



HAL
open science

Détection automatique de lésions par "intelligence artificielle" en TEP/TDM au [18F]FDG en routine clinique: quelle méthode est la plus performante?

Maxence Gambiez

► To cite this version:

Maxence Gambiez. Détection automatique de lésions par "intelligence artificielle" en TEP/TDM au [18F]FDG en routine clinique: quelle méthode est la plus performante?. Médecine humaine et pathologie. 2024. dumas-04618800

HAL Id: dumas-04618800

<https://dumas.ccsd.cnrs.fr/dumas-04618800v1>

Submitted on 20 Jun 2024

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



UNIVERSITÉ DE PICARDIE JULES VERNE
FACULTÉ DE MÉDECINE D'AMIENS

Année 2024

N°2024-19

**DETECTION AUTOMATIQUE DE LÉSIONS PAR « INTELLIGENCE
ARTIFICIELLE » EN TEP/TDM AU [¹⁸F]FDG EN ROUTINE CLINIQUE :
QUELLE MÉTHODE EST LA PLUS PERFORMANTE ?**

THÈSE POUR LE DIPLÔME D'ÉTAT DE DOCTEUR EN MÉDECINE
Spécialité : MÉDECINE NUCLÉAIRE

Présentée et soutenue publiquement le 12 avril 2024 par

Maxence GAMBIEZ

Président du jury :

Monsieur le Professeur **Marc-Etienne MEYER**

Membres du jury :

Monsieur le Professeur **Jean-Pierre MAROLLEAU**

Monsieur le Docteur **Laurent DERCLE**

Monsieur le Docteur **David MORLAND**

Monsieur le Docteur **Paul BLANC-DURAND**

Directeur de thèse :

Monsieur le Docteur **Antoine GIRARD**

DEDICACES ET REMERCIEMENTS

A Monsieur le Professeur Marc-Etienne MEYER, président du jury

Professeur des Universités-Praticien Hospitalier (Médecine nucléaire et traitement d'image)

Chef du service de médecine nucléaire et traitement de l'image, CHU Amiens-Picardie

Pôle Imagerie

Cher professeur,

Je vous remercie pour votre accompagnement tout le long de mes premières années d'internat, du premier jour où vous m'avez fait visiter le service, à celui-ci où vous me faites l'honneur de présider le jury de ma thèse.

J'espère que vous validerez tout de même mon diplôme malgré le fait que je ne sois jamais venu en vélo dans le service... les miracles arrivent ceci dit !

A Monsieur le Professeur Jean-Pierre MAROLLEAU

Professeur des Universités-Praticien Hospitalier (Hématologie, transfusion)

Chef du service d'hématologie clinique et de thérapie cellulaire, CHU Amiens-Picardie

Chef du Pôle « Oncopôle »

Cher Professeur,

Je vous remercie d'avoir instantanément accepté de participer au jury de ma thèse.

J'espère que ce travail, portant en partie sur une population de patients atteints de lymphome, suscitera un intérêt de votre part sur le sujet, et des collaborations futures sur les larges perspectives de ce domaine.

A Monsieur le Docteur Laurent DERCLE

MD, PhD, associate research scientist

Department of Radiology at Columbia University Irving Medical Center

Docteur,

Votre participation à distance au jury de cette thèse m'honore.

J'espère que vous serez intéressé par les résultats de notre travail, et qu'il participera à façonner de nombreux projets autour de l'intelligence artificielle dans notre spécialité outre-Atlantique.

A Monsieur le Docteur David MORLAND

MCU-PH (Médecine Nucléaire et traitement de l'image)

Institut Godinot, Reims

Docteur,

Je vous remercie d'avoir porté intérêt à cette thèse et de participer à ce jury.

Votre expertise en intelligence artificielle est un atout majeur pour l'appréciation de ce travail, qui je l'espère vous portera satisfaction, et motivera de futures collaborations.

A Monsieur le Docteur Paul BLANC-DURAND

MCU-PH (Médecine nucléaire et traitement de l'image)

AP-HP, Assistance Publique - Hôpitaux de Paris

Docteur,

Votre participation à ce jury était une évidence, et je vous remercie d'avoir fait le déplacement pour apporter votre regard indispensable à cette étude.

J'espère sincèrement que les données recueillies à l'issue de notre travail seront bénéfiques à la poursuite de vos travaux, et que cela en fera bénéficier l'ensemble de la profession.

A mon directeur de thèse, Monsieur le Docteur Antoine GIRARD

Praticien Hospitalier (Médecine nucléaire et traitement de l'image)

Pôle imagerie

Antoine,

Je mesure à quel point j'ai eu la chance de croiser ta route, et ne te remercierai jamais assez pour ton accompagnement sans faille, ta pédagogie, ta rigueur, tes conseils, et ta bonne humeur.

Tu fais partie de ces médecins que je considère être des exemples pour notre profession.

J'ai hâte de poursuivre mon parcours professionnel à tes côtés.

A toute l'équipe de Médecine Nucléaire du CHU d'Amiens, sans qui mon internat n'aurait pas été le même, Fanny, Sophie, Pascale, Florence, Laure, Constance, Louise, Manon, Lola, Eden, Rebecca, Isabelle, Catherine, Corinne, Véronique, Nathalie, Magali, Hélène, Pascal, Robin, Fred, Thomas, Laurens.

A Marie-Christine, ça me manque tellement de te demander de rajouter des patients en urgence...

A Mathilde et Marcelin, inconditionnels amis. Votre présence et soutien aura été précieux jusqu'au bout, et le sera toujours.

A Chalabia, pour ton rire communicatif, ta chapka culte, ton soutien sans faille, et surtout ta précieuse amitié.

A notre Virgile national, co-interne en or et ami, promis je t'amènerai un de ces jours une cafetière digne de toi !

A Etienne, pour ta motivation et ton sourire communicatifs.

A Ines, pour ton dynamisme et ta pédagogie.

A mes futurs co-internes Charles-Louis et Claire.

A Trevor et Julie, pour vos conseils avisés, votre pédagogie et ce plaisir de travailler avec vous.

A Michèle, pour ton travail précieux en cardiologie dans notre service.

A Romain, pour la personne et le médecin génial que tu es.

A Florian, pour tes enseignements en médecine nucléaire.

A Nicolas, toi t'es un chouette apothicaire !

A Jimmy, pour nos fous rires et nos vacances en radio qui me manquent.

A Charles, sans qui mon semestre en radio n'aurait jamais été le même.

A toute l'équipe de Saclay ! Vous me manquez.

A Albertine et Samia, merci d'avoir fait de ce Saclay ce qu'il a été. Vous êtes géniales !

A Farid, Hanene, et toute l'équipe de radiothérapie de Saint-Quentin.

A Zaïneb et Alice, pour ces souvenirs de covoit et ces six mois ensemble !

A Clem et Mart, vous êtes incroyables. Nos retrouvailles à Amiens n'étaient peut-être pas que le fruit du hasard finalement... Je vous adore et vous souhaite tout le bonheur du monde.

A Pierre-Antoine, mon Buzz l'Eclair favori et partenaire de crêpes !

A Maxime, merci d'avoir su rapidement tirer le meilleur de moi-même. Tu es un exemple pour moi.

A Maleka et Filippo, mes amis belges, je vous adore.

A Adrien, pour ces innombrables séances cinéma, ces repas à la Fossetta, et ta précieuse amitié.

A mes parents, je vous aime. Merci d'avoir été là jusqu'au bout.

A Valentin, pour la magnifique personne que tu es, et ces moments de bonheur vécus et à venir.

A toutes celles et ceux que je n'ai pas pu citer, mais qui ont été là pour moi et m'ont porté jusqu'ici. Merci.

SOMMAIRE

DEDICACES ET REMERCIEMENTS	1
SOMMAIRE	5
LISTE DES ABREVIATIONS	7
LISTE DES DOCUMENTS.....	9
Figures.....	9
Tableaux.....	9
INTRODUCTION.....	10
I. La Tomographie par Emission de Positons (TEP)	10
1.1 Historique de la TEP.....	10
1.2 Principes physiques.....	11
1.2.1 Généralités sur les radioisotopes employés en TEP	11
1.2.2 Les vecteurs ou traceurs	13
1.3 Détection en TEP	15
1.3.1 Principes généraux de détection	15
1.3.2 Scintillateurs	16
1.3.3 Photodétecteurs.....	16
1.4 Quantification en TEP : notions de SUV et de volumes métaboliques.....	17
1.4.1 SUV ou Standardized Uptake Value	17
1.4.2 Volumes métaboliques	17
1.4.3 Mesures des volumes métaboliques : techniques et limites.....	18
II. Lymphomes et mélanomes : quelle place pour la TEP ?.....	19
2.1 Lymphomes.....	19
2.1.1 Place de l'évaluation en TEP pour les lymphomes.....	19
2.1.2 Facteurs pronostiques de la TEP pour les lymphomes	21
2.1.3 Volume métabolique tumoral total (TMTV).....	22
2.1.4 Delta SUVmax (Δ SUVmax).....	22
2.2 Mélanomes.....	23
III. Etat du parc TEP en France	24
IV. Intelligence artificielle : bases théoriques et applications en médecine nucléaire.....	25
4.1 Définitions : intelligence artificielle, machine learning et deep learning	25
4.1.1 Intelligence artificielle (IA).....	25
4.1.2 Machine learning	25
4.1.3 Deep learning.....	26
4.2 Solutions d'intelligence artificielle en TEP	28
4.2.1 SubtlePET™	28
4.2.2 AUTOID et PAIRE	29
V. Objectif.....	31

MATERIALS AND METHODS	32
I. Population.....	32
II. Technical features of PET/CT	32
III. Detection of lesions : the 4 studied methods	33
IV. Reference standard	33
V. Statistical analysis	34
RESULTS.....	35
I. Sample characteristics	35
II. Considering each method independently.....	37
III. Considering each AI-based method (M2, M3 and M4) as a completing human reading (M1).....	41
IV. False negative lesions.....	42
V. False positive VOIs	43
DISCUSSION	45
I. Main results	45
II. Strengths and limitations of the study	46
III. Literature	46
IV. Perspectives.....	47
CONCLUSION.....	50
BIBLIOGRAPHIE	51
RESUME - ABSTRACT.....	55

LISTE DES ABREVIATIONS

BGO : Germanate de bismuth

BOM : Biopsie ostéo-médullaire

CHC : Carcinome hépato-cellulaire

CNN : Convolutional neural network (= réseau de neurones convolutifs)

FDG : Flurodésoxyglucose

FDG-6-P : Fluorodésoxyglucose-6-Phosphate

FES : Fluoroestradiol

FET : Fluoroéthyl-L-tyrosine

G6P : Glucose-6-Phosphate

IRM : Imagerie par Résonance Magnétique

IA : Intelligence Artificielle (= AI : Artificial Intelligence)

LBDGC : Lymphome B diffus à grandes cellules (= DLBCL : diffuse large B-cell lymphoma)

LH : Lymphome de Hodgkin

LF : Lymphome folliculaire

LNH : Lymphome non-Hodgkinien

LSO : Oxyorthosilicate de lutécium dopé cérium

LYSO : Oxyorthosilicate d'yttrium et de lutécium

MIP : Maximum Intensity Projection

MTV : Metabolic Tumor Volume (= Volume Métabolique Tumoral)

PSMA : Prostate Specific Membrane Antigene

ROI : Region of interest (= region d'intérêt)

Se- : Sensitivity (= sensibilité)

Sp- : Specificity (= spécificité)

SUV : Standardized Uptake Value

TEP : Tomographie par Emission de Positons (= PET : Positron Emission Tomography)

TAP : Thoraco-abdomino-pelvien

TDM : Tomodensitométrie (= CT : Computed Tomography)

TLG : Total Lesion Glycolysis (= Glycolyse Lésionnelle Totale)

TLV : Total Lesion Volume (= Volume Lésionnel Total)

TMTV : Total Metabolic Tumor Volume (= Volume Métabolique Tumoral Total)

TNE : Tumeur neuro-endocrine

TP : True Positive (= VP : Vrai Positif)

FP : False Positive (= FP : Faux Positif)

FN : False Negative (= FN : Faux Négatif)

PPV : Positive Predictive Value (= VPP : Valeur Prédictive Positive)

NPV : Negative Predictive Value (= VPN : Valeur Prédictive Négative)

VOI : Volume of Interest (= Volume d'Intérêt)

LISTE DES DOCUMENTS

Figures

- **Figure 1** : Premières images TEP chez l'homme en 1976.
- **Figure 2** : Structure chimique du [18F]FDG.
- **Figure 3** : Mécanismes cellulaires de captation et d'accumulation du glucose et du FDG.
- **Figure 4** : Principes généraux de détection en TEP.
- **Figure 5** : Protocole GAINED.
- **Figure 6** : Evolution du nombre d'examens TEP annuels réalisés en France entre 2009 et 2022.
- **Figure 7** : Evolution du nombre de médecins nucléaires en France (sources : atlas démographiques du CNOM).
- **Figure 8** : Exemple d'un CNN avec une architecture en U-Net en trois couches.
- **Figure 9** : Distribution of the number of lesions for each organ, per patient, according to the reference standard.
- **Figure 10** : Number of volumes of interest detected with each method independently.
- **Figure 11** : Example of a patient managed for melanoma for whom four evaluated methods provided mismatching VOI on [18F]FDG PET-CT images.
- **Figure 12** : Repartition of true positive and false negative lesions of each AI-based method as completing human reading.
- **Figure 13** : Number of lesions detected with each method compared to the reference standard (black horizontal lines), classified by organs.
- **Figure 14** : Number of false positive VOIs, classified by organs.

Tableaux

- **Tableau 1** : Principaux radioisotopes émetteurs β^+ utilisés en TEP.
- **Tableau 2** : Les traceurs utilisés en TEP et leurs applications (liste non exhaustive).
- **Tableau 3** : Recommandations internationales de la place de la TEP dans la prise en charge des lymphomes.
- **Tableau 4** : Avidité au FDG des types histologiques de lymphomes.
- **Tableau 5** : Classification d'Ann Arbor modifiée par Lugano.
- **Tableau 6** : Echelles de Deauville « classique » et modifiée.
- **Table 7** : Diagnostic performances of each method independently, per lesions
- **Table 8** : Comparison of diagnostic sensitivities of each method, for both "melanoma" and "lymphoma" subgroups, per lesion.
- **Table 9**: Diagnostic performances of each AI-based method as completing human reading, per lesions.

INTRODUCTION

I. La Tomographie par Emission de Positons (TEP)

1.1 Historique de la TEP

L'imagerie métabolique et fonctionnelle a, dès la deuxième moitié du XXème siècle, changé le paradigme selon lequel les processus physiologiques et pathologiques étaient observables uniquement par le biais de modifications morphologiques. Ils pouvaient désormais être appréciés par une étude directe et indirecte des mécanismes cellulaires, grâce à l'utilisation de la radioactivité.

C'est en 1952 qu'est né le premier appareil d'imagerie positonique utilisable dans une démarche clinique. G.L. Brownell et W.H. Sweet utilisaient alors deux détecteurs (compteur à scintillation) de iodure de sodium dopé au thallium (NaI(Tl)), en coïncidence, chez un patient présentant une tumeur cérébrale récurrente (1). Le radioisotope employé correspondait à du Gallium 68 (^{68}Ga), utilisé en routine clinique de nos jours, couplé à des vecteurs, par exemple pour l'étude des tumeurs neuroendocrines exprimant des récepteurs à la somatostatine (2).

En 1971, l'arrivée de la tomographie par émission de positons (TEP), constituée de deux plans de détecteur NaI(Tl) en coïncidence, selon un dispositif rotatif assisté par ordinateur (3), a permis la visualisation d'images tomographiques, reconstruites par rétroprojection filtrée (4). Trois années plus tard, Phelps et al. ont initié le design des tomographes actuels, basés sur 48 détecteurs de NaI(Tl) sur un réseau orthogonal, à usage clinique, commercialisés en 1976, réalisant alors les premières image TEP chez l'homme (5,6).

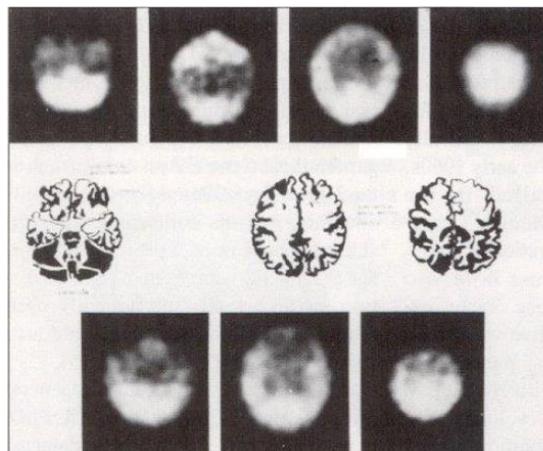


Figure 1 : Premières images TEP chez l'homme en 1976. En haut : étude du débit sanguin cérébral avec du $[^{13}\text{N}]\text{-NH}_3$. En bas : étude du métabolisme glucidique cérébral avec du $[^{11}\text{C}]\text{-Glucose}$ (5).

La visualisation de ces images était limitée par la faible résolution spatiale (environ 20 mm) (5) ne permettant pas des analyses optimisées telles que nous pouvons les avoir actuellement. Les évolutions technologiques successives des 40 dernières années continuent d'inclure et de généraliser l'utilisation de la TEP dans une routine clinique, notamment avec le passage de la TEP/TDM analogique à numérique, mais aussi avec le couplage avec l'imagerie par résonance magnétique (IRM).

1.2 Principes physiques

La TEP utilise le principe de détection de rayonnements gamma émis par un radioisotope émetteur de positons couplé à un vecteur, injecté au patient. Le traceur (grâce à sa partie vectrice) est destiné à étudier de manière plus ou moins spécifique une cible (récepteurs, transporteurs...), ayant des caractéristiques de bio-distribution et biologiques précises, traduisant des mécanismes physio-pathologiques.

1.2.1 Généralités sur les radioisotopes employés en TEP

1.2.1.1 La radioactivité

La radioactivité est un processus atomique purement nucléaire, la plupart du temps spontané, résultant d'une réaction à l'instabilité. Il s'agit d'un phénomène régi par les lois de probabilité (avec une décroissance exponentielle du nombre de noyaux radioactifs en fonction du temps), exoénergétique (libérant de l'énergie), et ayant pour conséquence l'émission de particules et/ou de rayonnements.

Sur les 331 nucléides (noyaux) naturels, 75 sont considérés instables et donc radioactifs : ils ont une durée de vie variant de la nanoseconde aux milliards d'années. Ils peuvent avoir une origine terrestre (ex : l'uranium 238 ou ^{238}U) ou cosmique (ex : le tritium ou ^3H). Les isotopes artificiels sont quant à eux tous radioactifs.

Différents types de rayonnements sont émis par les isotopes radioactifs, notamment :

- Désintégration β^+ et capture électronique : noyaux trop pauvres en neutrons
⇒ Gain de stabilité par la transformation d'un proton en neutron.
- Désintégration β^- : noyaux trop riches en neutrons
⇒ Gain de stabilité par la transformation d'un neutron en proton.
- Désintégration alpha : noyaux lourds ($Z \geq 83$)
⇒ Gain de stabilité par l'émission d'une particule α (^4He).
- Désintégration par fission spontanée : noyaux très lourds

Ces émissions sont pures lorsqu'un seul type de rayonnement est émis, ou à l'inverse impures, le plus souvent alors associées à l'émission d'un ou plusieurs rayonnements gammas, dont les raies sont caractéristiques d'un élément (passage d'un noyau père à l'état excité à un noyau fils). Chaque radioisotope est caractérisé par sa période ou demi-vie, temps nécessaire pour que la moitié des noyaux d'un échantillon se désintègrent naturellement.

1.2.1.2 Désintégration β^+ et principe d'annihilation.

Ce phénomène s'accompagne de l'émission d'un positon (e^+) et d'un neutrino (particule indétectable, sans charge et sans masse). Parmi les isotopes, certains sont purs (ex : Fluor 18) ou impurs (ex : Rubidium 82).

Le principe de détection par la TEP repose sur les propriétés d'annihilation faisant suite à une désintégration β^+ . Le positon émis lors de celle-ci parcourt une longueur de l'ordre du millimètre (dans l'eau et les tissus), durant laquelle il est constamment freiné (freinage électrostatique), et où il opère plusieurs ionisations dans son environnement proche. A la fin de cette longueur, avec une quantité de mouvement quasiment nulle par perte d'énergie cinétique, il rencontre un électron environnant. De cette collision résulte l'émission de deux rayonnements gamma, émis à 180° l'un de l'autre, emportant chacun une énergie de 511 keV. Ce sont donc ces rayonnements qui sont détectés, en coïncidence.

1.2.1.3 Les radioisotopes en TEP

Les radioisotopes employés en TEP sont des émetteurs de rayonnement β^+ . La production de ces isotopes utilisés en routine clinique ou en recherche résulte de l'usage :

- D'un cyclotron : par exemple pour la production de ^{18}F à partir de ^{18}O .
- D'une élution dans un générateur : par exemple pour la production de ^{68}Ga à partir de ^{68}Ge .

Isotope	Période (demi-vie)	Production
Fluor 18	109,8 minutes	Cyclotron
Rubidium 82	1,3 minutes	Générateur Sr82/Rb82
Oxygène 15	2,04 minutes	Cyclotron
Azote 13	9,96 minutes	Cyclotron
Carbone 11	20,4 minutes	Cyclotron
Cuivre 64	12,7 heures	Cyclotron
Gallium 68	68 minutes	Générateur Ge68/Ga68
Iode 124	4,18 jours	Cyclotron

Tableau 1 : Principaux radioisotopes émetteurs β^+ utilisés en TEP.

1.2.1.4 Exemple du fluor 18

L'isotope radioactif le plus souvent utilisé en pratique est le fluor 18 ou ^{18}F , émetteur β^+ pur, dont la demi-vie d'environ 110 minutes est adaptée à une utilisation clinique. Il est produit à partir d'accélération de protons sur des isotopes d'oxygène 18 ou ^{18}O au sein de cyclotrons, libérant du ^{18}F et un neutron (réaction de type $^{18}\text{O}(\text{p},\text{n})^{18}\text{F}$). Le ^{18}F suit alors une désintégration spontanée de type β^+ .

Le Fluor 18 a l'avantage de présenter une période relativement brève, lui conférant une sécurité dosimétrique pour le patient. Il s'agit d'un isotope léger incorporable dans des molécules organiques, de production aisée via des cyclotrons hors des murs hospitaliers.

1.2.2 Les vecteurs ou traceurs

1.2.2.1 Définition des vecteurs et traceurs

Le traceur, en médecine nucléaire, correspond à une molécule vectrice (ou vecteur) combinée à un radioisotope.

Les vecteurs atteignent d'autres cibles ayant des caractéristiques biologiques et/ou pharmacologiques spécifiques. Ils donnent une information précise d'un processus physiologique ou pathologique, en fonction de la molécule ciblée qui régit cette fonction.

En fonction du modèle étudié, le vecteur en TEP peut se présenter sous différentes formes, par exemple de :

- Molécule entrant dans un métabolisme (ex : FDG).
- Molécule se fixant sur des structures comme des récepteurs ou antigènes (ex : PSMA, choline...).

1.2.2.2 Exemple d'un traceur : le fluorodésoxyglucose ou fluor 18 couplé au glucose

En TEP, le traceur le plus utilisé est le FDG (vecteur) marqué au ^{18}F (radioisotope). Il se comporte comme analogue du glucose, suit la voie métabolique de celui-ci, et témoigne d'activités cellulaires mitotiques accrues ou anormalement basses.

Le glucose entre dans la cellule via un transporteur membranaire GLUT (7–9). Dans le cytoplasme, il est pris en charge par différentes enzymes, dont une hexokinase qui le transforme en glucose-6-phosphate ou G6P (réaction réversible via la glucose-6-phosphatase) (9,10) Le G6P suit en conditions aérobies la voie de la glycolyse via le cycle de Krebs.

Le FDG perd le groupement hydroxyle 2 de la molécule de glucose, à la place duquel se greffe le ^{18}F . Il rentre dans la cellule de manière analogue via le transporteur GLUT, puis est pris en

charge par l'hexokinase. Le FDG-6-P (fluorodésoxyglucose-6-phosphate) obtenu n'est alors plus en mesure d'être métabolisé et s'accumule dans la cellule (de par son hydrophilie), lui permettant d'être détecté (9).

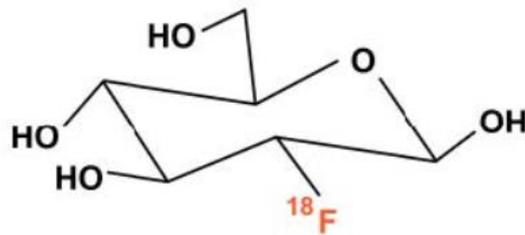


Figure 2 : Structure chimique du [18F]-FDG (11).

En cas de néoplasie ou de foyer inflammatoire, il existe en général à l'échelle cellulaire une accumulation excessive de FDG-6-P (8). A l'inverse, en cas de perte cellulaire anormale, une baisse voire une absence d'activité sera observée (12).

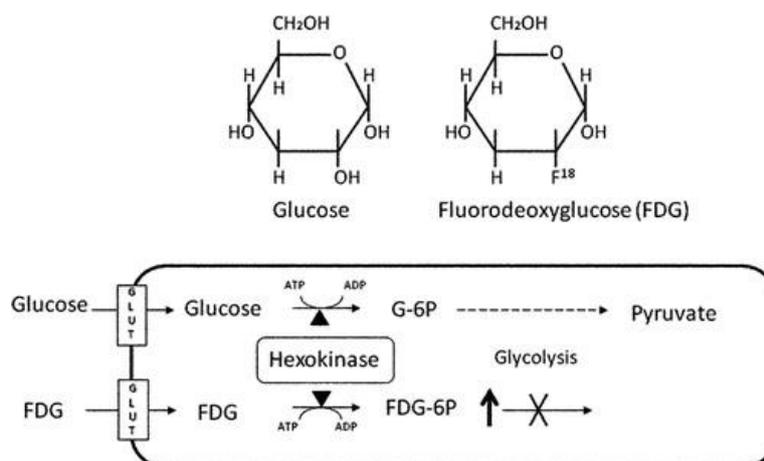


Figure 3 : Mécanismes cellulaires de captation et d'accumulation du glucose et du FDG (13).

Son utilisation présente une innocuité biologique et pharmacologique, et l'absence d'effets indésirables.

Les applications cliniques sont larges :

- En oncologie, ce traceur est utilisé pour la caractérisation des lésions, les bilans d'extension et le staging tumoral, les évaluations d'efficacité thérapeutique, l'optimisation en radiothérapie, les diagnostics de récurrence...
- Hors oncologie, il sera utilisé pour les maladies inflammatoires ou systémiques, les infections (9,10), les mesures de viabilité cardiaque (14) ou les pathologies neurodégénératives (12).

1.2.2.3 Exemples d'autres traceurs en TEP en routine clinique

Les applications de l'utilisation des radioisotopes en TEP sont larges et regroupent plusieurs domaines.

Radiotraceurs	Applications
18F-choline	Adénocarcinomes de prostate CHC Hyperparathyroïdie (adénomes)
18F-Na	Métastases osseuses de néoplasies pulmonaires, mammaires et de prostate
18F-DOPA	TNE Syndromes parkinsoniens Neuro-oncologie
18F-FET	Tumeurs cérébrales
18F-FES	Cancer du sein
18F-florbetapir, flutemetamol	Plaques amyloïdes (démences)
68Ga-PSMA 18F-PSMA	Adénocarcinomes de prostate
68Ga-DOTATOC	TNE Méningiomes

Tableau 2 : Les traceurs utilisés en TEP et leurs applications (liste non exhaustive).

1.3 Détection en TEP

1.3.1 Principes généraux de détection

La détection de deux photons émis à 180° l'un de l'autre est (quasi-)simultanée, par des détecteurs constitués de scintillateurs et de tubes photomultiplicateurs, disposés en anneau pour une meilleure sensibilité de détection. La majorité des systèmes de nos jours fonctionnent selon un mode d'imagerie en trois dimensions, avec une collimation électronique, permettant un gain en sensibilité (multiplié par 5 ou 8) par rapport aux techniques bi-dimensionnelles.

Le principe des détecteurs est de transformer l'information recueillie sous forme d'énergie transportée par un photon de 511 keV, en une information exploitable en signal électronique. L'autre objectif est de localiser l'endroit où a eu lieu l'annihilation. Les sites de l'émission du positon et de l'annihilation sont différents, séparés d'une distance de l'ordre du millimètre dans

l'eau et les tissus (en règle générale 0.6 mm), conférant la principale limite en termes de résolution spatiale des systèmes usuels.

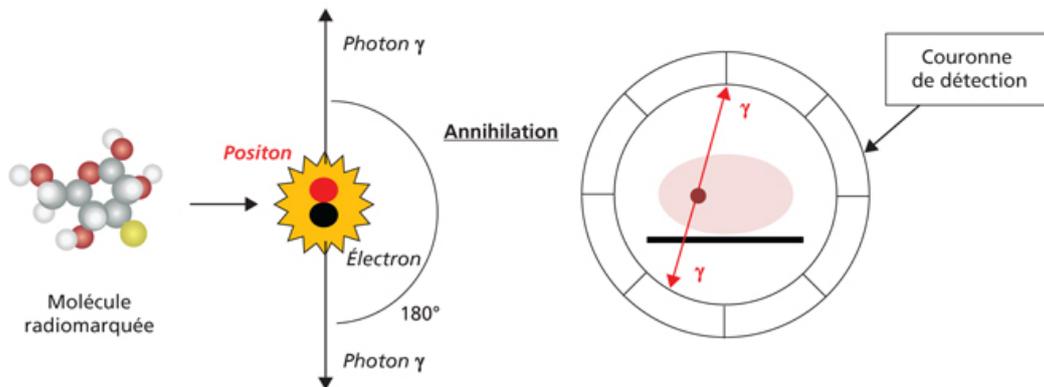


Figure 4 : Principes généraux de détection en TEP (15).

1.3.2 Scintillateurs

Le scintillateur (ou cristal scintillant) est le premier élément du détecteur, placé devant un photodétecteur, et reçoit le photon incident de 511 keV. Par le biais de phénomène d'ionisation et d'excitation des éléments du milieu du scintillateur, le photon incident est converti en photons lumineux.

Pour être adapté à l'utilisation en TEP, c'est-à-dire un flux important de photons de haute énergie, le cristal scintillant doit posséder certaines propriétés imposant une sensibilité de détection satisfaisante (densité élevée du cristal), une résolution en énergie et spatiale adaptée, et une bonne résolution temporelle. Cet ensemble doit être analysé en regard des coûts et disponibilités usuels de ces équipements.

Les principaux cristaux de scintillation utilisés en TEP avec les propriétés adéquates sont le BGO (germanate de bismuth), le LSO (oxyorthosilicate de lutécium dopé cérium) et le LYSO (oxyorthosilicate d'yttrium et de lutécium).

1.3.3 Photodétecteurs

Les photons lumineux arrivent ensuite sur la première portion du photodétecteur, à savoir une photocathode, qui les convertit en électrons.

Dans un système dit analogique, ceux-ci sont alors multipliés au sein d'un photomultiplicateur, avant de rencontrer une anode, ce qui permettra la production d'une impulsion électrique, dont l'amplitude est proportionnelle à l'énergie initiale des photons incidents (reflétant l'importance du flux incident).

Les photomultiplicateurs peuvent être remplacés par des semi-conducteurs plus compacts : des diodes à avalanche ou des « photomultiplicateurs » au silicium (SiPM). L'avantage résidant en une conversion directe des photons lumineux en un signal numérique, permettant un gain important en résolution temporelle (quelques picosecondes) et en sensibilité (pour les SiPM). Cela apporte un intérêt en routine clinique, par exemple l'amélioration du contraste ou un gain de détection des petites lésions.

1.4 Quantification en TEP : notions de SUV et de volumes métaboliques

1.4.1 SUV ou Standardized Uptake Value

La SUV est la grandeur la plus utilisée pour caractériser la fixation du FDG en TEP dans un processus pathologique. Elle correspond au rapport entre l'activité détectée dans un tissu d'intérêt (kBq/ml) et la dose injectée au patient (kBq), elle-même rapportée au poids du patient (g) (8). La SUV est une grandeur exprimée sans dimension (16).

Si la distribution du traceur à un temps donné était uniforme, la valeur du SUV serait identique en tous points. Tout écart par rapport à un 1 signifie que la distribution n'est plus uniforme.

La SUV s'exprime de plusieurs manières (17), le plus couramment :

- En SUV_{max} : valeur accordée au voxel le plus intense d'une région d'intérêt
- En SUV_{mean} : valeur moyenne des voxels d'une région d'intérêt
- En SUV_{peak} : valeur dans une sphère de 1 mL la plus intense contenue dans la région d'intérêt.

La SUL (standardized uptake value corrected for lean body mass) est une des variantes qui permet de corriger la fixation en fonction de la masse maigre, notamment utilisé dans les critères PERCIST (18).

1.4.2 Volumes métaboliques

La mesure de la SUV est soumise à de nombreuses variations (procédures et condition d'examen, conditions physiopathologiques du patient...) (16,19). De fait, pour mieux caractériser la tumeur et ses modifications intra-tumorales en tant que lésion en trois dimensions, ont été introduites les notions de volume métabolique tumoral (metabolic tumor volume (MTV)) et de glycolyse lésionnelle totale (total lesion glycolysis (TLG)).

Le MTV est obtenu par le contour de lésion (manuel ou automatique). Le TLG dérivant de cette mesure est obtenu par la multiplication du SUV_{mean} et du MTV (20), mais n'est pas pleinement utilisé en routine clinique par manque de validation (17).

1.4.3 Mesures des volumes métaboliques : techniques et limites

Un des objectifs majeurs de la TEP est de déterminer de manière optimale et reproductible les différents paramètres ayant un rôle dans la détermination de la sévérité de l'atteinte, son grade, mais aussi impactant le pronostic et la prise en charge. Si la SUV est désormais mesuré de façon robuste, il doit en être de même pour les volumes métaboliques.

La mesure de ces volumes pose aujourd'hui plusieurs problèmes, notamment en terme de non uniformité de la méthode de mesure et de temps consacré à cette tâche (17,21).

La délimitation manuelle est une technique assez spécifique (car excluant au mieux les aires de bruit de fond) et permettant un « contouring » en regard des limites morphologiques normales/pathologiques (via la TDM ou l'IRM en fusion), mais avec une variabilité inter-observateur et un temps de traitement d'image non négligeables, rendant son utilisation en routine clinique peu optimale (17).

L'utilisation d'un seuil est la technique la plus employée, et regroupe plusieurs méthodes (17) :

- Seuil fixe (exemple : SUV à 2,5 ou SUV_{mean} hépatique + 2 déviations standards (DS)), simple et rapide d'utilisation, mais dépend de la valeur du SUV seuil, rend difficile l'évaluation de tumeurs hétérogènes et prend peu en compte la taille de la tumeur et le bruit de fond (BDF) environnant.
- Seuil en pourcentage (exemple : 40 à 50 % du SUV_{max}), également simple et rapide d'utilisation, mais de sensibilité variable en fonction du seuil choisi.
- Seuil adaptatif, où le seuil est adapté au ratio tumeur-BDF.
- Mesure de la différence de gradient entre la tumeur et le BDF environnant.
- Utilisation d'algorithmes avancés incluant la croissance de régions, basés sur des modèles statistiques ou du machine learning.

En pratique, les trois dernières techniques de mesure sont peu disponibles dans les logiciels commercialisés.

Le développement de l'IA (Intelligence Artificielle) doit de fait répondre aux enjeux de la robustesse de détection des lésions et de la mesure des volumes métaboliques, en prenant en compte le gain de temps apporté au traitement de l'image. Se sont ainsi principalement développées deux solutions de deep learning répondant à ces objectifs : AUTOID (Siemens Healthineers, Erlangen, Allemagne) et PAIRE (PAIRE, Paris, France).

II. Lymphomes et mélanomes : quelle place pour la TEP ?

2.1 Lymphomes

2.1.1 Place de l'évaluation en TEP pour les lymphomes

Les lymphomes représentent un groupe hétérogène de pathologies lymphoprolifératives, plus ou moins indolentes ou agressives (22). Nous les distinguons en deux groupes :

- Les lymphomes Hodgkiniens (LH), qui représentent 10 à 15 % de l'ensemble des lymphomes (23).
- Les lymphomes non Hodgkiniens (LNH), eux-mêmes divisés en plusieurs types, dont les principaux sont : les lymphomes B diffus à grandes cellules (32 %), les lymphomes folliculaires (17 %), les lymphomes T (10-15 %), et les lymphomes à cellules du manteau (4 %) (22).

L'évaluation à différentes étapes de la prise en charge des LNH et LH par la TEP/TDM au 18F-FDG est d'utilisation courante en routine clinique, et est de mieux en mieux codifiée.

Stade l'évaluation	Types de lymphomes concernés	Objectifs
Bilan initial	Lymphomes avides en FDG : LH, LBDGC et LF Transformation histologique d'un lymphome non avide en FDG en un lymphome agressif (Syndrome de Richter)	Stadification initiale : modification du stade chez 10 à 30 % des patients Permet de surseoir à la BOM si lésions focales de la moëlle osseuse suspectes de localisation de LH ou de LNH agressif. Permet de guider la biopsie en cas de suspicion clinique et/ou biologique et/ou radiologique de syndrome de Richter
En cours de traitement : • Précoce : après 2 cycles de CT • Intermédiaire (iPET) : après 4 cycles de CT	Identification des patients répondeurs et non répondeurs set adaptation thérapeutique : • LBDGC • LH	Etude de chimio-sensibilité Impact sur la suite du traitement : • Suivi : réponse complète

		<ul style="list-style-type: none"> • Consolidation : réponse partielle • Rattrapage : progression, stabilité
En fin de traitement	Lymphomes avides en FDG	Confirmation de la rémission complète Pose l'indication de biopsie en cas de masse résiduelle (LH et DLBCL)
Surveillance systématique en cas de rémission complète	Non recommandé	

Tableau 3 : Recommandations internationales de la place de la TEP dans la prise en charge des lymphomes (24,25).

Avidité au FDG	Types histologiques de lymphomes
Elevée	Lymphome B diffus à grandes cellules Lymphome de Burkitt Lymphome T anaplasique à grandes cellules Lymphome extra-ganglionnaire à cellules T/NK
Modérée	Leucémie-lymphome à cellules T de l'adulte Lymphome cutané à cellules T
Basse	Mycosis fungoïde (syndrome de Sézary)
Variable	Lymphome folliculaire : <ul style="list-style-type: none"> • Grades 1 et 2 : basse à modérée • Grade 3 : modérée à élevée Lymphome à cellules du manteau : basse à modérée Lymphome de la zone marginale (dont MALT) : aucune à élevée Lymphome à petites cellules lymphocytaires : aucune à élevée

Tableau 4 : Avidité en FDG des types histologiques de lymphomes (21,23).

2.1.2 Facteurs pronostiques de la TEP pour les lymphomes

2.1.2.1 Classifications d'Ann Arbor

La classification la plus connue est celle d'Ann Arbor modifiée par Lugano, établie en 2014 (24–27), utilisée pour la stadification initiale (l'atteinte de l'anneau de Waldeyer, amygdalienne et splénique étant considérée comme ganglionnaire).

Stade	Atteinte ganglionnaire	Atteinte extra-ganglionnaire
Limité		
Stade I	Un ganglion ou un groupe de ganglions adjacents	Lésion extra-ganglionnaire unique sans atteinte ganglionnaire
Stade II	Deux groupes ganglionnaires ou plus du même côté du diaphragme	Stade I ou II par l'extension ganglionnaire avec atteinte extra-ganglionnaire contiguë limitée
Stade II « bulky »	Stade II avec mention « bulky » (volumineux) : <ul style="list-style-type: none">• Masse ganglionnaire de plus de 10 cm de grand axe• Et/ou masse médiastinale supérieure au tiers du diamètre thoracique	
Avancé		
Stade III	Ganglions des deux côtés du diaphragme Ganglion(s) sus-diaphragmatique(s) avec atteinte splénique	Non applicable
Stade IV	Atteinte ganglionnaire et extra-ganglionnaire non contiguë	Non applicable

Tableau 5 : Classification d'Ann Arbor modifiée par Lugano (26).

2.1.2.2 Echelle de Deauville

L'échelle de Deauville de 2009 sera quant à elle pertinente pour les évaluations intermédiaires et en fin de traitement, et est aussi intégrée au sein de la classification de Lugano (24,28).

Elle correspond à une évaluation visuelle et quantitative des SUV des lésions résiduelles par rapport à l'activité métabolique d'un organe de référence (médiastin et foie), et ce quelle que soit la taille de la lésion.

Son utilisation dans l'évaluation intermédiaire est limitée à deux cadres :

- Recommandée pour détecter les patients répondeurs et non répondeurs : LH et LBDGC
- Recommandée pour l'adaptation thérapeutique :
 - ✓ Du LH : échelle de Deauville « classique », ou modifiée selon le protocole AHL 2011 (LH diffus de stade II « Bulky », III et IV chez les patients de moins de 65 ans traités selon BEACOPP_{escalated}) (29,30).
 - ✓ Du LBDGC si la SUVmax initiale est inférieure à 10 : échelle de Deauville « classique ».

Dans le cadre de l'évaluation de fin de traitement, l'échelle de Deauville « classique » est utilisée pour tous types histologiques de lymphomes avides en FDG (24). Un score de 1, 2 ou 3 définit une réponse métabolique complète, et 4 ou 5 une réponse partielle ou progression notamment si apparition de nouvelle(s) lésion(s).

Score	Echelle de Deauville « classique »	Echelle de Deauville modifiée
1	Pas d'hyperfixation	
2	Fixation ≤ médiastin	
3	Médiastin < fixation ≤ foie	Médiastin < fixation < 140 % foie
4	Fixation > foie	Fixation ≥ 140 % foie
5	Fixation >> foie / progression	Fixation ≥ 200 % foie

Tableau 6 : Echelles de Deauville « classique » et modifiée.

2.1.3 Volume métabolique tumoral total (TMTV)

Le TMTV est calculé en additionnant les MTV de chacune des lésions.

Les valeurs prédictives du TMTV pour la survie sans progression ont été étudiées pour de nombreux types de lymphomes : les lymphomes folliculaires, primitifs du médiastin, à cellules du manteau, T périphériques, B diffus à grandes cellules et le lymphome de Hodgkin (17).

Néanmoins, de nombreuses variations au sein des différentes revues d'études montrent l'intérêt d'adopter des méthodes de mesure et de seuillage uniques et reproductibles, et de les étudier sur des populations homogènes, notamment en termes de stades d'évolution du lymphome (17). L'hétérogénéité de méthodologie rend à ce jour son utilisation en routine non consensuelle.

2.1.4 Delta SUVmax (Δ SUVmax)

Le Δ SUVmax correspond à la soustraction entre le SUVmax lésionnel le plus élevé de la TEP initiale et le SUVmax lésionnel le plus élevé de la TEP intermédiaire ou précoce.

En pratique, il est utilisé pour guider le traitement des lymphomes diffus à grandes cellules B, notamment selon le protocole GAINED, chez des patients de moins de 65 ans candidats au traitement intensif (24,31–34), où le Delta SUVmax est utilisé dans le cadre des évaluations intermédiaires des LDBGC (à 2 et 4 cures de chimiothérapie), pour l'adaptation thérapeutique en consolidation.

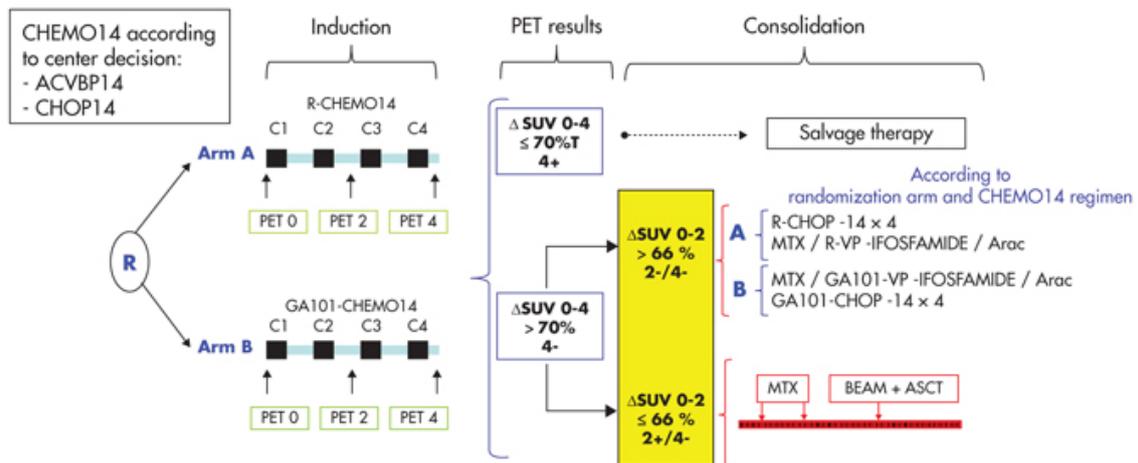


Figure 5 : Protocole GAINED (34).

2.2 Mélanomes

Pour le bilan d'extension initial (35), la TEP/TDM au 18F-FDG trouve son intérêt devant des mélanomes :

- A haut risque de dissémination : Breslow ≥ 4 mm, ulcération, et atteinte ganglionnaire clinique.
- Aux stades avancés : III (N1) ou IV (M1), avec une Se de 85 à 91 % et une Sp de 92 à 98 % (36,37).

L'objectif à ce stade est de déterminer les lésions indéterminées par l'imagerie conventionnelle, les localisations occultes et de cartographier les lésions (35).

Dans le cadre du suivi, en pratique, il existe un intérêt de la TEP pour les stades avancés ou les patients à haut risque de récurrence (35). Néanmoins, elle pourrait aussi être généralisée aux stades plus localisés, du fait de son excellente VPN notamment chez des patients asymptomatiques ou de stades I/II (35,38).

En cas de récurrence connue, qu'elle soit cutanée, ganglionnaire, ou à distance, l'imagerie corps entier est recommandée (TDM TAP ou TEP et IRM cérébrale) (35).

III. Etat du parc TEP en France

D'après une enquête de la Société Française de Médecine Nucléaire (SFMN), nous pouvons constater une nette évolution du nombre d'examen TEP, pour un nombre estimé à un peu plus de 805 000 examens en 2022, soit environ plus 17 % par rapport à 2021.

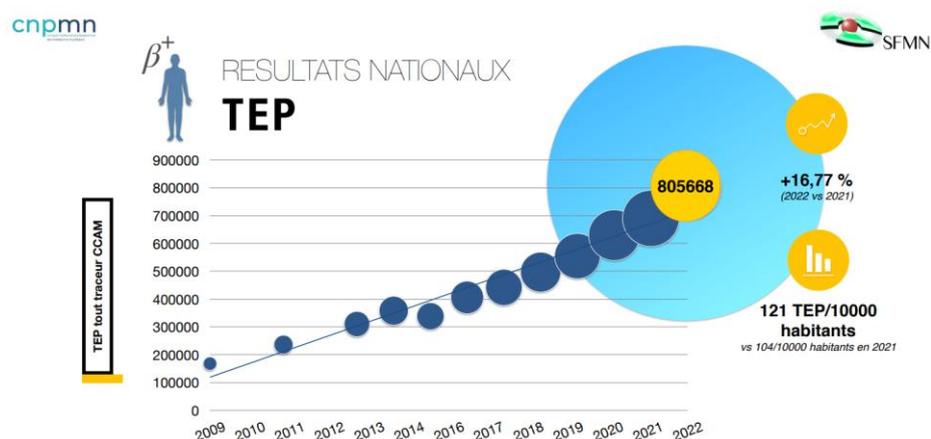


Figure 6 : Evolution du nombre d'examen TEP annuels réalisés en France entre 2009 et 2022.

Une quantité non négligeable d'actes, alors que le nombre de médecins dans la spécialité n'augmente que très peu selon les atlas démographiques du Conseil National de l'Ordre des Médecins (CNOM) (+ 26 % en 10 ans).

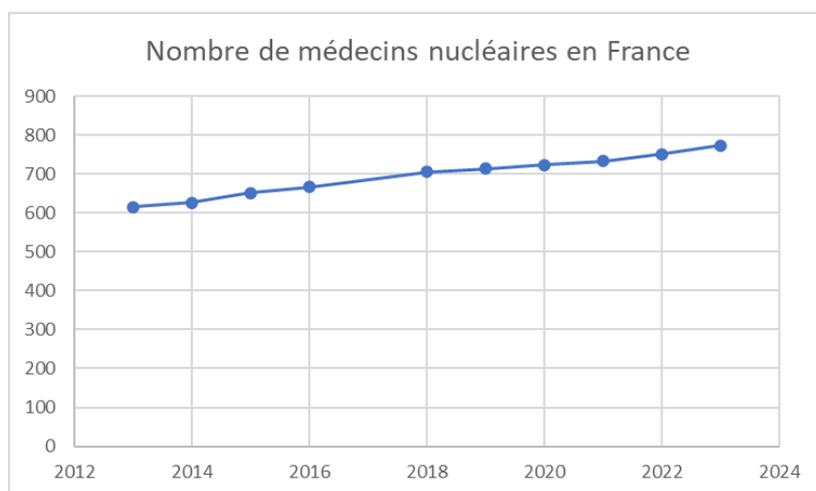


Figure 7 : Evolution du nombre de médecins nucléaires en France (sources : atlas démographiques du CNOM).

Ces quelques données démographiques doivent souligner, face à une demande croissante, l'importance de l'optimisation de l'interprétation des examens, notamment en terme d'efficience, dont le temps accordé à cette tâche. Certains outils sont à l'étude, comme les solutions d'IA.

IV. Intelligence artificielle : bases théoriques et applications en médecine nucléaire

4.1 Définitions : intelligence artificielle, machine learning et deep learning

4.1.1 Intelligence artificielle (IA)

L'IA est une branche des sciences informatiques, dont l'objectif est de créer des systèmes ayant pour but d'opérer des tâches nécessitant habituellement l'intelligence humaine (39), dont l'apprentissage. C'est un terme large qui regroupe plusieurs champs et techniques, dont le machine learning.

4.1.2 Machine learning

Le machine learning est l'approche mathématique et statistique permettant à un système d'apprendre à partir de données extraites, dans le but de résoudre des tâches, sans en avoir été préalablement et explicitement programmé pour chacune.

L'extraction de données (ou index) se fait en général au moyen de l'expertise humaine, qui discerne et encode leurs caractéristiques, avant de les soumettre à une analyse statistique qui les organise en clusters en fonction de ces mêmes caractéristiques. Pour des tâches plus complexes, l'extraction d'index peut faire appel à du deep learning, qui est une forme de machine learning.

En imagerie, les données extraites peuvent correspondre à des caractéristiques morphologiques et spatiales (index de forme 3D et localisation), des analyses de texture et la reconnaissance de motifs (matrices de texture) ou une intensité de voxels. Ce sont des index radiomiques dits « classiques », dont l'extraction peut être réalisée via des logiciels comme Pyradiomics ou LIFEEx (40), utilisés pour la recherche de biomarqueurs.

Une fois les données extraites, le machine learning opère selon un apprentissage supervisé avec différents algorithmes, pouvant être entraînés (40) :

- Prédicteurs linéaires : régressions linéaires, logistiques (LASSO), elastic net...
- Arbres de décisions : forêts aléatoires, boosting...
- Méthode des plus proches voisins
- Machines à vecteurs support
- Deep learning : réseaux de neurones

Il existe deux autres types de jeux de données, en plus de ceux utilisés pour l'entraînement des algorithmes. Un pour l'optimisation de l'algorithme, et un jeu de test indépendant permettant sa validation finale et sa généralisation (40).

En imagerie médicale, les données d'entraînement extraites peuvent passer par une étape de normalisation (par exemple pour rendre comparables des images issues de patients et avec protocoles d'acquisitions variés) et de réduction de dimension (pour éviter le surentraînement), avant l'apprentissage supervisé.

4.1.3 Deep learning

4.1.3.1 Apprentissage de représentations (feature learning ou representation learning)

Le deep learning est une technique plus récente, marquant une évolution clé du machine learning, dérivant de l'apprentissage de représentation (39). Les données et caractéristiques ne sont alors plus extraites par un algorithme en fonction de l'expertise humaine, mais directement par le système, capable alors de traiter des données brutes et d'en déduire les clusters.

Cela permet au système informatique d'apprendre et d'optimiser cet apprentissage des caractéristiques de manière autonome en limitant l'intervention humaine.

4.1.3.2 De la représentation des réseaux de neurones, à l'architecture d'un réseau de neurones convolutifs (convolutional neural networks ou CNN)

Les réseaux de neurones servant de base au CNN ont été largement étudiés dès les années 50 (39). Ceux participant à l'élaboration des CNN s'inspirent des réseaux neuronaux dans le cortex visuel des animaux, dans lequel les réponses (= sortie ou output) aux stimuli visuels (= entrée ou input) résultent de l'interaction entre de nombreuses couches de systèmes neuro-dendritiques et synaptiques, excitateurs et/ou inhibiteurs, et ce plus ou moins simultanément. La réponse dépend de l'atteinte d'un seuil dans le corps cellulaire, au-delà duquel un potentiel d'action est enregistré.

Ce schéma simplifié a été repris pour mettre en œuvre les technologies de deep learning, dont l'essor a pris de l'ampleur dès 2012 (39,41). Le terme « deep » renvoie à la notion de couches profondes et donc multiples, selon un réseau de neurones convolutifs (CNN).

D'un point de vue schématique, l'information est traitée par des couches de neurones artificiels (ou nœuds) successifs interconnectés entre eux. Les deux couches extrêmes de cette représentation correspondent respectivement aux entrées (= input layers, qui sont des données brutes plus ou moins complexes comme une intensité de voxel ou des indices de texture) et sorties (= output layers, qui sont des valeurs cibles comme des données classifiées). Les couches profondes permettent les traitements intermédiaires des données selon les processus d'inférence (39). L'objectif étant d'obtenir une prédiction ayant une forte probabilité d'appartenir à la classe

« réelle » de l'entrée, ce modèle s'appuyant en grande partie sur des fonctions de probabilité bayésiennes.

4.1.3.3 Optimisation d'un système

Les bases fondamentales de l'IA et in fine du deep learning reposent sur des principes mathématiques de minimisation de l'erreur. L'objectif étant de proposer une architecture de CNN permettant de prédire le plus précisément la ou les sortie(s) fiable(s).

A une valeur prédite correspond une ou plusieurs entrées. Dans un modèle humain standard, ce sont des observations simples qui produisent ces prédictions à partir de données brutes, dont la distribution peut être approchée par une fonction mathématique, qu'il s'agira de trouver. Cela se fait en appliquant plus ou moins aléatoirement des poids à des paramètres d'une fonction, et en calculant la distance (ou coût) entre la prédiction de cette fonction et sa distribution réelle. Tant que ce coût est élevé, il y a une erreur (ou un défaut de précision) élevée, et les paramètres sont ajustés et mis à jour.

Cette correction repose sur la rétropropagation du gradient, et est répétée autant de fois que nécessaire, jusqu'à obtenir les paramètres qui obtiennent la fonction qui explique au mieux la distribution des prédictions et optimiser la précision du modèle : c'est la minimisation du coût (39). Dans le deep learning, la fonction peut être approchée par les couches profondes de neurones (ou nœuds) intermédiaires, et la minimisation de l'erreur (ou optimisation) s'opère en faisant varier le « poids » des paramètres (selon l'importance de leur contribution dans l'erreur).

Différents algorithmes d'optimisation sont utilisés, comme par exemple SGD (Stochastic gradient descent), Adam, Adagrad...(42).

4.1.3.4 Convolution et application dans un réseau de neurones en U-NET

Un noyau de convolution agit sur une image comme un filtre (passe-haut ou passe-bas), dont la finalité est à la fois (39,43) :

- De réduire les dimensions d'une image pour faciliter les calculs du système et éviter un sur-apprentissage.
- D'extraire plus facilement des caractéristiques de forme ou des bords d'une image, où l'objectif sera d'opérer des tâches de segmentation et de classification.

Dans un réseau de neurones convolutifs, une grande partie des neurones artificiels des couches profondes se comportent tels des noyaux de convolution. C'est l'entraînement du modèle qui permet d'optimiser le poids de leurs éléments. Un des procédés de convolution est le « max

pooling », qui permet de réduire les dimensions d'une image sans en altérer ses caractéristiques principales.

L'architecture courante d'un système d'imagerie pour les tâches de segmentation est le U-NET, qui permet le traitement d'images bi-dimensionnelles, et est composé de deux chemins :

- Encodeur de contraction, composé d'une succession de couches de convolution et de « max pooling » : réduit la dimension de l'image et en détecte les caractéristiques.
- Décodeur avec des couches de déconvolution transposée : évite une perte importante de données suite à la contraction et permet une localisation précise des objets de l'image. Il restitue également une taille de l'image identique à l'initiale.

Ce système est complété par un goulot d'étranglement, permettant de relier les informations entre l'encodeur et le décodeur.

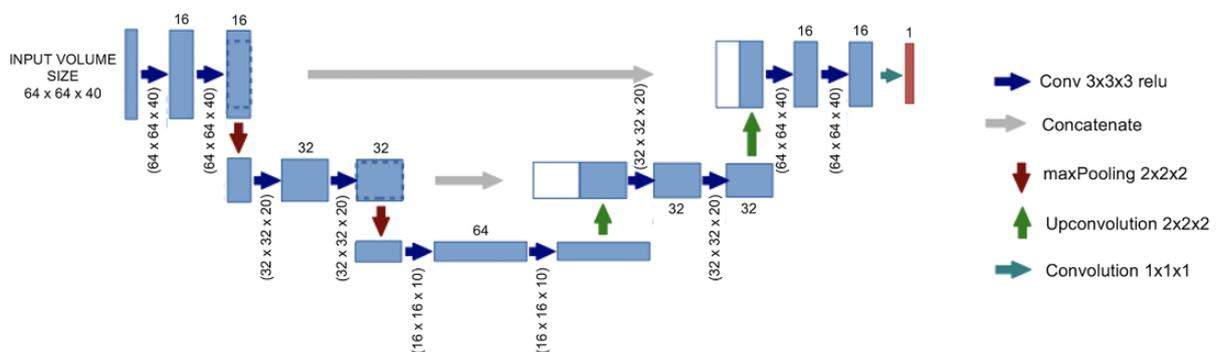


Figure 8 : Exemple d'un CNN avec une architecture en U-Net en trois couches (44).

Il existe des variantes du système U-Net :

- Le 3D-U-Net (45,46), qui permet de traiter des images tri-dimensionnelles, en remplaçant les pixels par des voxels.
- Le U-Net++ (« Nested U-Net »), permettant d'augmenter la précision de la segmentation, avec un coût d'entraînement élevé (en partie par un nombre de paramètres étudiés plus conséquent).

4.2 Solutions d'intelligence artificielle en TEP

4.2.1 SubtlePET™

L'enjeu de nombreux centres est de trouver des solutions pour optimiser les acquisitions, en assurant une sécurité dosimétrique supplémentaire pour le patient (réduction d'activité injectée, diminution du temps d'examen), compatible avec une qualité d'interprétation sans perte d'information.

SubtlePET™ (Subtle Medical, Stanford, États-Unis) est une solution d'IA permettant une diminution de la dose injectée aux patients avec un temps d'acquisition diminué. Elle comprend une architecture en 2.5 dimensions de CNN en U-Net, dont les entrées correspondent à des images de qualité dégradée (soit par réduction du temps d'acquisition, ou de la dose injectée au patient), et les sorties à des images de qualité quasiment entièrement restituée (proche d'une image réalisée avec une pleine dose et/ou de durée non diminuée).

Une diminution de la dose injectée entraîne une majoration du bruit de fond : dans une étude de 2022, Bonardel et al. ont permis de mettre en évidence qu'une diminution de moitié de l'activité initiale injectée de 18F-FDG au patient associée à un post-traitement de débruitage par SubtlePET restituait à la fois des images de qualité (rapport signal/bruit et perte d'informations) comparable à celles effectuées à pleine dose sans post-traitement, et des SUVmax comparables, sur des machines TEP de dernière génération (SiPM et TOF). Ces résultats n'étaient pas retrouvés en cas d'une diminution de 66 % de la dose (47).

Une autre étude par Weyts et al. testait l'algorithme sur des acquisitions dont la durée était divisée par 2 (45 secondes / position de lit), afin de voir si la qualité d'image était dégradée sans l'application du débruitage, et comparable à un examen effectué sans réduction de durée d'acquisition (90 secondes / position de lit), sur des caméras SiPM. Elle montrait un indice visuel de qualité comparable à un examen réalisé sans réduction de durée d'acquisition et des bonnes performances semi-quantitatives (48).

4.2.2 AUTOID et PAIRE

4.2.2.1 AUTOID

Auto-ID est une solution d'IA, développée comme une extension de « Lesion Scout » par Siemens (syngo.via MM Oncology).

« Lesion Scout » propose une segmentation automatique du volume des lésions selon des critères :

- D'identification de lésion, en choisissant comme paramètre le SUVmax, le SUVpeak, PERCIST (foie, aorte descendante), ou une valeur définie manuellement. A noter que le PERCIST (PET response criteria in solid tumor) foie, largement utilisé, correspond à une valeur de $1.5 \times \text{SUVmean hépatique} + 2\text{DS}$ (21).
- De segmentation, en utilisant un seuil en pourcentage du SUVmax local, le seuil défini dans les critères d'identification de lésions ou défini manuellement.

A partir de cette technique, AUTOID propose de sélectionner les lésions à inclure ou non dans la prise en compte des limites tumorales, par exemple en excluant les activités correspondant à une activité physiologique (comme l'excrétion urinaire, l'activité neurologique/cardiaque normale...) ou non tumorale.

L'architecture du CNN utilisé comprend une alternance de noyaux de convolution et de pooling maximum, avec comme inputs les reconstructions multi-planaires TEP et TDM, le MIP (maximum intensity projection) du TEP, la position sur l'atlas, et comme outputs la prédiction de la classification (tumoral à inclure ou activité à exclure) et de la localisation.

Quelques études ont étudié l'utilisation de cette solution d'IA en pratique. Karimjee et al. ont montré la reproductibilité, la fiabilité et le gain en temps de travail dans l'utilisation d'AUTOID pour la mesure du TMTV et du TLG chez des patients atteints de LBDGC (21). De même, Capobianco et al. ont montré une bonne corrélation des TMTV de LBDGC, segmentés et mesurés par l'IA, par rapport à des TMTV réalisés par des experts (49).

4.2.2.2 PAIRE

PAIRE est une solution d'intelligence artificielle récemment mise au point dans le service de médecine nucléaire du CHU Henri-Mondor à Créteil (AP-HP), et propose, au même titre qu'AUTOID, de réaliser des segmentations et des quantifications automatiques (SUV, MTV et TLG) (50).

Il s'agit également d'un réseau de neurones convolutifs, qui présente une architecture U-Net en trois dimensions basée sur le langage de programmation Python (51).

PAIRE a, à ce jour, été peu étudié. Blanc-Durand et al. ont étudié son utilisation dans la segmentation volumique des gliomes en TEP-FET (44). Mechti et al ont évalué les performances à la fois d'AUTOID et PAIRE dans la détection d'anomalie et de segmentation dans une cohorte monocentrique du cancer du poumon, et ont montré notamment une bonne corrélation avec la vérité terrain pour les deux logiciels dans le calcul de volumes (52). Van Der Gucht et al. ont montré des résultats en faveur d'une haute sensibilité de détection, des excellentes capacités de quantification, et une rapidité des calculs pour la détection et la quantification automatique de lésions (50). Enfin, Blanc-Durand et al. ont montré la pertinence de l'utilisation de PAIRE dans la détection et la segmentation des lésions de LBDGC, malgré une légère sous-estimation du TMTV (51).

V. Objectif

Les solutions commerciales de détection et de contournage de lésion en TEP basées sur l'IA se présentent avec comme double avantage de (1) permettre un gain de temps médical substantiel pour la mesure du TMTV et (2) constituer un filet de sécurité épaulant le médecin pour la détection des lésions. A ce jour, la littérature scientifique évaluant l'apport et la robustesse de telles solutions reste pauvre.

Cette étude a pour objectif d'évaluer les performances de détection de lésions sur la TEP/TDM au [¹⁸F]FDG par deux algorithmes d'IA utilisant un réseau de neurones convolutifs, AUTOID et PAIRE, lorsqu'ils sont utilisés seuls ou en complément de la lecture humaine.

MATERIALS AND METHODS

I. Population

One hundred and fifty-one [¹⁸F]FDG-PET/CT of 151 patients from three different French institutions were retrospectively analyzed (121 from the Eugène Marquis center of Rennes, 20 from the Antoine Lacassagne center of Nice, and 10 from the University hospital of Amiens-Picardy).

Ninety-four (62%) of these patients performed their [¹⁸F]FDG-PET/CT between January 2015 and July 2019, before the start of immune checkpoints inhibitors (anti-PD-1 alone or in combination with ipilimumab) in a first or second line of treatment for metastatic melanoma. The other 57 (38%) of patients performed their [¹⁸F]FDG-PET/CT between October 2018 and October 2021, before being treated with CART-cells as part of the management of large B-cell lymphoma.

Non-inclusion criteria were unrecoverable or uninterpretable [¹⁸F]FDG-PET/CT, people under guardianship or curatorship or deprived of liberty.

No patient had multiple PET/CT analyzed.

II. Technical features of PET/CT

The acquisition field was “whole-body” for 50 (33.1%) patients, while it spread from the vertex to mi-tight for the remaining 101 (66.9%). For 40 (26%) of them, the CT was performed after injection of iodinated contrast agents.

One hundred and twenty-three patients (81.5%) performed their PET/CT on an integrated Biograph mCT system, 2 (1.3%) on a Biograph Vision (Siemens Healthineers), 13 (8.6%) on a Discovery ST and 12 (7.9%) on a Discovery MI (GE HealthCare), and one (0.7%) on a Vereos (Philipps).

The mean \pm SD injected activity was 226.2 ± 58.0 MBq, corresponding to 3.1 ± 0.4 MBq/kg (53). The injection-acquisition delay was 69.2 ± 16.7 minutes.

III. Detection of lesions : the 4 studied methods

Four different methods for detection of lesions were used.

First, the manual human detection (M1) stands for the current standard of care. It was performed by senior nuclear medicine physicians, using the LIFEx software v5.10 and v7.20 (54).

Then, three automated detection methods were used to be compared to M1. Two of these automated methods were performed using the AUTOID algorithm with different identification settings with either an absolute threshold fixed at SUV 2.5 (M2) or according to PERCIST criteria (M3). For the later, an hepatic volume of interest of 14 cm³ (diameter of 30 mm) was automatically placed in the hepatic parenchyma (the reference volume was moved if automatically placed in a lesion), then, an absolute threshold of $1.5 * \text{hepatic SUV}_{\text{mean}} + 2$ standard deviations was applied.

The last method (M4) was based on the fully automated PAIRE algorithm. Unlike with AUTOID, there was no preliminary step of defining threshold settings. Subsequently, target volumes were generated as a mask (.nii files).

For each method, delineation was performed using relative thresholds of 40% and 41% of the SUV_{max} of each lesion, respectively for patients with melanoma and lymphoma, according to current guidelines (26,55,56). The volumes provided by each method as suspect of lesions are called volumes of interest (VOI). Multifocal disseminated lesions were numbered independently, while a diffuse involvement (for bone or peritoneum, for instance) was considered as a single lesion.

IV. Reference standard

To define the reference standard at lesion scale, all VOIs identified with at least one of the four investigated detection methods were reviewed consensually by two nuclear medicine physicians, based on PET/CT images, blind to clinical context or follow-up. Volumes generated with M1 and M4 were reviewed using the LifeX software v7.1, and those generated with M2 and M3 were reviewed with Syngo.Via VB60. (Siemens HealthineersTM, Erlangen, Germany).

Identified VOIs were labelled as “true positive” if they were considered suspect for primary or secondary neoplastic lesions (melanoma or lymphoma) according to the consensus review. If not, VOIs corresponding to non-pathological findings were considered as “false positive” findings. Additionally, VOIs corresponding to suspected other pathological findings (traumatic

lesions, infection, lesions suspicious for a second cancer ...) were labeled separately in a third category.

V. Statistical analysis

Statistical tests were performed using MedCalc software version 12.5.0.0 (© MedCalc Software Ltd). Distribution of the number of lesions per patients was presented with the mean and the standard deviation (mean \pm SD). Number of “true positive”, “false positive”, and “false negative” lesions were pairwise compared using the paired samples Student “t” test. Comparison of proportion were performed with the Chi-squared test. The statistical tests were two-sided, and their results were considered statistically significant if $p < 0.05$.

RESULTS

I. Sample characteristics

A total of 1544 “true positive” lesions in 151 patients were analyzed, including 1020 lesions in 94 melanoma patients and 524 lesions in 57 lymphoma patients.

The average number of lesions per patient was 10.2 ± 17.0 [min: 0 ; max: 115].

A total of 128 (84.8%) patients presented at least one lesion, while 23 patients had no lesions according to the standard reference.

The lymph nodes were the most frequently affected organs (Figure 9), with 94 (62.3%) patients involved, compared to 54 (35.8%) patients for the skin and subcutaneous tissue, 28 (18.5%) for the lungs, 25 (16.6%) for the bones, 23 (15.2%) for the liver, 21 (13.9%) for the muscles, 20 (13.2%) for the peritoneum, 8 (5.3%) for the adrenals, 7 (4.6%) for the pleura and the spleen, respectively, 5 (3.3%) for the thyroid, 4 (2.6%) for the pancreas and the brain, respectively, 3 (2.0%) for the intestines, 1 (0.7%) for the testis, and 12 (7.9%) for all the others combined.

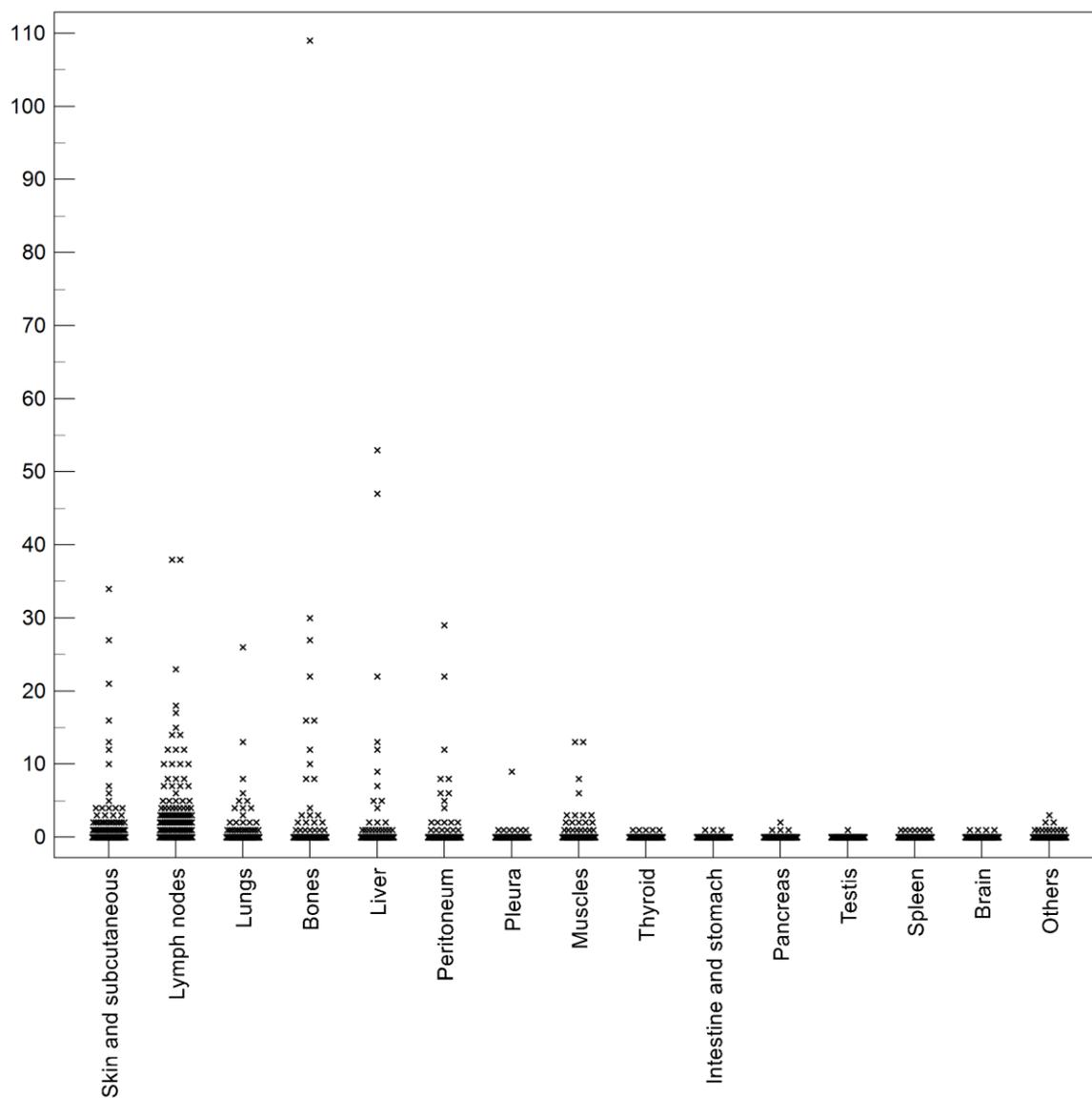


Figure 9 : Distribution of the number of lesions for each organ, per patient, according to the reference standard.

II. Considering each method independently

The diagnostic performances of each detection method taken independently are detailed in **Table 7** and **Figure 10**.

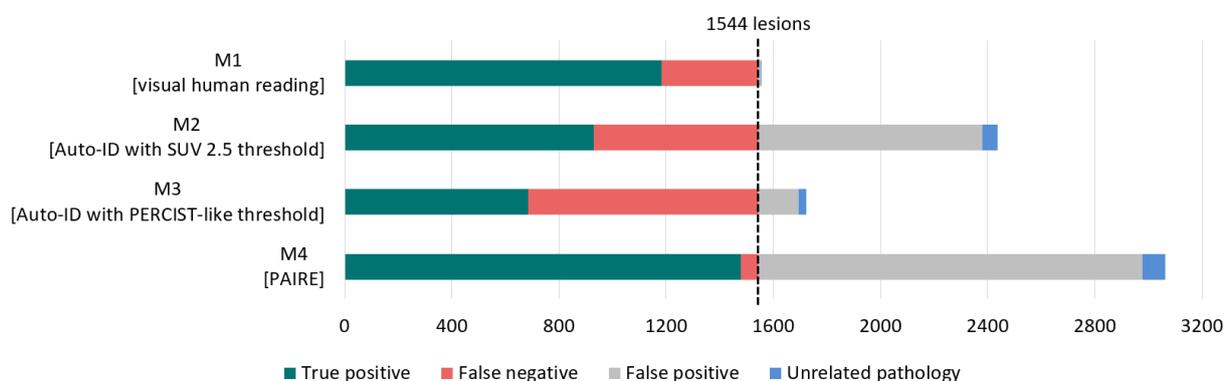


Figure 10 : Number of volumes of interest detected with each method independently.

	Detected VOI	True positive	False positive	Sensitivity [95%CI]	PPV [95%CI]
M1 [visual human reading]	1194	1182	10	76.6% [74.4, 78.7]	99.2% [99.1, 99.2]
M2 [AUTOID with SUV 2.5 threshold]	1822	929	837	60.2% [57.8, 62.6]	52.6% [51.9, 53.4]
M3 [AUTOID with PERCIST-like threshold]	864	684	151	44.3% [41.8, 46.8]	81.9% [81.3, 82.5]
M4 [PAIRE]	2997	1478	1435	95.7% [94.6, 96.7]	50.7% [50.7, 50.8]

Table 7 : Diagnostic performances of each method independently, per lesions.

In a pairwise comparison, M1 detected significantly more “true positive” lesions (mean 7.8 ± 12.1) than M2 (mean 6.2 ± 10.4 , $p < 0.001$) and than M3 (mean 4.5 ± 8.6 , $p < 0.001$), but significantly less than M4 (mean 9.8 ± 16.7 , $p < 0.001$).

M2 detected significantly more “true positive” lesions (mean 6.2 ± 10.4) than M3 (mean 4.5 ± 8.6 , $p < 0.001$), but significantly less than M4 (mean 9.8 ± 16.7 , $p < 0.001$).

M3 detected significantly fewer “true positive” lesions (mean 4.5 ± 8.6) than M4 (mean 9.8 ± 16.7 , $p < 0.001$).

An example of one patient with mismatching detection results between different methods is presented in Figure 11.

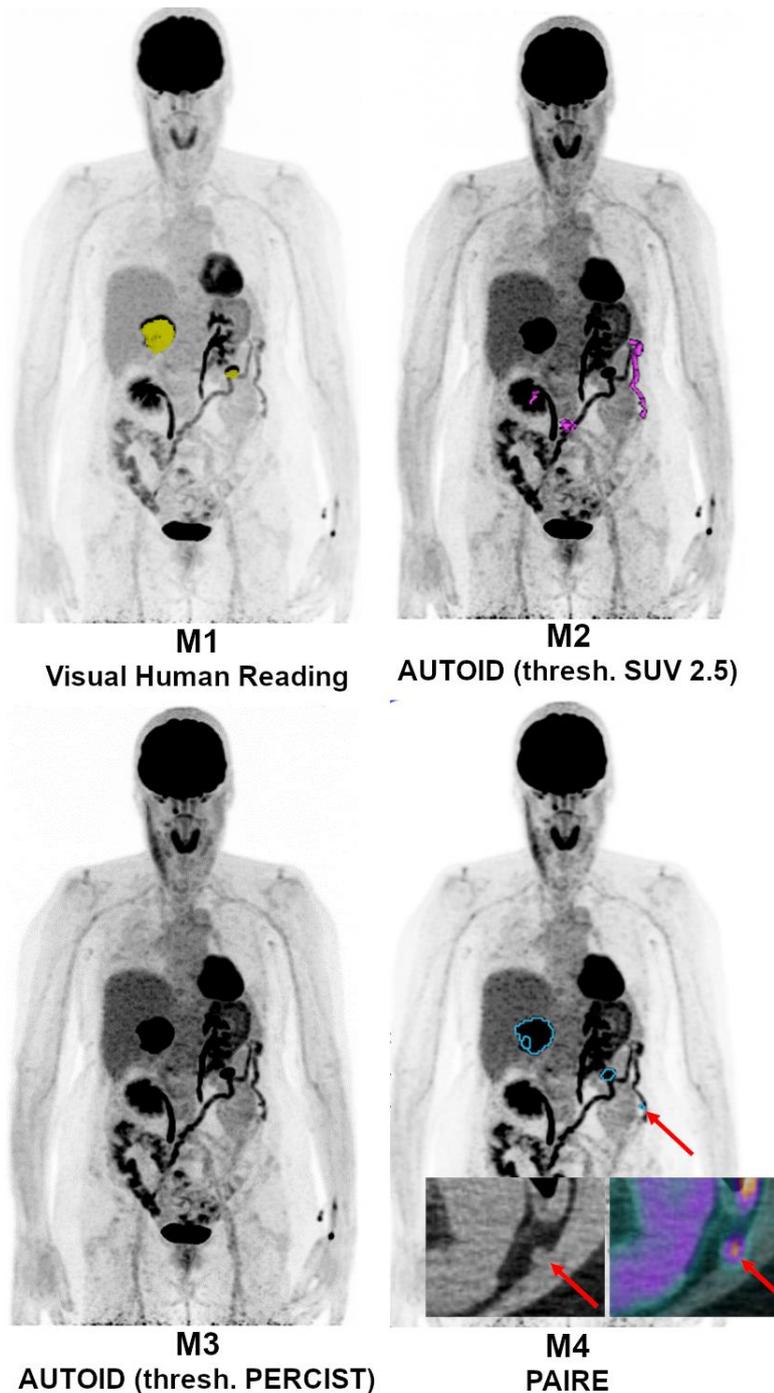


Figure 11: Example of a patient managed for melanoma for whom four evaluated methods provided mismatching VOI on $[^{18}\text{F}]\text{FDG}$ PET-CT images. The visual human reading (M1) found 2 out the 3 lesions determined with the reference standard, and no false positive VOI. The methods relying on the AUTOID (M2 and M3) algorithm revealed none of the lesions, but M2 falsely detected physiological activity from the digestive tract. The fourth methods (M4), using the PAIRE algorithm correctly identified all of the 3 lesions, including one subcentimetric parietocolic nodule (SUV_{max} 4.9) missed with M1 (red arrow).

Regarding the negative predictive value, at the patient scale, a total of 4 patients had no VOI detected with M2 (all 4 had no lesion according to reference standard), 35 with M3 (14 had no lesion according to reference standard), and 0 with M4.

Regarding subgroups, the visual human reading (M1) provided a significantly higher sensitivity for melanoma lesions than for lymphoma lesions. In contrast, regarding automated methods, only M4 presented a significantly higher sensitivity for lymphoma lesions than for melanoma lesions (Table 8).

	Melanoma patients (n = 1020 lesions)	Lymphoma patients (n = 524 lesions)	p (chi-squared)
	Sensitivity [95%CI]	Sensitivity [95%CI]	
M1 [visual human reading]	79.8% [77.2, 82.2]	70.2% [66.1, 74.1]	<0.001*
M2 [AUTOID with SUV 2.5 threshold]	58.5% [55.4, 61.6]	63.4% [59.1, 67.5]	0.063
M3 [AUTOID with PERCIST-like threshold]	42.8% [39.8, 46.0]	47.1% [42.8, 51.5]	0.107
M4 [PAIRE]	94.7% [93.1, 96.0]	97.7% [96.0, 98.8]	0.006*

Table 8 : Comparison of diagnostic sensitivities of each method, for both “melanoma” and “lymphoma” subgroups, per lesion. (* : $p < 0.05$)

III. Considering each AI-based method (M2, M3 and M4) as a completing human reading (M1)

Diagnostic performances of each combination method as usable in clinical workflow are detailed in **Table 9** and **Figure 12**.

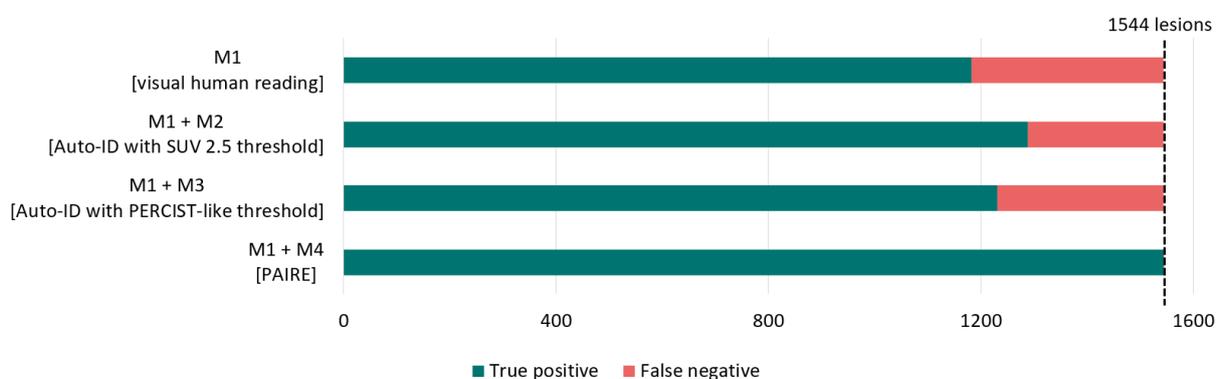


Figure 12 : Repartition of true positive and false negative lesions of each AI-based method as completing human reading.

	True positive	False negative	Sensitivity [95%CI]
M1 [visual human reading]	1182	362	76.6% [74.4, 78.7]
M1+M2 [AUTOID with SUV 2.5 threshold]	1288	256	83.4% [81.5, 85.2]
M1+M3 [AUTOID with PERCIST-like threshold]	1231	313	79.7% [77.6, 81.7]
M1+M4 [PAIRE]	1543	1	99.9% [99.6, 100]

Table 9 : Diagnostic performances of each AI-based method as completing human reading, per lesions.

When pairwise compared, all combinations of methods detected significantly more lesions per patient than M1 alone. “M1 + M2” detected a mean of 8.5 ± 13.6 “true positive” lesions per patient ($p < 0.001$), “M1 + M3” reported 8.2 ± 12.8 lesions ($p = 0.004$), and “M1 + M4” found 10.2 ± 17.0 lesions ($p < 0.001$).

When pairwise comparing combination of methods “M1 + M4” detected significantly more lesions than “M1 + M3” ($p < 0.001$) and “M1 + M2” ($p < 0.001$), respectively. While when considering AUTOID as completing the human, “M1 + M2” detected significantly more lesions than “M1 + M3” ($p < 0.001$).

IV. False negative lesions

According to the human reading (M1) 362 lesions were missed, including 41% (48/116) of the peritoneal lesions, 32% (75/231) of cutaneous and subcutaneous lesions, 29% (20/70) of the muscular lesions, 22% (22/101) of pulmonary lesions, 22% (63/286) of bone lesions, 19% (89/481) of the nodal lesions, and 19% (37/193) of the hepatic lesions (Figure 13).

M2 provided 615 false negative lesions, including all the 4 cerebral lesions, 66% (76/116) of the peritoneal lesions, 64% (148/231) of the cutaneous and subcutaneous lesions, 60% (42/70) of the muscular lesions, 40% (4/10) of the adrenal lesions, 35% (101/286) of the osseous lesions, 33% (33/101) of the pulmonary lesions, 29% (140/481) of the nodal lesions, and 25% (49/193) of the hepatic lesions.

M3 provided 860 false negative lesions, including all the 4 cerebral lesions, 78% (91/116) of the peritoneal lesions, 74% (172/231) of the cutaneous and subcutaneous lesions, 67% (47/70) of the muscular lesions, 67% (68/101) of the pulmonary lesions, 60% (6/10) of the adrenal lesions, 51% (146/286) of the osseous lesions, 49% (234/481) of the nodal lesions, and 37% (71/193) of the hepatic lesions.

M4 provided a total of 66 false negative lesions, including 20% (2/10) of the adrenal lesions and 6% (22/481) of the nodal lesions. None of the 4 cerebral lesions was missed.

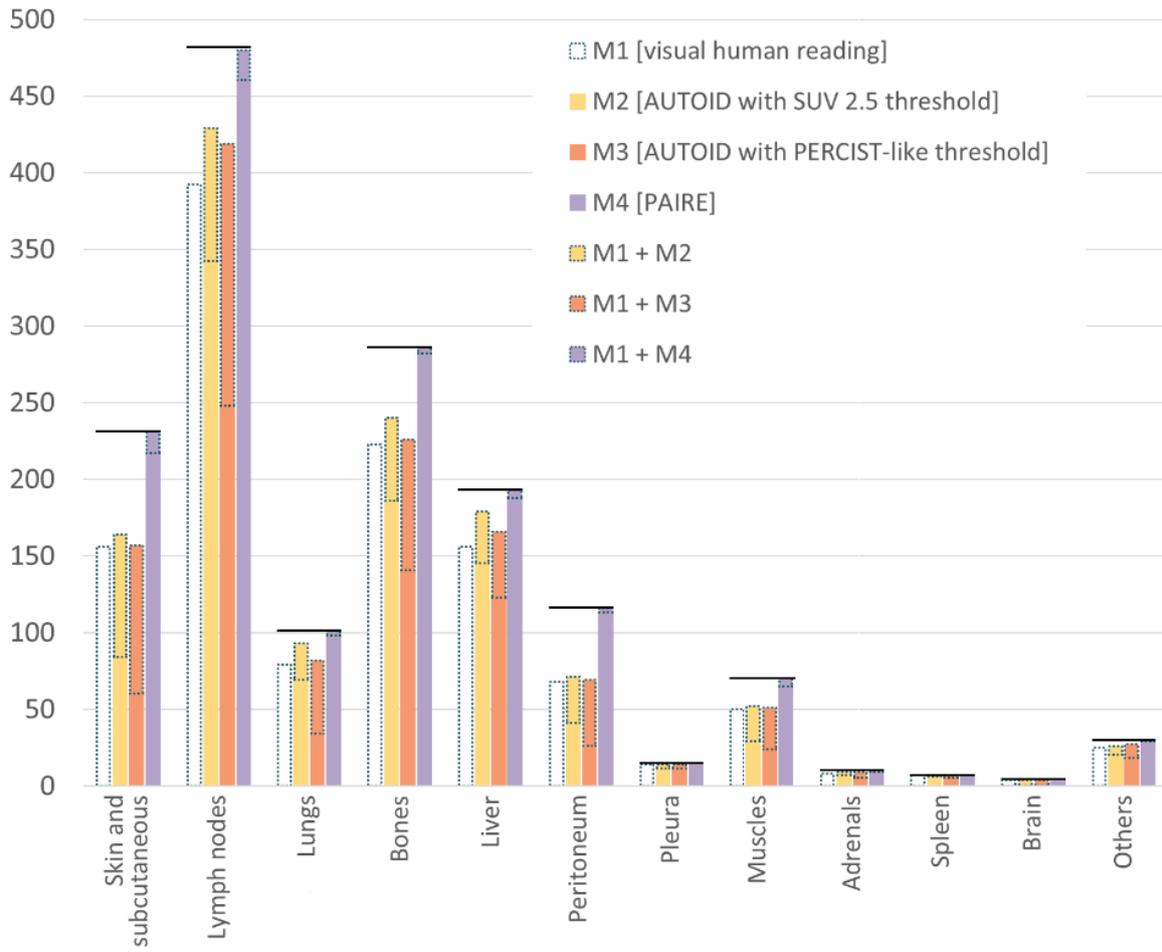


Figure 13 : Number of lesions detected with each method compared to the reference standard (black horizontal lines), classified by organs.

V. False positive VOIs

Regarding automated methods, M4 reported the most numerous false positive VOIs (1435), that were located in a large variety of tissues (Figure 14). M2 provided 837 false positive VOIs, located mainly in the vessels and the bone marrow. In contrast, M3 provided a few false positive VOIs (151), mainly located in the bone marrow.

Additionally, M2, M3 and M4 revealed unexpected pathologies (infections, second primary neoplastic lesion ...) with 56, 29 and 84 VOIs, respectively.

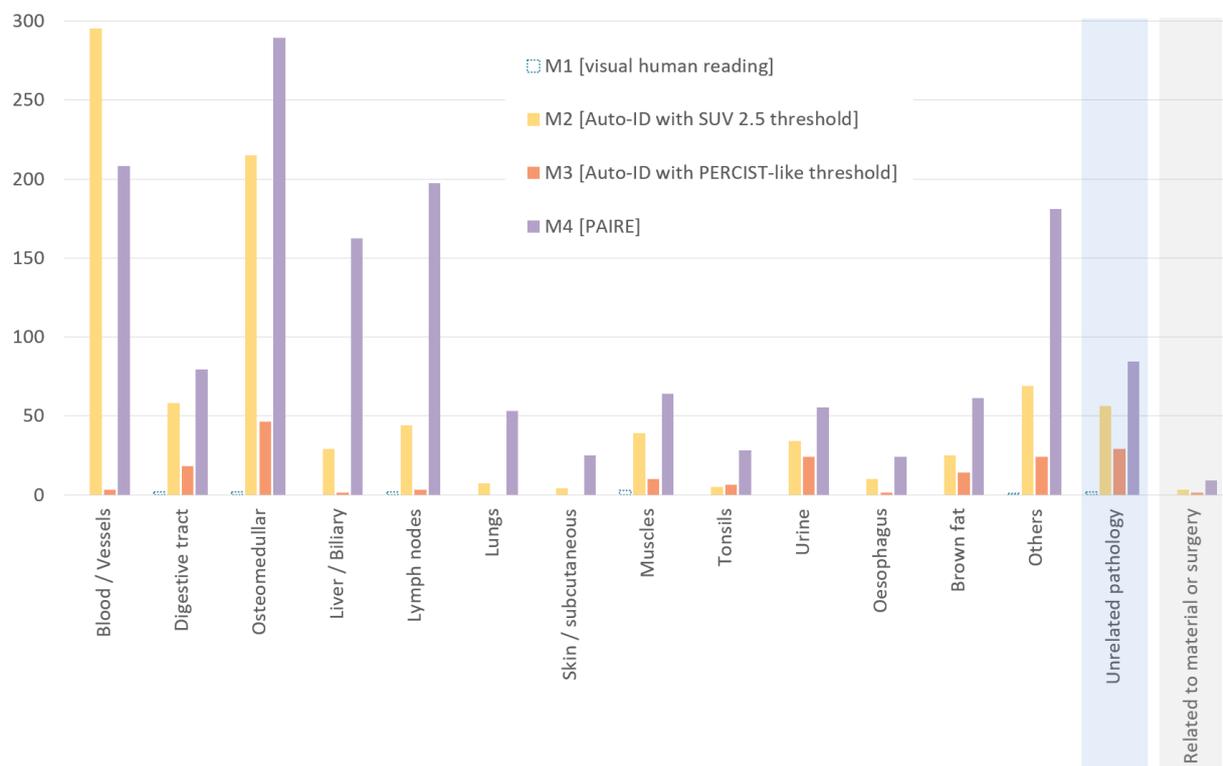


Figure 14 : Number of false positive VOIs, classified by organs.

DISCUSSION

I. Main results

This study aimed to compare the performance of lesions detection on [¹⁸F]FDG PET/CT using two AI algorithms based on convolutional neural networks (AUTOID and PAIRE), when employed either independently or in conjunction with human interpretation. The ultimate purpose was to select the most suitable solution to assist the nuclear medicine physician as a safety net for lesions detection.

When comparing method individually, PAIRE (M4) detected significantly more lesions than any of the other three methods (M1 : manual, M2 : using AUTOID with a 2.5 threshold, and M3 using AUTOID with PERCIST-like threshold). The sensitivity of PAIRE (M4) reached 95.7%, compared to 76.6%, 60.2% and 44.3% for M1, M2 and M3, respectively, and was slightly higher in lymphoma (97.7%) patients than in melanoma patients (94.7%).

When PAIRE was completing the human reading (M1 + M4), it reached an excellent sensitivity of 99.9%, missing only one nodal lesion. It detected significantly more lesions than the two other combinations of methods using AUTOID that provided sensitivities of 83.4% and 79.7% for “M1 + M2” and “M1 + M3”, respectively. The two latter enhanced significantly but moderately the sensitivity compared to the human reading alone (M1).

Regarding the undetected lesions, AUTOID and the human reading missed many lesions, mostly located in the same organs. A substantial proportion (>15%) of lesions remained undetected in numerous tissues including the peritoneum, cutaneous and subcutaneous tissues, muscles, lungs, bones, lymph nodes and the liver. Additionally, AUTOID missed all the 4 cerebral lesions. In contrast, M4 did not miss more than 10% of lesions in any organ, except the adrenals (2/10).

The large amount of false positive VOIs is a major limitation for automated detection methods. PAIRE (M4) provided the highest ratio of false positivity, with 48.8% of generated VOIs being related to physiological activity, compared to 45.9% with M2 and 17.5% with M3. For all, the physiological vascular and osteomedullar activity were the main cause of false positivity. Additionally, PAIRE (M4) reported a large amount of false positive VOIs in the liver and the lymph nodes.

II. Strengths and limitations of the study

This multicentric study is the first to investigate the performances of these two algorithms, across different devices in patients with two types of pathologies (melanomas and DLBCL). The large sample of patients presenting more than 1500 lesions allowed us to obtain robust statistical results.

This work has some limitations that should be highlighted. First, the subjective human consensus was used as the reference standard to differentiate lesion from false positive VOIs. Even considering the experience of the nuclear medicine physicians, classification errors remain possible. Moreover, only lesions detected by at least one of the four methods were analyzed. Some lesions might have been undetected by all methods (i.e., weakly or non-avid for [¹⁸F]FDG) and not being taken into account in the statistical analysis. Therefore, the sensitivity may have been overestimated for each of the four methods. Second, when using AUTOID (M2 and M3), only VOIs greater than 0.5 mL (diameter of 10 mm) were kept, independently of the thresholding method. This may have limited the detection of small lesions and underestimated the sensitivity. In contrast, the use of PAIRE does not require the user to define a cut-off for VOI sizes, and so doesn't restrict the detection of small lesions. Finally, in the same manner, the use of AUTOID required to set an intensity threshold for detection. Empirically, thresholds of SUV 2.5 (M2) based on the PERCIST criteria (M3) were chosen based on previous studies. Different cut-off values could have led different results. The lack of recommendations on the use of AUTOID leads to variability in the definition of pre-processing parameters. As a result, this hampers reproducibility between readers and between centers, which is one of the main benefits expected from the use of AI.

III. Literature

The literature on the topic is very emerging and mainly focusing on the application of such algorithms for lesions' segmentation. In 2020, Blanc-Durand et al. studied the use of PAIRE for segmentation in 733 patients treated for DLBCL, compared to manual segmentation after adaptive thresholding at 41% of SUV_{max} (51). The authors showed good performances. The similarity between two datasets was moderate-high with a Dice coefficient (\pm SD) of 0.73 (\pm 0.20).

Mechti et al. studied PAIRE and AUTOID in a single-center sample including [¹⁸F]FDG PET/CT scans performed for initial staging of 66 patients with pulmonary neoplasms (52). Unlike our work, there was no manual human contouring of lesions, and the ground truth was

established on AUTOID predictions after manual correction. The AUTOID detection technique was using a PERCIST-like threshold, with VOIs segmented based at 42% of the local SUV_{max}. VOIs with volumes less than 2 cm³ were excluded. The published results are consistent with ours, with a reported sensitivity of 92.7% with PAIRE and 49.3% using AUTOID. The “false positives/true positives” ratio reached 50 % for PAIRE and 48.9 % for AUTOID.

Van der Gucht et al. studied the performance of PAIRE in a population of 97 patients from 3 different centers who performed [¹⁸F]FDG-PET/CT scans for oncological indications (50). The classification of VOIs was also done by the review nuclear medicine physicians. This study showed an excellent sensitivity of the algorithm, reaching 97.4%, with a moderate PPV of 67.4%.

IV. Perspectives

PAIRE seems to show a high sensitivity in lesions detection, but with a predictive value inferior to other techniques. Associated with manual detection by an expert physician, using PAIRE enabled the detection of all of the 1544 lesions except one in the present sample. The question arises as to its future role in routine practice. To date, PAIRE cannot completely replace the human eye, as evidenced by some false negatives of this algorithm, seen in classic visual detection. In light of the study results, its real advantage could lie in a robust "safety net" to ensure no target is missed.

However, several cases need to be distinguished, for example :

- Neoplastic involvement is limited, or the examination is considered "unremarkable" from a neoplastic point of view: there would be more interest in having to detect any lesion. For these cases, in our study, the number of false positives was limited, and some lesions were not manually seen and "caught" by PAIRE (for example, small peritoneal lesions or suspicious pulmonary micronodules) (Figure 11). It is certainly in these indications that this algorithm, combined with visual detection, would find the most sense at this stage.
- There are a multitude of lesions, and the relevance of describing them all in practice is questionable (not done in the vast majority of cases), especially in terms of impact on patient management. PAIRE could help in this case to find lesions in neglected organs (e.g., brain parenchyma in cases from our study) but less for an exhaustive description of each target.

On the other hand, concerning AUTOID, within the limit of detecting lesions with a volume strictly greater than 0.5 mL, its use in practice would be limited by low sensitivity, regardless of the choice of a manual threshold or segmentation according to PERCIST.

While one of the strengths of our study was to include a large number of patients, representing a fairly heterogeneous distribution of lesions, two pathologies were studied, diffuse large B-cell lymphomas and melanomas. A large study dedicated to lesion detection in a larger population in terms of neoplastic primary should be considered to approach closer to the reality of the field.

Our study aimed to investigate lesion detection performance and did not focus on the utility of volume measurement, which could be the subject of further research. Blanc-Durand et al. had highlighted, in their previously mentioned 2020 study, an underestimation of TMTV in training cohort ($-12 \text{ mL (2.8\%) } \pm 263$ ($p = 0.27$)) and external validation cohort ($-116 \text{ mL (20.8\%) } \pm 425$ ($p = 0.01$)) (51). Karimjee et al. studied the reproducibility of results and time saved in daily tasks, using AUTOID to measure metabolic volumes (TMTV and TLG) in 51 patients managed for DLBCL (21). They found an excellent reproducibility between semi-automatic and automatic methods by PERCIST thresholding and 41% of SUVmax, as long as manual correction was applied by the nuclear medicine physician (ICC of 0.99). Furthermore, in the previous Mehti et al. study, the mean metabolic volumes were, respectively, 58 and 80 mL for PAIRE and AUTOID, compared to 72 mL after correction, without significant differences ($p = 0.10$ and $p = 0.45$), and were correlated with the ground truth ($\rho = 0.81$ and 0.76 for PAIRE and AUTOID, respectively) (52). Van der Gucht et al., in the previously mentioned study, not only showed good correlation of MTV measured by PAIRE with that of physicians ($R^2 = 0.92$), but also with SUVmax values ($R^2 = 0.99$), and the major axis ($R^2 = 0.91$) (50).

Similarly, only [^{18}F]FDG was studied, but one could consider expanding the algorithm's training on PET exams of other modalities. For example, as mentioned in the introduction, Blanc-Durand et al. used a full 3D U-Net Convolutional Neural Network in automatic segmentation of gliomas in [^{18}F]FET-PET/CT in 37 patients, 26 (70%) in the training set and 11 (30%) in the validation set. Ground truth segmentations were obtained using manual delineation and thresholding ($1.3 * \text{background}$). At voxel level, great performances were obtained (Se 88%, Sp 99%, PPV 78%, and NPV 99.9%) (44). After 150 epochs (a single epoch corresponding to a review of all elements by the system), Dice Coefficient reached 0.7924 in the training set and 0.7911 in the validation set. This coefficient increased at 0.8231 after modifications of the system (morphological dilatation and fixed thresholding of the predicted U-Net mask).

As stated in the introduction, the data on the demographic evolution of physicians, in relation to the rapidly increasing number of nuclear medicine exams, imply studying the interest of applying these algorithms in terms of time saved during a work shift. Other data could specify the place of these software in practice. Karimjee et al. studied this parameter in their previously

mentioned study, demonstrating that there was a gain of about 2 minutes 30 per measurement between semi-automatic and automatic methods with correction (21).

One of the future challenges will be also to assess the impact on patient management.

The enrichment of the PET machine park mainly concerns digital equipment, studying the performance of these algorithms would warrant a comparison with their application on analog PET scans, still widely used in our territory.

One of the latest challenges we can mention, is the possibility of including a final classification step based on AI results to decrease the number of false positives, for example, by categorizing targets according to the probability of malignancy and/or the affected organ/tissue.

CONCLUSION

The use of artificial intelligence in nuclear medicine offers some usable tools in routine practice, but has only proven its performance in a limited number of studies. Our work highlights the interest such algorithms, not for replacing the physician, but for securing routine interpretation of [¹⁸F]FDG-PET/CT scan.

Among the studied algorithms, PAIRE provides the highest added value as a “security net” when associated with the expertise of the nuclear medicine physician.

Being aware of the strengths (high sensitivity) and weaknesses (high number of false positive findings) will drive the nuclear medicine community to define the place of such algorithms in routine practice in a nuclear medicine departments. .

BIBLIOGRAPHIE

1. Sweet WH, Brownell GL. LOCALIZATION OF INTRACRANIAL LESIONS BY SCANNING WITH POSITRON-EMITTING ARSENIC. JAMA. 1955;157(14):1183–1188. doi:10.1001/jama.1955.02950310009002.
2. Bozkurt MF, Virgolini I, Balogova S, Beheshti M, Rubello D, Decristoforo C, et al. Guideline for PET/CT imaging of neuroendocrine neoplasms with 68Ga-DOTA-conjugated somatostatin receptor targeting peptides and 18F-DOPA. Eur J Nucl Med Mol Imaging. août 2017;44(9):1588-601.
3. C. A. Burnham and G. L. Brownell, « A Multi-Crystal Positron Camera, » in IEEE Transactions on Nuclear Science, vol. 19, no. 3, pp. 201-205, June 1972, doi: 10.1109/TNS.1972.4326726.
4. D.A. Chesler, “Three-dimensional activity distribution from multiple positron scintigraphs,” Journal of Nuclear Medicine, 1971, 347-348.
5. M. E. Phelps, E. J. Hoffman, N. A. Mullani, C. S. Higgins and M. M. T. Pogossian, « Design Considerations for a Positron Emission Transaxial Tomograph (PETT III), » in IEEE Transactions on Nuclear Science, vol. 23, no. 1, pp. 516-522, Feb. 1976.
6. Nutt R. The History of Positron Emission Tomography. Molecular Imaging & Biology. 1 janv 2002;4(1):11-26.
7. Meyer HJ, Wienke A, Surov A. Associations between GLUT expression and SUV values derived from FDG-PET in different tumors—A systematic review and meta analysis. PLoS One. 17 juin 2019;14(6):e0217781.
8. Higashi T, Tamaki N, Torizuka T, Nakamoto Y, Sakahara H, Kimura T, et al. FDG Uptake, GLUT-1 Glucose Transporter and Cellularity in Human Pancreatic Tumors. Journal of Nuclear Medicine. 1 oct 1998;39(10):1727-35.
9. Love C, Tomas MB, Tronco GG, Palestro CJ. FDG PET of Infection and Inflammation. RadioGraphics. sept 2005;25(5):1357-68.
10. Annunziata S, Delgado Bolton RC, Kamani CH, Prior JO, Albano D, Bertagna F, et al. Role of 2-[18F]FDG as a Radiopharmaceutical for PET/CT in Patients with COVID-19: A Systematic Review. Pharmaceuticals (Basel). 10 nov 2020;13(11):377.
11. Fluorodésoxyglucose F-18 — acadpharm [Internet]. [cité 21 mars 2024]. Disponible sur: https://dictionnaire.acadpharm.org/w/Fluorod%C3%A9soxyglucose_F-18
12. Ou YN, Xu W, Li JQ, Guo Y, Cui M, Chen KL, et al. FDG-PET as an independent biomarker for Alzheimer’s biological diagnosis: a longitudinal study. Alzheimers Res Ther. 29 juin 2019;11:57.
13. Kobayashi K, Bhargava P, Raja S, Nasseri F, Al-Balas HA, Smith DD, et al. Image-guided Biopsy: What the Interventional Radiologist Needs to Know about PET/CT. RadioGraphics. sept 2012;32(5):1483-501.
14. Khalaf S, Al-Mallah MH. Fluorodeoxyglucose Applications in Cardiac PET: Viability, Inflammation, Infection, and Beyond. Methodist Debakey Cardiovasc J. 2020;16(2):122-9.
15. Lafforgue C, Lassailly G, Baillet C, Huglo D, Dharancy S. Intérêt pronostique du TEP-TDM dans la prise en charge du carcinome hépatocellulaire. Hépatogastro & Oncologie Digestive. 1 nov 2019;26(9):896-904.
16. Buvat I. Les limites du SUV. Médecine Nucléaire. avr 2007;31(4):165-72.

17. Kostakoglu L, Chauvie S. Metabolic Tumor Volume Metrics in Lymphoma. *Seminars in Nuclear Medicine*. janv 2018;48(1):50-66.
18. O JH, Lodge MA, Wahl RL. Practical PERCIST: A Simplified Guide to PET Response Criteria in Solid Tumors 1.0. *Radiology*. août 2016;280(2):576-84.
19. Sridhar P, Mercier G, Tan J, Truong MT, Daly B, Subramaniam RM. FDG PET Metabolic Tumor Volume Segmentation and Pathologic Volume of Primary Human Solid Tumors. *American Journal of Roentgenology*. mai 2014;202(5):1114-9.
20. Pak K, Cheon GJ, Nam HY, Kim SJ, Kang KW, Chung JK, et al. Prognostic Value of Metabolic Tumor Volume and Total Lesion Glycolysis in Head and Neck Cancer: A Systematic Review and Meta-Analysis. *J Nucl Med*. juin 2014;55(6):884-90.
21. Karimjee M, Delaby G, Huglo D, Baillet C, Willaume A, Dujardin S, et al. Evaluation of a convolution neural network for baseline total tumor metabolic volume on [18F]FDG PET in diffuse large B cell lymphoma. *Eur Radiol*. 4 janv 2023;33(5):3386-95.
22. Zanoni L, Bezzi D, Nanni C, Paccagnella A, Farina A, Broccoli A, et al. PET/CT in Non-Hodgkin Lymphoma: An Update. *Seminars in Nuclear Medicine*. mai 2023;53(3):320-51.
23. Brice P, De Kerviler E, Friedberg JW. Classical Hodgkin lymphoma. *The Lancet*. oct 2021;398(10310):1518-27.
24. Barrington SF, Mikhaeel NG, Kostakoglu L, Meignan M, Hutchings M, Müller SP, et al. Role of Imaging in the Staging and Response Assessment of Lymphoma: Consensus of the International Conference on Malignant Lymphomas Imaging Working Group. *JCO*. 20 sept 2014;32(27):3048-58.
25. J. Bröckelmann* P, A. Eichenauer* D, Jakob T, Follmann M, Engert A, Skoetz N. Hodgkin Lymphoma in Adults. *Dtsch Arztebl Int*. août 2018;115(31-32):535-40.
26. Cheson BD, Fisher RI, Barrington SF, Cavalli F, Schwartz LH, Zucca E, et al. Recommendations for Initial Evaluation, Staging, and Response Assessment of Hodgkin and Non-Hodgkin Lymphoma: The Lugano Classification. *JCO*. 20 sept 2014;32(27):3059-67.
27. Cottreau AS, Meignan M, Nioche C, Capobianco N, Clerc J, Chartier L, et al. Risk stratification in diffuse large B-cell lymphoma using lesion dissemination and metabolic tumor burden calculated from baseline PET/CT†. *Annals of Oncology*. mars 2021;32(3):404-11.
28. McCarten KM, Nadel HR, Shulkin BL, Cho SY. Imaging for diagnosis, staging and response assessment of Hodgkin lymphoma and non-Hodgkin lymphoma. *Pediatr Radiol*. oct 2019;49(11):1545-64.
29. Casasnovas RO, Bouabdallah R, Brice P, Lazarovici J, Ghesquieres H, Stamatoullas A, et al. PET-adapted treatment for newly diagnosed advanced Hodgkin lymphoma (AHL2011): a randomised, multicentre, non-inferiority, phase 3 study. *The Lancet Oncology*. 1 févr 2019;20(2):202-15.
30. Rossi C, André M, Joubert C, Fortpied C, Brice P, Deau-Fisher B, et al. Stage IIb High Risk Hodgkin Lymphoma Treated in the H10 and AHL2011 Trials: Similar Efficacy of Both Strategies and Prognostic Impact of Baseline Tmtv and PET2 Response. *Blood*. 13 nov 2019;134:128.
31. Le Gouill S, Ghesquières H, Oberic L, Morschhauser F, Tilly H, Ribrag V, et al. Obinutuzumab vs rituximab for advanced DLBCL: a PET-guided and randomized phase 3 study by LYSA. *Blood*. 29 avr 2021;137(17):2307-20.

32. Meignan M, Gallamini A, Itti E, Barrington S, Haioun C, Polliack A. Report on the Third International Workshop on Interim Positron Emission Tomography in Lymphoma held in Menton, France, 26–27 September 2011 and Menton 2011 consensus. *Leukemia & Lymphoma*. oct 2012;53(10):1876-81.
33. Itti E, Blanc-Durand P, Berriolo-Riedinger A, Kanoun S, Kraeber-Bodéré F, Meignan M, et al. Validation of the Δ SUVmax for Interim PET Interpretation in Diffuse Large B-Cell Lymphoma on the Basis of the GAINED Clinical Trial. *Journal of Nuclear Medicine*. 1 nov 2023;64(11):1706-11.
34. Alcantara M, Philippe L. Lymphome diffus à grandes cellules B. *Hématologie*. 1 janv 2017;23(1):27-34.
35. Vercellino L, Rivas A, Baroudjian B, Lebbé C, Merlet P. Rôle de la TEP FDG dans l'évaluation de l'extension locorégionale et à distance du mélanome. *Médecine Nucléaire*. nov 2020;44(5-6):305-12.
36. Xing Y, Bronstein Y, Ross MI, Askew RL, Lee JE, Gershenwald JE, et al. Contemporary Diagnostic Imaging Modalities for the Staging and Surveillance of Melanoma Patients: a Meta-analysis. *JNCI Journal of the National Cancer Institute*. 19 janv 2011;103(2):129-42.
37. Arrangoiz R, Papavasiliou P, Stransky CA, Yu JQ, Tianyu L, Sigurdson ER, et al. Preoperative FDG-PET/CT Is an Important Tool in the Management of Patients with Thick (T4) Melanoma. *Dermatology Research and Practice*. 2012;2012:1-6.
38. Lee HH, Paeng JC, Cheon GJ, Lee DS, Chung JK, Kang KW. Recurrence of Melanoma After Initial Treatment: Diagnostic Performance of FDG PET in Posttreatment Surveillance. *Nucl Med Mol Imaging*. oct 2018;52(5):327-33.
39. Chartrand G, Cheng PM, Vorontsov E, Drozdal M, Turcotte S, Pal CJ, et al. Deep Learning: A Primer for Radiologists. *RadioGraphics*. nov 2017;37(7):2113-31.
40. Sun R, Deutsch E, Fournier L. Intelligence artificielle et imagerie médicale. *Bulletin du Cancer*. janv 2022;109(1):83-8.
41. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM*. 24 mai 2017;60(6):84-90.
42. Ojha V, Nicosia G. Backpropagation Neural Tree. *Neural Networks*. mai 2022;149:66-83.
43. Sarigül M, Ozyildirim BM, Avci M. Differential convolutional neural network. *Neural Networks*. août 2019;116:279-87.
44. Blanc-Durand P, Van Der Gucht A, Schaefer N, Itti E, Prior JO. Automatic lesion detection and segmentation of 18F-FET PET in gliomas: A full 3D U-Net convolutional neural network study. Sherman JH, éditeur. *PLoS ONE*. 13 avr 2018;13(4):e0195798.
45. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation [Internet]. arXiv; 2016 [cité 10 mars 2024]. Disponible sur: <http://arxiv.org/abs/1606.06650>
46. Hsu LM, Wang S, Walton L, Wang TWW, Lee SH, Shih YYI. 3D U-Net Improves Automatic Brain Extraction for Isotropic Rat Brain Magnetic Resonance Imaging Data. *Front Neurosci*. 16 déc 2021;15:801008.
47. Bonardel G, Dupont A, Decazes P, Queneau M, Modzelewski R, Coulot J, et al. Clinical and phantom validation of a deep learning based denoising algorithm for F-18-FDG PET images from

- lower detection counting in comparison with the standard acquisition. *EJNMMI Phys.* 11 mai 2022;9:36.
48. Weyts K, Lasnon C, Ciappuccini R, Lequesne J, Corroyer-Dulmont A, Quak E, et al. Artificial intelligence-based PET denoising could allow a two-fold reduction in [18F]FDG PET acquisition time in digital PET/CT. *Eur J Nucl Med Mol Imaging.* sept 2022;49(11):3750-60.
 49. Capobianco N, Meignan M, Cottereau AS, Vercellino L, Sibille L, Spottiswoode B, et al. Deep-Learning ¹⁸F-FDG Uptake Classification Enables Total Metabolic Tumor Volume Estimation in Diffuse Large B-Cell Lymphoma. *J Nucl Med.* janv 2021;62(1):30-6.
 50. Van Der Gucht A, Mantzarides M, Lerman L, Itti E, Blanc Durand P. Évaluation en routine clinique de PaIRe, outil d'intelligence artificielle pour la détection et la quantification automatique des lésions en TEP/TDM au 18F-FDG. *Médecine Nucléaire.* 1 juill 2021;45(4):192.
 51. Blanc-Durand P, Jégou S, Kanoun S, Berriolo-Riedinger A, Bodet-Milin C, Kraeber-Bodéré F, et al. Fully automatic segmentation of diffuse large B cell lymphoma lesions on 3D FDG-PET/CT for total metabolic tumour volume prediction using a convolutional neural network. *Eur J Nucl Med Mol Imaging.* mai 2021;48(5):1362-70.
 52. Mechti M, Itti E, Blanc Durand P. Comparaison de deux outils d'intelligence artificielle pour la détection d'anomalies et la segmentation en TEP/TDM au 18F-FDG dans une cohorte monocentrique de cancers de poumons. *Médecine Nucléaire.* mars 2023;47(2):70.
 53. Boellaard R, Delgado-Bolton R, Oyen WJG, 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.* 1 févr 2015;42(2):328-54.
 54. Nioche C, Orhac F, Boughdad S, Reuzé S, Goya-Outi J, Robert C, et al. LIFEx: A Freeware for Radiomic Feature Calculation in Multimodality Imaging to Accelerate Advances in the Characterization of Tumor Heterogeneity. *Cancer Research.* 14 août 2018;78(16):4786-9.
 55. Seban RD, Nemer JS, Marabelle A, Yeh R, Deutsch E, Ammari S, et al. Prognostic and theranostic 18F-FDG PET biomarkers for anti-PD1 immunotherapy in metastatic melanoma: association with outcome and transcriptomics. *Eur J Nucl Med Mol Imaging.* 1 oct 2019;46(11):2298-310.
 56. Seban RD, Moya-Plana A, Antonios L, Yeh R, Marabelle A, Deutsch E, et al. Prognostic 18F-FDG PET biomarkers in metastatic mucosal and cutaneous melanoma treated with immune checkpoint inhibitors targeting PD-1 and CTLA-4. *Eur J Nucl Med Mol Imaging.* sept 2020;47(10):2301-12.

RESUME - ABSTRACT

Détection automatique de lésions par « intelligence artificielle » en TEP/TDM au [¹⁸F]FDG en routine clinique : quelle méthode est la plus performante ?

Introduction – Des algorithmes d'IA ont été commercialisés depuis peu pour les épauler médecins nucléaires dans la détection des lésions en routine. Cette étude a pour objectif d'évaluer les performances de détection de lésions sur la TEP/TDM au [¹⁸F]FDG par deux algorithmes d'IA utilisant un réseau de neurones convolutifs, lorsqu'ils ont utilisés seuls ou en complément de la lecture humaine.

Matériels et méthodes – 151 TEP/TDM au [¹⁸F]FDG de patients pris en charge pour des mélanomes ou des lymphomes dans 3 centres français ont été rétrospectivement analysées. La détection des lésions a été effectuée manuellement (M1), par AUTOID (seuil à SUV 2,5 (M2) ou selon PERCIST (M3)) et par PAIRE (M4). L'ensemble des volumes d'intérêt (VOI) mis en évidence pour chacune de 4 méthodes était revu par un consensus d'experts. Les VOIs jugés correspondant à des lésions néoplasiques en lien avec la pathologie étudiée étaient labélisés « vrais positifs » (VP). Les sensibilités de chacune des 4 méthodes de détection étaient comparées deux à deux en utilisant le test t de Student pour échantillons appariés.

Résultats – Au total, 1544 lésions ont été considérées comme VP selon le standard de référence. Les sensibilités respectives de PAIRE (M4), d'AUTOID avec seuil à un SUV à 2,5 (M2) et selon PERCIST (M3) et de la méthode manuelle (M1), étaient respectivement de 95,7%, 60,2%, 44,3% et 76,6%. La sensibilité de PAIRE et de la détection manuelle associés atteignaient 99,9% et étaient significativement plus élevée que celle de toute autre méthode seule ou en combinaison. PAIRE (M4) rapporterait le plus grand nombre (1435) de VOIs faux positifs, contre 837 pour M2 et 151 avec M3.

Conclusion – L'utilisation de PAIRE pourrait constituer un filet de sécurité robuste pour les médecins nucléaires dans la détection de lésions en TEP/TDM au [¹⁸F]FDG. Son intégration en routine est limitée par son taux élevé de faux positifs et doit être confortée par d'autres études.

Mots-clés : TEP/TDM ; [¹⁸F]FDG ; IA ; AUTOID ; PAIRE

Automatic detection of lesions by "artificial intelligence" in [¹⁸F]FDG PET/CT in clinical routine: which method is the most efficient?

Introduction - AI algorithms have recently been commercialized to support nuclear physicians in routine lesion detection. The aim of this study was to evaluate lesion detection performance on [¹⁸F]FDG PET/CT by two AI algorithms using a convolutional neural network, when used alone or in addition to human reading.

Materials and methods - 151 [¹⁸F]FDG PET/CT scans of patients managed for melanoma or lymphoma in 3 French centers were retrospectively analyzed. Lesion detection was performed manually (M1), by AUTOID (threshold at SUV 2.5 (M2) or PERCIST-like (M3)) and by PAIRE (M4). All volumes of interest (VOIs) identified by each of the 4 methods were reviewed by a consensus of experts. VOIs judged to correspond to neoplastic lesions related to the pathology studied were labelled "true positives" (TP). The sensitivities of each of the 4 detection methods were compared in pairs using Student's t-test for paired samples.

Results - A total of 1544 lesions were considered TP according to the reference standard. The respective sensitivities of PAIRE (M4), AUTOID with threshold at SUV of 2.5 (M2) and PERCIST-like (M3), and the manual method (M1), were 95.7%, 60.2%, 44.3% and 76.6% respectively. The sensitivity of PAIRE combined with the human detection reached 99.9% and was significantly higher than that of any other method alone or in combination. PAIRE (M4) reported the highest number (1435) of false-positive VOIs, compared with 837 for M2 and 151 with M3.

Conclusion - The use of PAIRE could provide a robust safety net to physicians for lesion detection on [¹⁸F]FDG PET/CT. Its routine use is limited by its high false-positive rate and needs to be confirmed by further studies.

Key words: PET/CT; [¹⁸F]FDG; AI; AUTOID; PAIRE