UNICAEN, CEA, CNRS, ISTCT/CERVOxy group, GIP CYCERON, Caen, France

Background: With a constantly increasing number of diagnostic images performed each year, Artificial Intelligence (AI) denoising methods offer an opportunity to respond to the growing demand. However, it may affect information in the image in an unknown manner. This study quantifies the effect of AI-based denoising on FDG PET textural information in comparison to a convolution with a standard gaussian postfilter (EARL1).

Methods: The study was carried out on 113 patients who underwent a digital FDG PET/CT (VEREOS, Philips Healthcare). 101 FDG avid lesions were segmented semi-automatically by a nuclear medicine physician. VOIs in the liver and lung as reference organs were contoured. PET textural features were extracted with pyradiomics. Texture features from AI denoised and EARL1 *versus* original PET images were compared with a Concordance Correlation Coefficient (CCC). Features with CCC values ≥ 0.85 threshold were considered concordant. Scatter plots of variable pairs with R2 coefficients of the more relevant features were computed. A Wilcoxon signed rank test to compare the absolute values between AI denoised and original images was performed.

Results: The ratio of concordant features was 90/104 (86.5%) in AI denoised *versus* 46/104 (44.2%) with EARL1 denoising. In the reference organs, the concordant ratio for AI and EARL1 denoised images was low, respectively 12/104 (11.5%) and 7/104 (6.7%) in the liver, 26/104 (25%) and 24/104 (23.1%) in the lung. SUVpeak was stable after the application of both algorithms in comparison to SUVmax. Scatter plots of variable pairs showed that AI filtering affected more lower *versus* high intensity regions unlike EARL1 gaussian post filters, affecting both in a similar way. In lesions, the majority of texture features 79/100 (79%) were significantly ($p<0.05$) different between AI denoised and original PET images.

Conclusions: Applying an AI-based denoising on FDG PET images maintains most of the lesion's texture information in contrast to EARL1-compatible Gaussian filter. Predictive features of a trained model could be thus the same, however with an adapted threshold. Artificial intelligence based denoising in PET is a very promising approach as it adapts the denoising in function of the tissue type, preserving information where it should.

Keywords: denoising, AI, PET, radiomics, medical imaging, convolutional neural network, VEREOS

OPEN ACCESS

Edited by:
Laurent Dercle, Columbia University Irving Medical Center, United States

Reviewed by:
Suyash P. Awate, Indian Institute of Technology Bombay, India
Ziren Kong, Chinese Academy of Medical Sciences, China

****Correspondence:***
Cyril Jaudet
c.jaudet@baclesse.unicancer.fr
Aurélien Corroyer-Dulmont
a.corroyerdulmont@baclesse.unicancer.fr

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *09 April 2021*
Accepted: *26 July 2021*
Published: *24 August 2021*

Citation:
Jaudet C, Weyts K, Lechervy A, Batalla A, Bardet S and Corroyer-Dulmont A (2021) The Impact of Artificial Intelligence CNN Based Denoising on FDG PET Radiomics. Front. Oncol. 11:692973. doi: 10.3389/fonc.2021.692973

INTRODUCTION

Imaging modalities are nowadays an essential diagnostic tool in medicine. From 2009 to 2019 the number of exams in the USA has increased by about 18%, 42% and 105% for CT, MRI and PET respectively (1). This increasing demand has exceeded the actual offer leading to unreasonable delay, weeks or even months for MRI and PET scans in France/Europe (2). An appropriate image denoising may help to reduce scanning time or even reduce injected dose for PET. It may allow to increase the number of examinations without impacting too much working hours or requiring the installation of new medical imaging devices. Deep learning as a subdivision of artificial intelligence (AI) allows to build promising denoising models.

We focused on PET imaging as it will benefit of denoising because of its long scanning time. Although many studies are actually investigating the clinical performance of this method, it may also impact other emerging fields such as imaging based predictive models, radiomics and other AI applications (3).

Medical images are basically a visual representation of different grey levels based on their density (CT), magnetic properties (MRI) or functional information (PET/SPECT). The distribution of the grey values characterizes the heterogeneity of the information. A fast-evolving field called radiomics provide a methodology to extract different features based on intensity, shape, texture from images in order to build predictive models (4). This approach holds great promises as being able to predict patient outcomes. They might allow personalized treatment. As an example, an overall survival predictive model including radiomics features was computed in lung cancer (5) This field is increasing with an annual growth rate of published papers of 177.82% between 2013 and 2018 (6). The models are very promising but there are still some efforts to be made to translate and implement them in a routine clinical setting (7).

Artificial intelligence is in the early phase of application in medical imaging. In this article, we used deep learning and more specifically convolutional neural network approaches which represent a subdivision of AI techniques. Today deep learning has a key role in image reconstruction, processing (denoising, segmentation), analysis and predictive modelling. These applications will develop even more in the future (8). In most of these tasks, they often outperformed a more traditional approach (9). A comparison of this type of AI based denoising algorithm on a PET/MR with clinical data show an increase of the contrast over noise ratio by $46.80 \pm 25.23\%$ compared to $18.16 \pm 10.02\%$ for a Gaussian filter only (10)]. Other methods studied in (10) like guided nonlocal means, block matching 4D or deep decoder improve the CNR oby$24.35 \pm 16.30\%$, $38.31 \pm 20.26\%$ and $41.67 \pm 22.28\%$ respectively. Denoising may also be performed during reconstruction, however this cannot be implemented on an existing machine. The most important limitation is the lack of FDA or CE certification of all those approaches. We focus our study on Subtle PET™ (Subtle Medical, Stanford, USA provided by Incepto, France). It is a post-processing FDA and CE approved denoising software for FDG PET (11), based on convolutional neural networks (CNN), the most common deep learning architecture for image processing.

AI denoising and radiomics are two very promising fields in medical imaging. However, we are the first, to the best of our knowledge, to try to combine these two approaches for PET Imaging. We question whether a radiomics model using PET [^{18}F] FDG trained on classical data is still valid after an AI denoising method. This study measured the stability of basic and radiomics PET features in lesions and normal reference organs when applying an AI denoising solution. We also wanted to provide an intuitive understanding on how images are affected by AI compared to a reference gaussian post filter routinely used in our center to generate EARL1 compatible PET series.

MATERIALS AND METHODS

This retrospective study was approved by the local institution review board. 113 patients referred to our oncological institution for an initial or follow-up [^{18}F] FDG PET/CT exam between January and March 2020 were retrospectively included. We obtained an informed consent (non-opposition) from all patients. This observational study was in line with MR 004, a national French institution (INDS) defining health research conduct guidelines. The study population characteristics are shown in **Table 1**.

Our PET center is accredited by EANM research limited (EARL) (12) and EANM imaging guidelines (13) were respected. The patients were injected with 4MBq/kg of [^{18}F] FDG IV. PET images from skull base to mid-thighs were acquired on a digital

TABLE 1 | Description of the patient cohort.

Patients (N)		**133**	**Number**
Sex		68%	female%
Age(Y)		61.5±13.5	mean±SD
		[24-89]	[range]
Weight (kg)		74±16	mean±SD
		[35-110]	[range]
BMI(kg/m2)		27±6	mean±SD
		[15-42]	[range]
Indication			
Oncologic	95 (84%)		Number (%)
	Breast	36 (32%)	
	Lung	17 (15%)	
	Gynaecologic (except breast)	14 (12%)	
	other malignancies (lymphoma, anal, colorectal, bladder, thyroid, head and neck, melanoma, myeloma) or mixed	28 (25%)	
Diagnostic benign versus malignant	14 (12%)		
Miscellaneous	5 (4%)		

PET/CT (VEREOS 2018, Philips Healthcare) during 1min/bed position. Once acquired, PET images were reconstructed with an 3D OSEM algorithm, 4 iterations, 4 subsets with point spread function (PSF) correction. Scatter and attenuation correction was applied. The spacing and matrix size were respectively of 2x2x2 mm^3 and 288x288 pixels. An EARL1 reconstruction was also generated with the same parameters but convolved with a gaussian post filter of 7.2mm. CT scan parameters were 100-140 kV (BMI adaptive), with variable mAs according to an index dose right of 14 and an iterative reconstruction I dose 4; 64x0.625mm slice collimation, pitch of 0.83, rotation time 0.5 s, 3D modulation, matrix 512x512 and voxel size 0.97x0.97x 3 mm^3. The PET mean dose was 5.32 mSv for a patient of 70 Kg. CT had a CTDI median value of 4.8 mGy and a DLP of 431.5 mGy.cm.

The originally reconstructed PET images (with PSF modelling) were denoised with a convolutional neural network (CNN) approach by a commercially available software, Subtle PET® by Subtle Medical. SubtlePET™ uses a multi-slice 2.5D (5 slices) encoder-decoder U-Net DCNN to perform denoising. The software takes a low count PET image as the input and generates a high-quality PET image (close to full dose image) as the output. Accreditation from FDA and CE required robustness. The denoising model was trained on PET images from different centers and vendors. It employs a CNN-based method in a pixel's neighborhood to reduce noise and increase image quality. Using a residual learning approach and optimized for quantitative (L1 norm) as well as structural similarity (SSIM), the software learns to separate and suppress the noise components while preserving and enhancing non-noise components. The images were directly sent from the PET console to a specific local server. Once transferred they were anonymized, denoised, deanonymized and pushed back to a clinical viewer. The mean treatment time was 45 s on a NVIDIA 1080 GPU processor.

All contours were performed in 3D slicer version 4.10 (14) on original PSF PET images and copied on AI denoised and EARL1 PET series. Spherical volumes of interest (VOI) were drawn in the reference organs: liver (3 cm radius, avoiding upper parts, tissue boundaries and major vessels) and lung (1.5 cm radius, drawn in the upper parts). Up to five FDG avid lesions per patient (including only the most intense ones), in total 101 lesions, were segmented by an experienced nuclear medicine physician. Segmented lesions consisted only of authentic malignant primary and metastatic lesions in solid tumors or lymphoma. A semi-automatic tool was employed to segment lesions. A VOI was created by clicking on the original PET image. This 3DSlicer module (PETTumors Segmentation) is based on a highly automated optimal surface segmentation approach, which is a variant of the layered optimal graph image segmentation of multiple objects and surfaces segmentation (15). The VOI was than inspected and manually adjusted with a brush if needed. An automatic donut of 2 voxels diameters was grown around the lesion to calculate the lesion over background ratio. The mean analyzed metabolic volume was 20 (1-162) ml. The same VOI were used for original, AI denoised and EARL1like images.

The extraction of radiomics features was automatically carried out with the pyradiomics package (16) thus mostly complying with the Image Biomarker Standardisation Initiative (17). Images had a native isotropic spacing of 2x2x2 mm^3 so an interpolation step was not necessary. As there is no consensus about the intensity discretization, a fixed bin number of 64 was used (18). A python code using simpleITK (19) was developed to extract all the radiomics features and is accessible in the supplementary information. Eight groups of radiomics features were computed. The intensity class contains first-order data, describing the distribution of voxel intensities within the image region defined by the VOI. They are commonly used and basic images metrics. The shape class is constituted of the 3D size and shape of the VOI. These shape features were excluded as the VOI was the same in all the images. A Grey Level Co-occurrence Matrix (GLCM) class describes the second-order joint probability function of an image region. Grey Level Size Zone Matrix (GLSZM) features quantify grey level zones in an image. A grey level zone is defined as the number of connected voxels that share the same grey level intensity (3D). The Grey Level Run Length Matrix (GLRLM) class testifies of grey level runs, which are defined as the length in number of pixels, of consecutive pixels that have the same grey level value (1D). Neighboring Grey Tone Difference Matrix (NGTDM) is a descriptor of the difference between a grey value and the average grey value of neighbors. A Grey Level Dependence Matrix (GLDM) characterizes the number of connected voxels within a distance from the center voxel in function of their grey level. Most features used in this study are in compliance with Imaging Biomarker Standardization Initiative (IBSI) (IBSI reference manual).IQ wavelets class contains two features, a local analyzing just the VOI and a global of the whole image. These metrics characterize image quality as the ratio between high and low wavelet frequencies.

The Concordance Correlation Coefficients (CCC) (20) were evaluated comparing the post processing IA denoised and EARL1 images to the original PET. CCC values of +/-1 describe a perfect positive/negative correlation and 0 no correlation. Features with a minimum CCC of 0.85 were considered as statistically reproducible and concordant (21). Scatter plots of variable pairs with R^2 value was displayed for the coefficient of variation (CV) and mean SUV values to understand the difference of CCC's in lesions and in liver when an AI denoising or EARL1 filter are used on original images. Mean SUV in lesions is presented using boxplots with minimum, maximum, 1st quartile and 3rd quartile to highlight the difference between the 3 series. A paired Wilcoxon signed rank test was used to compare features in original and AI denoised, and original and EARL1 images. P-values <0.05 were considered statistically significant. All the statistical analyzes were performed using python (22) and scipy.stats library. All the data and the python code of the analysis are available on https://github.com/AurelienCD/RadiomicsIA_PET_Depository_Manuscript-ID-692973.

RESULTS

A visual comparison of AI denoised (B) *versus* original images (A) shows that the AI approach seems to decrease noise in healthy tissues while preserving the intensity distribution in the

lesion in **Figure 1**. In the EARL1-PET image (C) background noise is reduced, but also in the lesion the uptake intensity and distribution are affected. Similar observations can be deduced from the second patient's images (**Figures 1D–F**).

The concordance correlation coefficient (CCC) testifying of the stability of the features comparing denoised to original images is presented in **Figure 2**. In lesions, 90/104 (86.5%) with AI and 46/104 (44%) with EARL1 denoising stayed stable. All stable features in the EARL1 images were also stable in AI images. For the basic intensity class parameters, SUVpeak, SUVmean and SUVmedian kept a CCC≥0.85 in the two denoising approaches. SUVmax and SUVmin CCC values stayed stable for the AI denoised images in the lesions, but fell below the significant threshold for EARL1 images. The NGTDM features were less affected by both denoising methods. In the reference organs, for AI and EARL1 respectively, 12/104 (11.5%) and 7/104 (6.7%) in liver and 26/104 (25%) and 24/104 (23.1%) in lung had a CCC value at least of 0.85. The majority of the features in reference organs are less stable then in lesions for the two denoising methods. For the basic intensity parameters, SUV mean was overall stable for both denoising methods while SUV peak in both liver and lung for AI denoising, *versus* only in the lung for EARL1.

Concerning AI denoising, CV values in lesions before and after processing were very similar. In the lesions the values were slightly below and parallel to the identity line with R^2 = 0.992. EARL1 showed a lower correlation and greater distance from the identity line(**Figure 3A**). In healthy liver (**Figure 3B**), the behavior was different. CV was reduced by a magnitude order of 2 for both denoising methods. With IA denoising, the points were also more scattered for liver (R^2 = 0.884) than for lesions (R^2 = 0.992). EARL1 denoising showed less differences (R^2 = 0.851 *vs* 0.893). The SUV mean value displayed in **Figures 3C, D** showed high correlation in lesions as well as in the healthy tissue. In **Figure 3D** SUV mean in liver is not modified by a EARL1 gaussian postfilter (R^2 = 1). Scatter plots of variable pairs for all the features are accessible in the supplementary materials.

Figure 4 testifies of the difference of SUV mean in lesions between AI and EARL1 denoising compared to the original images. AI denoising not significantly modified the SUV mean values with a p=0.06. EARL1 post filter led to a significantly lower mean SUV in lesions (p<0.001).

The results of the paired Wilcoxon signed rank test between original and AI denoised images are presented in **Table 2**. Almost all the features 79/100 (79%) were significantly different. Wavelets features were not studied. In the intensity class only 4/27 were not significantly different. SUV mean and median values were not significantly different between the AI denoised image and the original one. **Table 2** shows in blue the 18 features that had a CCC>0.85 and were not significantly different.

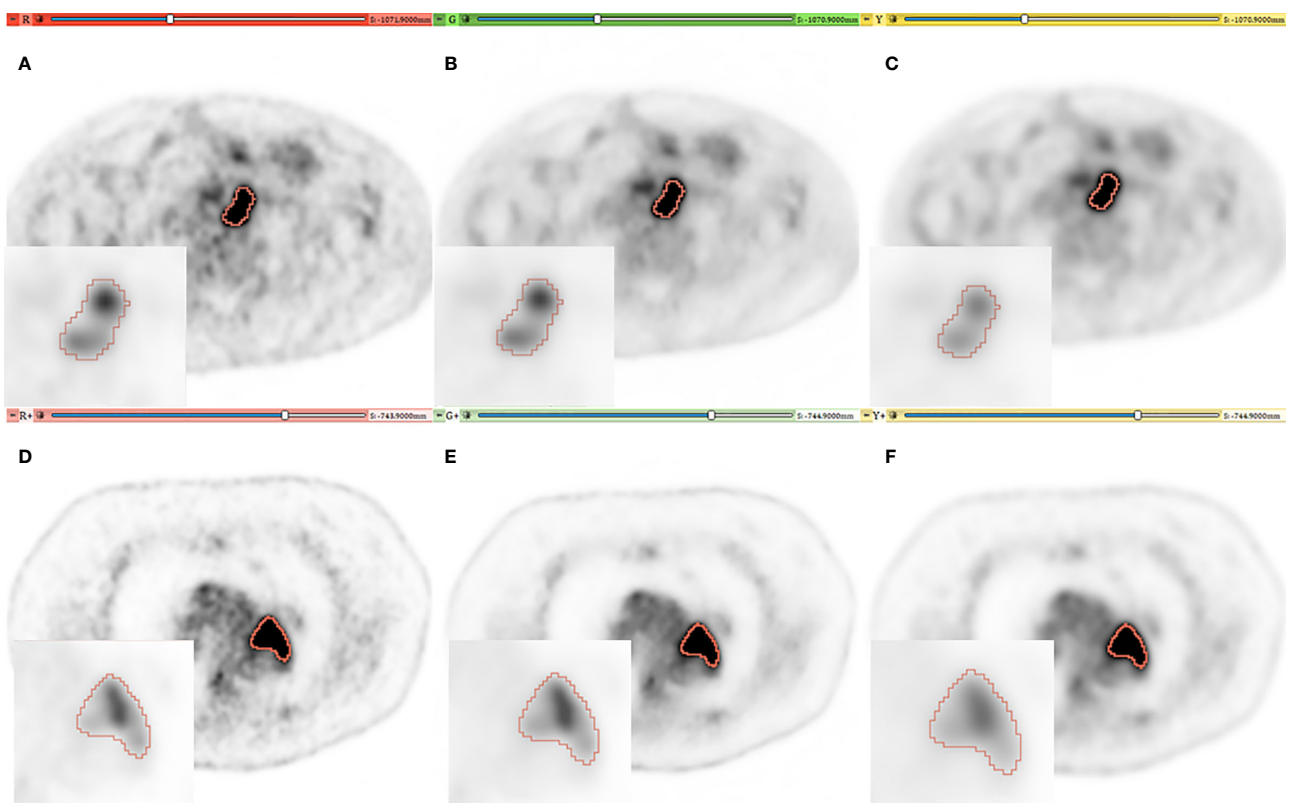


FIGURE 1 | Representative PET imaging of two lesions in different patients with a SUV windowing of (0–5). **(A, D)** [red], **(B, E)** [green], **(C, F)** [yellow] for original, AI and EARL images respectively. A Zoom is added on each image with SUV windowing between (0–25) and (0–30) for the first and second patient.

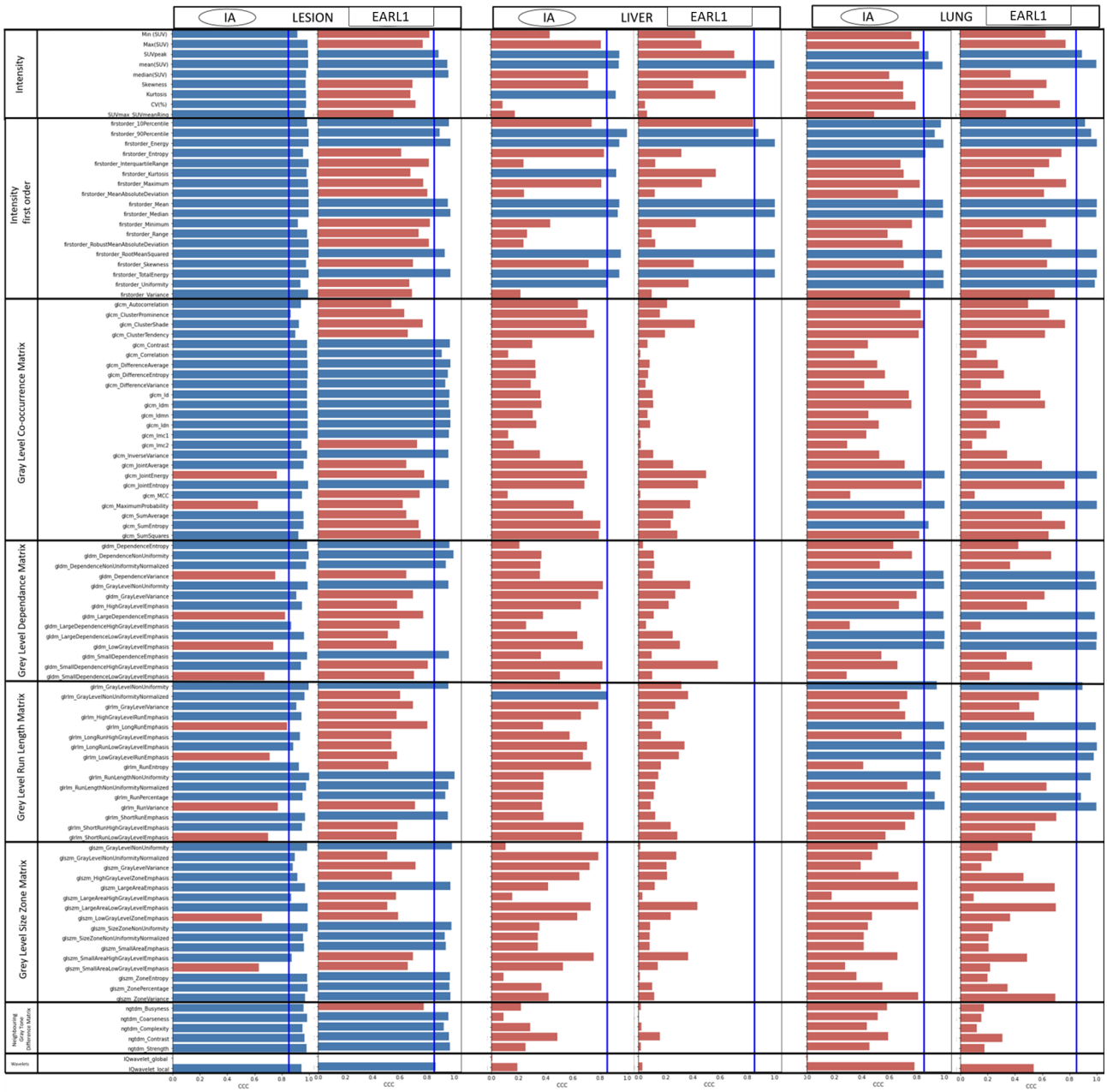


FIGURE 2 | CCC of all the features from AI and EARL *versus* original images. A threshold is display by a line with a CCC = 0.85. Blue bar indicates a CCC≥0.85, red bar CCC < 0.85.

DISCUSSION

We evaluated the impact of AI denoising on the stability of radiomics features computed in FDG PET images, the standard being the clinical images. We also concurrently evaluated the effect of EARL1 gaussian filtering. To the best of our knowledge, it is the first clinical study on the impact of artificial intelligence denoising on PET radiomics.

Texture features used in radiomics models describe the pattern distribution of voxels and quantify intra-tumor heterogeneity in all 3 dimensions (4). 86.5% showed a stable behavior for intensity and radiomics classes. The stability criterium was based on a CCC≥0.85 (20). In lesions, values were significantly different in 71.1% of the features after AI denoising. An AI denoising approach like CNN seems to change the absolute values of most of the features but keep the correlation between them.

Advanced applications aim at the correlation of image features, like radiomics, with clinical endpoints (4, 23). Radiomics models derived from CT correlated with a prognostic value, overall survival in lung cancer patient (5). In baseline PET of locally advanced rectal

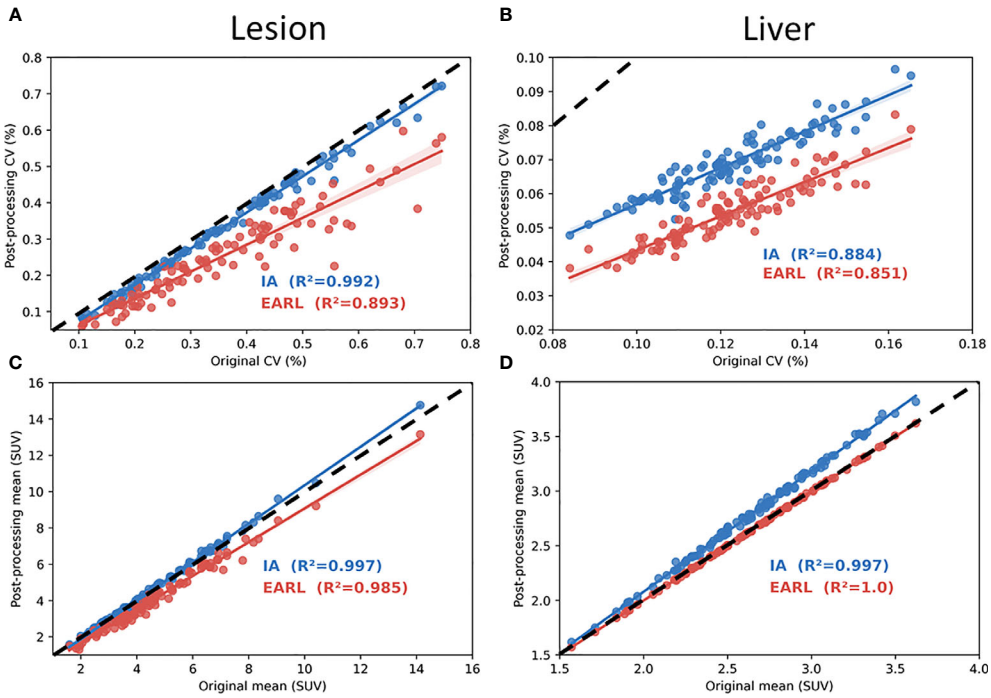


FIGURE 3 | Coefficient of correlation plot with R^2 value in lesions **(A, C)** and healthy liver **(B, D)**. **(A–D)** show respectively coefficient of variation (CV) and mean SUV calculated from AI and EARL1-like image in function of original images. Dotted line represents the identity line.

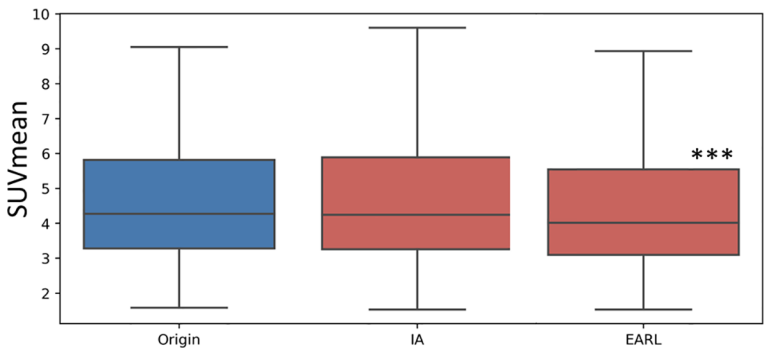


FIGURE 4 | Box plot of the mean SUV value in lesion in Original, AI and EARL1-like images. The distribution difference between the original images and AI is not significant (p=0.06) while it is significant (p<0.001 ***) between original and EARL1-like images.

cancer ^{18}F-FDG PET/CT texture features provide strong independent predictors of survival in patients (24). These models are very promising however there are several pitfalls to overcome (25) such as study design, data acquisition, segmentation, features calculation and modeling by the radiomics community. This study allows a better understanding of the behavior of predictive models when an AI denoising method is employed. A predictive model based on this type of information can be built from MRI, PET, CT or a combination of image modalities.

Deep learning AI techniques have been used to perform denoising on PET images for example by generating a full-dose PET images from low-dose images (26) or to directly filter reconstructed PET images (27). We used an AI denoising approach based on DCNN (11). This approach seems to be able to reduce the acquisition time activity product by a factor of 2 to 4. We used it directly on the studied PET image without activity or time reduction because we want to characterize the effect of AI denoising while not compensating for count losses.

Denoising will be more and more used but may also generate pitfalls to build a radiomics predictive model as the 3d texture information may be modified. Studying the stability of features with a test-retest approach has been performed in PET (28). The number

TABLE 2 | Result of Wilcoxon signed rank test of all the features between AI denoised and original images in lesions.

Classes	Features	P value	Median [min; max]	Classes	Features	P value	Median [min; max]
Intensity	SUVmin	<0.001 (***)	0.25 [-0.13 ; 1.09]	Gray Level Size Zone Matrix	glszm_GrayLevelNonUniformity	<0.001 (***)	-0.48 [-24.09 ; 1.34]
	SUVmax	<0.001 (***)	-0.53 [-4.52 ; 0.0]		glszm_GrayLevelNonUniformityNormalized	<0.001 (***)	-0.0 [-0.01 ; 0.01]
	SUVpeak (SUV)	<0.001 (***)	-0.15 [-1.3 ; 0.91]		glszm_GrayLevelVariance	<0.001 (***)	15.79 [-21.5 ; 60.88]
	SUVmean	0.13	0.03 [-0.34 ; 0.64]		glszm_HighGrayLevelZoneEmphasis	0.11	13.64 [-132.09 ; 206.28]
	SUVmedian	0.17	-0.04 [-0.75 ; 1.18]		glszm_LargeAreaEmphasis	0.11	0.02 [-95.71 ; 1321.6]
	Skewness	<0.01 (**)	-0.02 [-0.34 ; 0.21]		glszm_LargeAreaHighGrayLevelEmphasis	<0.001 (***)	162.72 [-1058.51 ; 112710.32]
	Kurtosis	<0.001 (***)	-0.13 [-1.46 ; 0.15]		glszm_LargeAreaLowGrayLevelEmphasis	<0.001 (***)	0.01 [-16.91 ; 46.07]
	CV(%)	<0.001 (***)	-0.02 [-0.09 ; -0.0]		glszm_LowGrayLevelZoneEmphasis	<0.001 (***)	0.0 [-0.03 ; 0.03]
	SUVmax/SUVmeanRing	<0.001 (***)	-0.55 [-3.74 ; -0.18]		glszm_SizeZoneNonUniformity	0.84	1.69 [-182.77 ; 68.57]
	firstorder_10Percentile	<0.001 (***)	0.1 [-0.15 ; 0.58]		glszm_SizeZoneNonUniformityNormalized	<0.05 (*)	0.0 [-0.1 ; 0.13]
	firstorder_90Percentile	<0.01 (**)	-0.11 [-0.67 ; 0.83]		glszm_SmallAreaEmphasis	<0.05 (*)	0.0 [-0.08 ; 0.07]
	firstorder_Energy	<0.01 (**)	-45.92 [-3961.59 ; 133754.46]		glszm_SmallAreaHighGrayLevelEmphasis	0.4	8.44 [-159.75 ; 258.22]
	firstorder_Entropy	<0.001 (***)	0.04 [-0.19 ; 0.23]		glszm_SmallAreaLowGrayLevelEmphasis	<0.001 (***)	0.0 [-0.03 ; 0.03]
	firstorder_InterquartileRange	<0.001 (***)	-0.08 [-0.43 ; 0.25]		glszm_ZoneEntropy	0.27	0.0 [-0.26 ; 0.18]
	firstorder_Kurtosis	<0.001 (***)	-0.13 [-1.47 ; 0.15]		glszm_ZonePercentage	0.55	0.0 [-0.07 ; 0.08]
	firstorder_Maximum	<0.001 (***)	-0.53 [-4.52 ; 0.0]		glszm_ZoneVariance	<0.05 (*)	0.04 [-95.59 ; 1276.34]
	firstorder_MeanAbsoluteDeviation	<0.001 (***)	-0.07 [-0.35 ; 0.18]	Gray Level Run Length Matrix	glrlm_GrayLevelNonUniformity	<0.001 (***)	-0.94 [-56.51 ; 4.11]
	firstorder_Mean	0.13	0.03 [-0.34 ; 0.64]		glrlm_GrayLevelNonUniformityNormalized	<0.001 (***)	-0.0 [-0.01 ; 0.0]
	firstorder_Median	<0.01 (**)	0.03 [-0.3 ; 0.71]		glrlm_GrayLevelVariance	<0.001 (***)	18.76 [-11.94 ; 65.78]
	firstorder_Minimum	<0.001 (***)	0.25 [-0.13 ; 1.09]		glrlm_HighGrayLevelRunEmphasis	<0.001 (***)	22.09 [-110.68 ; 235.67]
	firstorder_Range	<0.001 (***)	-0.81 [-5.13 ; -0.13]		glrlm_LongRunEmphasis	0.1	0.0 [-0.04 ; 1.12]
	firstorder_RobustMeanAbsoluteDeviati	<0.001 (***)	-0.04 [-0.15 ; 0.16]		glrlm_LongRunHighGrayLevelEmphasis	<0.001 (***)	34.26 [-118.94 ; 310.16]
	firstorder_RootMeanSquared	0.62	-0.0 [-0.46 ; 0.58]		glrlm_LongRunLowGrayLevelEmphasis	<0.001 (***)	0.0 [-0.04 ; 0.06]
	firstorder_Skewness	<0.01 (**)	-0.02 [-0.34 ; 0.22]		glrlm_LowGrayLevelRunEmphasis	<0.001 (***)	0.0 [-0.03 ; 0.04]
	firstorder_TotalEnergy	<0.01 (**)	367.4 [-31692.67 ; 1070036.9		glrlm_RunEntropy	<0.001 (***)	0.04 [-0.13 ; 0.28]
	firstorder_Uniformity	<0.001 (***)	-0.0 [-0.01 ; 0.02]		glrlm_RunLengthNonUniformity	0.85	0.9 [-772.5 ; 39.45]
	firstorder_Variance	<0.001 (***)	-0.16 [-5.37 ; 1.74]		glrlm_RunLengthNonUniformityNormalized	0.43	0.0 [-0.08 ; 0.01]
Gray Level Co-occurrence Matrix	glcm_Autocorrelation	<0.001 (***)	32.8 [-108.96 ; 302.57]		glrlm_RunPercentage	0.63	-0.0 [-0.1 ; 0.01]
	glcm_ClusterProminence	<0.001 (***)	3901.61 [-175021.65 ; 674754.		glrlm_RunVariance	<0.05 (*)	0.0 [-0.02 ; 0.64]
	glcm_ClusterShade	<0.001 (***)	1452.16 [-1515.44 ; 6965.97]		glrlm_ShortRunEmphasis	0.56	0.0 [-0.05 ; 0.01]
	glcm_ClusterTendency	<0.001 (***)	68.57 [-48.76 ; 212.39]		glrlm_ShortRunHighGrayLevelEmphasis	<0.001 (***)	20.21 [-109.38 ; 228.93]
	glcm_Contrast	<0.001 (***)	4.22 [-20.95 ; 63.31]		glrlm_ShortRunLowGrayLevelEmphasis	<0.001 (***)	0.0 [-0.03 ; 0.04]
	glcm_Correlation	<0.001 (***)	0.02 [-0.06 ; 0.09]	Neighbouring GrayTone Difference Matrix	ngtdm_Busyness	<0.01 (**)	0.01 [-0.19 ; 0.12]
	glcm_DifferenceAverage	<0.001 (***)	0.2 [-0.55 ; 1.95]		ngtdm_Coarseness	<0.01 (**)	-0.0 [-0.0 ; 0.01]
	glcm_DifferenceEntropy	<0.001 (***)	0.02 [-0.16 ; 0.25]		ngtdm_Complexity	<0.001 (***)	214.16 [-1087.88 ; 3109.34]
	glcm_DifferenceVariance	<0.001 (***)	1.0 [-7.23 ; 21.33]		ngtdm_Contrast	<0.001 (***)	0.03 [-0.65 ; 0.34]
	glcm_Id	<0.01 (**)	-0.0 [-0.03 ; 0.07]		ngtdm_Strength	<0.001 (***)	0.73 [-8.51 ; 6.65]
	glcm_Idm	<0.05 (*)	-0.0 [-0.02 ; 0.08]	Gray Level Dependence Matrix	gldm_DependenceEntropy	<0.001 (***)	0.04 [-0.19 ; 0.23]
	glcm_Idmn	<0.001 (***)	-0.0 [-0.01 ; 0.0]		gldm_DependenceNonUniformity	0.9	1.43 [-188.75 ; 44.87]
	glcm_Idn	<0.001 (***)	-0.0 [-0.02 ; 0.01]		gldm_DependenceNonUniformityNormalize	0.16	0.0 [-0.05 ; 0.09]
	glcm_Imc1	<0.001 (***)	-0.01 [-0.06 ; 0.02]		gldm_DependenceVariance	<0.05 (*)	0.04 [-1.53 ; 27.27]
	glcm_Imc2	<0.001 (***)	0.01 [-0.0 ; 0.03]		gldm_GrayLevelNonUniformity	<0.001 (***)	-0.94 [-58.39 ; 148.9]
	glcm_InverseVariance	<0.05 (*)	-0.0 [-0.04 ; 0.03]		gldm_GrayLevelVariance	<0.001 (***)	19.73 [-12.98 ; 66.4]
	glcm_JointAverage	<0.01 (**)	0.37 [-2.3 ; 4.37]		gldm_HighGrayLevelEmphasis	<0.001 (***)	22.35 [-109.79 ; 239.0]
	glcm_JointEnergy	<0.01 (**)	-0.0 [-0.0 ; 0.02]		gldm_LargeDependenceEmphasis	0.19	0.02 [-1.63 ; 64.99]
	glcm_JointEntropy	<0.001 (***)	0.03 [-0.56 ; 0.23]		gldm_LargeDependenceHighGrayLevelEmph	<0.001 (***)	382.84 [-802.27 ; 5417.16]
	glcm_MCC	<0.001 (***)	0.02 [-0.02 ; 0.07]		gldm_LargeDependenceLowGrayLevelEmph	<0.05 (*)	0.0 [-2.3 ; 3.47]
	glcm_MaximumProbability	0.34	-0.0 [-0.0 ; 0.09]		gldm_LowGrayLevelEmphasis	<0.001 (***)	0.0 [-0.03 ; 0.04]
	glcm_SumAverage	<0.01 (**)	0.74 [-4.6 ; 8.74]		gldm_SmallDependenceEmphasis	0.41	0.0 [-0.05 ; 0.08]
	glcm_SumEntropy	<0.001 (***)	0.06 [-0.2 ; 0.25]		gldm_SmallDependenceHighGrayLevelEmph	0.38	3.4 [-125.21 ; 227.65]
	glcm_SumSquares	<0.001 (***)	18.8 [-17.43 ; 66.07]		gldm_SmallDependenceLowGrayLevelEmph	<0.001 (***)	0.0 [-0.03 ; 0.02]

*$p < 0.05$, **$p < 0.01$ and ***$p < 0.001$. Yellow filling means that the features are not significantly different $p>0.05$ and have a CCC>0.85.*

of features selected based on their stability was 71% (CCC>0.8) in PET NSCLC patients. In this study the stability of FDG PET radiomics features in lesions was 86.5% (CCC≥0.85) between AI denoised and original images. These values are at least of the same order of magnitude highlighting the performance of AI for denoising in PET imaging. As a consequence, a predictive model built on standard PET images could be transposed on AI denoised images, especially concerning the features we have shown as stable in this study. However, the threshold values will have to be recomputed. On the other hand, in healthy tissues as liver and lung most of the features were unstable. Stable features were even less frequent in the liver (11.5%) than in the lung (25%). The effect of denoising on these tissues seems more drastic than on lesions. We hypothesize that the AI algorithm recognizes similar healthy features and changes their intensity value and distributions. As a consequence, the ratio of the lesion over liver uptake should be

transposed with care in clinical PET evaluation, as this ratio is altered for AI denoised versus original PET images.

The difference of behaviors in lesions and healthy tissue is one of the main advantages of AI based methods compared to an EARL1 gaussian post filter method. AI denoising maintains in the lesion the textural information and FDG uptake more stable while modifying healthy tissue. CV measures noise but also grey levels and is correlated to NECR/image quality in PET (29). As shown in **Figures 3A, B**, AI denoising had almost no effect on CV in lesions but reduced it in liver. On the contrary, EARL1 Gaussian postfilter reduced CV similarly in lesions and liver. Gaussian post filter will apply denoising accounting for neighbor all over the images whereas AI may be more selective in amplitude of denoising depending on noise vs non-noise components. The distribution of SUV mean in lesions has a different behavior between AI (paired t-test p=0.06) and EARL1 post filter (paired t-test p<0.001) compared to original images.

Interestingly EARL1 gaussian postfilter led to no modification of SUV mean values in liver (**Figure 3D**). It is mainly due to the increase of point spread function caused by the application of a gaussian postfilter. In a large homogenous area SUVmean was not modified while the noise (CV) was reduced. In smaller, more heterogeneous areas it will melt the grey levels of the different neighbors, lesions and healthy voxels (30). The modification of all the tissues in the image by the gaussian postfilter also appeared in **Figure 2**. Even in lesions only 44% of the features remained stable in EARL1 compatible versus original PET images.

In this analysis we tried to minimize the bias inherent of a radiomics workflow. We use pyradiomics which is mostly compatible with IBSI initiative. Each AI and EARL1 denoised images were extracted from the same images. The same VOI were used on all the series. One could however point out the use of the same contours for lesions in the 3 images as a possible study drawback. Re-segmentation of lesions on each image could have led to different contours and feature values. There is no gold standard for a segmentation method in PET radiomics. It remains also unclear to which extent this can affect radiomics values and predictive models (31). We chose a resampling of 64 bins instead of a fixed bin width (32) even if it showed a better reproducibility. As we directly compared images before and after denoising (minimum and maximum values of the image changed) resampling with a fixed bin width could lead to a different number of bins just due to noise reduction and not to texture based information. In a future work we would apply the same methodology with bin width resampling to strengthen our outcome. We didn' t split the data into training, validation and test cohorts in this study due to the relatively small number of patients and lesions (33). A test-retest radiomics study on patient in CT showed that 446/542 features had a higher CCC for patients with lung cancer than for those with rectal cancer (34). Our study was based on 113 patients, which is a small number. Pooling however different primary malignancies and lesions' nature and size might have helped to reduce overfitting. The main next challenge will be to validate our findings on different and heterogenous patient cohorts and other PET protocols and systems. It might be very risky to apply the same selection of features on other PET or even MRI or CT systems (25). Also, the mechanism of AI denoising recognizing successfully non noise *versus* noise components has to be further investigated on other camera types and PET protocols.

Numerical PET/CT's have a better spatial and temporal resolution leading to a more contrasted activity distribution in lesions than analog systems (35). As this study was carried out on a digital PET/CT we can expect that it will have been more sensitive to variations in texture compared to one on an analog system.

CONCLUSION

Applying an AI, CNN denoising on FDG PET images maintains most of the lesion's texture information in contrast to a EARL1-compatible Gaussian postfilter. The predictive texture features of a trained model could be transposed, however with an adapted threshold. Artificial intelligence in PET is a very promising approach as it adapts the denoising for noise *versus* non-noise components preserving information where it should.

DATA AVAILABILITY STATEMENT

All the data and the python code of the analysis are available on https://github.com/AurelienCD/RadiomicsIA_PET_Depository_Manuscript-ID-692973.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

CJ has designed the study, computed the radiomics and written the majority of the manuscript. KW included study patients, performed image analysis including segmentation, wrote manuscript parts and substantially revised it. AC-D has computed the statistics and written manusript sections. AL gave feedback on the AI algorithm. AB helped to design the study. SB gave advice and feedback on the study. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We would like to thank Henry Austins for the diligent correction of the English language.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.692973/full#supplementary-material

REFERENCES

1. OECD statistique. Available at: https://stats.oecd.org/.
2. SNITEM. Available at: https://www.snitem.fr/en/node/69277.
3. Forghani R, Savadjiev P, Chatterjee A, Muthukrishnan N, Reinhold C, Forghani B. Radiomics and Artificial Intelligence for Biomarker and Prediction Model Development in Oncology. *Comput Struct Biotechnol J* (2019) 17:995. doi: 10.1016/j.csbj.2019.07.001
4. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: Extracting More Information From Medical Images Using Advanced Feature Analysis. *Eur J Cancer* (2012) 48:441–6. doi: 10.1016/j.ejca.2011.11.036. C., Gillies, R., Boellard, R. Dekker, A., and Aerts, HJ.
5. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding Tumour Phenotype by Noninvasive Imaging Using a Quantitative Radiomics Approach. *Nat Commun* (2014) 5:4006. doi: 10.1038/ncomms5006. 6, Haibe-Kains B, Rietveld D, Hoebers F, Rietbergen MM, Leemans CR. Dekker A, Quackenbush J, Gillies RJ, Lambin P.
6. Song J, Yin Y, Wang H, Chang Z, Liu Z, Cui L. A Review of Original Articles Published in the Emerging Field of Radiomics. *Eur J Radiol* (2020) 108991. doi: 10.1016/j.ejrad.2020.108991
7. Rogers W, Thulasi Seetha S, Refaee TA, Lieverse RI, Granzier RW, Ibrahim A, et al. Radiomics: From Qualitative to Quantitative Imaging. *Br J Radiol* (2020) 93(1108):20190948. doi: 10.1259/bjr.20190948
8. Visvikis D, Le Rest CC, Jaouen V, Hatt M. Artificial Intelligence, Machine (Deep) Learning and Radio (Geno) Mics: Definitions and Nuclear Medicine Imaging Applications. *Eur J Nucl Med Mol Imaging* (2019) 46(13):2630–7. doi: 10.1007/s00259-019-04373-w
9. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A Survey on Deep Learning in Medical Image Analysis. *Med Image Anal* (2017) 42:60–88. doi: 10.1016/j.media.2017.07.005
10. Cui J, Gong K, Guo N, Wu C, Meng X, Kim K, et al. PET Image Denoising Using Unsupervised Deep Learning. *Eur J Nucl Med Mol Imaging* (2019) 46 (13):2780–9. doi: 10.1007/s00259-019-04468-4
11. Subtlepet. . Available at: https://subtlemedical.com/usa/subtlepet/.
12. Boellaard R, Hristova I, Ettinger S, Stroobants S, Chiti A, Bauer A, et al. Initial Experience With the EANM Accreditation Procedure of FDG PET/CT Devices. *Eur J Cancer* (2011) 47(Suppl. 4):S8 Abstract. doi: 10.1016/S0959-8049(11)72621-1
13. Boellaard R, Delgado-Bolton R, Oyen WJ, Giammarile F, Tatsch K, Eschner W, et al. Fdg PET/CT: EANM Procedure Guidelines for Tumour Imaging: Version 2.0. *Eur J Nucl Med Mol Imaging* (2015) 42(2):328–54. doi: 10.1007/s00259-014-2961-x
14. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S, et al. 3d Slicer as an Image Computing Platform for the Quantitative Imaging Network. *Magnetic Resonance Imaging* (2012) 30(9):1323–41. doi: 10.1016/j.mri.2012.05.001
15. Beichel RR, Van Tol M, Ulrich EJ, Bauer C, Chang T, Plichta KA, et al. Semiautomated Segmentation of Head and Neck Cancers in 18F-FDG PET Scans: A Just-Enough-Interaction Approach. *Med Phys* (2016) 43 (6Part1):2948–64. doi: 10.1118/1.4948679
16. Van Griethuysen JJ, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res* (2017) 77(21):e104–7. doi: 10.1158/0008-5472.CAN-17-0339
17. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJ, Andrearczyk V, Apte A, et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-throughput Image-Based Phenotyping. *Radiology* (2016) 295(2):328–38. doi: 10.1148/radiol.2020191145
18. Park JE, Park SY, Kim HJ, Kim HS. Reproducibility and Generalizability in Radiomics Modeling: Possible Strategies in Radiologic and Statistical Perspectives. *Korean J Radiol* (2019) 20(7):1124. doi: 10.2967/jnumed.110.082404
19. Yaniv Z, Lowekamp BC, Johnson HJ, Beare R. SimpleITK Image-Analysis Notebooks: A Collaborative Environment for Education and Reproducible Research. *J Digital Imaging* (2018) 31(3):290–303. doi: 10.1007/s10278-017-0037-8. 31(3):290-303.
20. Lawrence I, Lin K. A Concordance Correlation Coefficient to Evaluate Reproducibility. *Biometrics* (1989) 45(1):255–68. doi: 10.2307/2532051
21. Peerlings J, Woodruff HC, Winfield JM, Ibrahim A, Van Beers BE, Heerschap A, et al. Stability of Radiomics Features in Apparent Diffusion Coefficient Maps From a Multi-Centre Test-Retest Trial. *Sci Rep* (2019) 9(1):1–10. doi: 10.1038/s41598-019-41344-5
22. Anaconda Software Distribution. *Anaconda Documentation*. Anaconda Inc. (2020). Available at: https://docs.anaconda.com/.
23. Seifert R, Weber M, Kocakavuk E, Rischpler C, Kersting D. AI and Machine Learning in Nuclear Medicine: Future Perspectives. In: *Seminars in Nuclear Medicine*. WB Saunders (2020). doi: 10.1053/j.semnuclmed.2020.08.003
24. Lovinfosse P, Polus M, Van Daele D, Martinive P, Daenen F, Hatt M, et al. Fdg PET/CT Radiomics for Predicting the Outcome of Locally Advanced Rectal Cancer. *Eur J Nucl Med Mol Imaging* (2018) 45(3):365–75. doi: 10.1007/s00259-017-3855-5
25. Hatt M, Le Rest CC, Antonorsi N, Tixier F, Tankyevych O, Jaouen V, et al. Radiomics in PET/CT: Current Status and Future Ai-Based Evolutions. In: *Seminars in Nuclear Medicine*. WB Saunders Seminars in Nuclear Medicine (2020). doi: 10.1053/j.semnuclmed.2020.09.002
26. Kaplan S, Zhu YM. Full-Dose PET Image Estimation From Low-Dose PET Image Using Deep Learning: A Pilot Study. *J Digital Imaging* (2019) 32 (5):773–8. doi: 10.1007/s10278-018-0150-3
27. Gong K, Guan J, Liu CC, Qi J. PET Image Denoising Using a Deep Neural Network Through Fine Tuning. *IEEE Trans Radiat Plasma Med Sci* (2018) 3 (2):153–61. doi: 10.1109/TRPMS.2018.2877644
28. Leijenaar RT, Carvalho S, Velazquez ER, Van Elmpt WJ, Parmar C, Hoekstra OS, et al. Stability of FDG-PET Radiomics Features: An Integrated Analysis of Test-Retest and Inter-Observer Variability. *Acta Oncol* (2013) 52(7):1391–7. doi: 10.3109/0284186X.2013.812798
29. Reynés-Llompart G, Sabaté-Llobera A, Llinares-Tello E, Martí-Climent JM, Gámez-Cenzano C. Image Quality Evaluation in a Modern PET System: Impact of New Reconstructions Methods and a Radiomics Approach. *Sci Rep* (2019) 9(1):1–9. doi: 10.1038/s41598-019-46937-8
30. Soret M, Bacharach SL, Buvat I. Partial-volume Effect in PET Tumor Imaging. *J Nucl Med* (2007) 48(6):932–45. doi: 10.2967/jnumed.106.035774
31. Cook GJ, Azad G, Owczarczyk K, Siddique M, Goh V. Challenges and Promises of PET Radiomics. *Int J Radiat Oncol Biol Phys* (2018) 102 (4):1083–9. doi: 10.1016/j.ijrobp.2017.12.268
32. Leijenaar RT, Nalbantov G, Carvalho S, Van Elmpt WJ, Troost EG, Boellaard R, et al. The Effect of SUV Discretization in Quantitative FDG-PET Radiomics: The Need for Standardized Methodology in Tumor Texture Analysis. *Sci Rep* (2015) 5(1):1–10. doi: 10.1038/srep11075
33. Steyerberg EW. Validation in Prediction Research: The Waste by Data Splitting. *J Clin Epidemiol* (2018) 103:131–3. doi: 10.1016/j.jclinepi.2018.07.010
34. van Timmeren JE, Leijenaar RT, van Elmpt W, Wang J, Zhang Z, Dekker A, et al. Test–Retest Data for Radiomics Feature Stability Analysis: Generalizable or Study-Specific? *Tomography* (2016) 2(4):361. doi: 10.18383/j.tom.2016.00208
35. Van Sluis J, De Jong J, Schaar J, Noordzij W, Van Snick P, Dierckx R, et al. Performance Characteristics of the Digital Biograph Vision PET/CT System. *J Nucl Med* (2019) 60(7):1031–6. doi: 10.2967/jnumed.118.215418

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Jaudet, Weyts, Lechervy, Batalla, Bardet and Corroyer-Dulmont. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.