

## Metabolically active volumes automatic delineation methodologies in PET imaging: review and perspectives

Mathieu Hatt, Nicolas Boussion, Catherine Cheze-Le Rest, Olivier Pradier,  
Dimitris Visvikis

► **To cite this version:**

Mathieu Hatt, Nicolas Boussion, Catherine Cheze-Le Rest, Olivier Pradier, Dimitris Visvikis. Metabolically active volumes automatic delineation methodologies in PET imaging: review and perspectives. Cancer Radiothérapie, Elsevier Masson, 2011, in press. <inserm-00610621>

**HAL Id: inserm-00610621**

**<http://www.hal.inserm.fr/inserm-00610621>**

Submitted on 22 Jul 2011

**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.

Titre : Méthodologies de définition automatique des volumes métaboliquement actifs en imagerie TEP : évaluation et perspectives

Title : Metabolically active volumes automatic delineation methodologies in PET imaging: review and perspectives

Auteurs : M. Hatt<sup>A</sup>, N. Boussion<sup>A,B</sup>, C. Cheze-Le Rest<sup>A</sup>, D. Visvikis<sup>A</sup>, O. Pradier<sup>A,B</sup>

<sup>A</sup> INSERM U650 LaTIM, CHU Morvan, 5 avenue Foch, Brest, France

<sup>B</sup> Département de radiothérapie, CHU Morvan, 5 avenue Foch, Brest, France

Auteur correspondant :

Mathieu HATT

LaTIM, INSERM U650

CHU MORVAN

5 avenue Foch

29609 Brest

France

e-mail : hatt@univ-brest.fr

Tel : +33 2 98 01 81 11

Fax : +33 2 98 01 81 24

Titre courant : définition automatique des volumes TEP

**Résumé :**

L'imagerie TEP est dorénavant un outil de référence en routine clinique en oncologie, notamment dans le cadre du diagnostic. Des applications plus récentes, telles la prise en charge et le suivi thérapeutique, ou la définition de cibles en radiothérapie nécessitent une détermination rapide, précise et robuste des volumes métaboliquement actifs sur les images d'émission, ce qui ne peut être obtenu par contours manuels. Ce besoin clinique a motivé de nombreux développements ces dernières années pour la mise au point de méthodes automatiques. Cette revue propose une vue d'ensemble de ces méthodologies, et discute leur mise en œuvre et leur validation méthodologique et/ou cliniques. Des perspectives sur les travaux encore à accomplir sont également suggérées.

Mots clés: TEP ; segmentation ; contours ; radiothérapie ; suivi thérapeutique ; validation

**Abstract:**

PET imaging is now considered a gold standard tool in clinical oncology, especially for diagnosis purposes. More recent applications such as therapy follow up or tumor targeting in radiotherapy require a fast, accurate and robust metabolically active tumor volumes on emission images, which cannot be obtained through manual contouring. This clinical need has sprung a large number of methodological developments regarding automatic methods to defined tumor volumes on PET images. This paper reviews most of the methodologies that have been recently proposed and discusses their framework and methodological and/or clinical validation. Perspectives regarding the future work to be done are also suggested.

**Keywords:** PET ; segmentation ; contouring ; radiotherapy ; therapy follow up ; validation

## Définition automatique des volumes TEP ; M. Hatt

### 1. Introduction

L'imagerie TEP est dorénavant un outil de référence en routine clinique en oncologie, notamment pour le diagnostic (1). Des applications plus récentes de cette imagerie fonctionnelle concernent la prise en charge et le suivi thérapeutique (2), ou l'identification et la définition des cibles en radiothérapie (3). Ces dernières applications sont encore au stade du développement. Elles nécessitent des efforts en termes de standardisation (4), notamment multicentriques, ainsi que la mise au point d'outils permettant une quantification robuste, précise et reproductible. En particulier, la mise à disposition de méthodes permettant la définition automatique, rapide et fiable des volumes métaboliquement actifs (VMA) des tumeurs a été identifié comme un besoin important et urgent (2, 5). Les applications visées sont notamment la mise au point de nouveaux indices pronostiques (6-7) ou prédictifs de la réponse (8-9), le suivi thérapeutique et l'évaluation de la réponse (2), et la définition des contours des GTV (*Gross Tumor Volumes*) en radiothérapie (10). Ce besoin a notamment émergé suite aux efforts déployés depuis le début des années 2000 pour tenter de mettre en œuvre le « volume cible biologique », avec l'utilisation de radiotraceurs différents et plus spécifiques que le  $^{18}\text{F}$ -FDG, dont l'utilisation reste largement majoritaire en oncologie. Les traceurs permettant de visualiser la prolifération cellulaire ( $^{18}\text{F}$ -FLT) ou l'hypoxie ( $^{64}\text{Cu}$ -ATSM,  $^{18}\text{F}$ -FMISO) suscitent un intérêt particulier en radiothérapie car ils permettent d'identifier des régions plus agressives ou radiorésistantes. Ils prennent une importance accrue en combinaison avec les nouvelles techniques d'irradiation ciblée comme la radiothérapie conformationnelle par modulation d'intensité (RCMI) et les scénarii d'escalade de dose (11) et de "dose painting" (12). Toutefois la méthode standard de définition des GTV reste le contour manuel coupe par coupe par le radiothérapeute sur les volumes obtenue par TDM (TomoDensitoMétrie) dosimétrique. Cette dernière bénéficie d'une résolution spatiale élevée de l'ordre du mm, supérieure à celle offerte par la TEP (de

l'ordre de 4 à 6 mm). Plusieurs travaux ont étudié l'impact de l'utilisation des images fusionnées TEP/TDM sur la reproductibilité des contours ainsi que sur la taille et la forme des GTV. Par rapport à l'utilisation de l'image anatomique seule, les contours réalisés sur la fusion TEP/TDM mène à des GTV en général plus reproductibles (13-14), et souvent significativement plus petits ou plus grands en fonction des localisations et des cas (15-16). Actuellement, l'imagerie TEP est utilisée en routine clinique au mieux comme un guide visuel à la définition des GTV. La planification réellement fondée sur l'imagerie fonctionnelle (avec le traceur FDG seul ou combinant différents traceurs), bien qu'ayant été suggérée (17), est loin d'être une réalité clinique. Une des limites majeures à la réalisation d'une telle méthodologie est sa complexité : si la variabilité inter et intra utilisateurs de la détermination manuelle des GTV sur l'imagerie anatomique peut être significative (14), celle constatée sur l'imagerie fonctionnelle est plus importante et incompatible avec une pratique clinique (18).

Nous proposons dans cette revue une évaluation critique des solutions qui ont été proposées jusqu'à présent pour assister le clinicien dans la démarche chronophage et complexe de définition des contours du VMA sur les images TEP. Nous discuterons des manques actuels qu'il faut combler pour parvenir à des solutions pouvant être mises en pratique en routine clinique.

## 2. Evaluation critique

### 2.1 La problématique des méthodes fondées sur des seuillages

Les premiers travaux visant à établir une définition des VMA en TEP sont fondés majoritairement sur des méthodes de seuillage déterministe (fixe ou adaptatif) des valeurs des voxels de l'image. Une liste détaillée de ces travaux peut être consultée dans la revue de Dewalle et al. (19). Le cas de l'imagerie TEP est particulier dans le sens où les seuillages ont pris une grande importance en termes de popularité auprès des cliniciens et en conséquence un grand nombre de publications les utilisant. Le principal attrait de ces approches est leur simplicité méthodologique et leur facilité d'utilisation. Elles reposent

toutefois sur des hypothèses simplificatrices, en particulier concernant la distribution du radiotracer au sein du VMA. Cela explique la grande variabilité de valeurs de seuils indiquées dans les publications (de 20 à 75% du SUV maximum) et leur manque de précision et de robustesse (18, 20-24). Afin d'apporter une réponse au problème fondamental de l'utilisation d'une valeur fixe de seuil, il a été proposé d'adapter cette valeur aux caractéristiques du VMA et de son environnement immédiat. Ces « seuils adaptatifs », se fondent sur une estimation approximative du contraste entre le VMA et le « fond » physiologique, en lien avec une optimisation propre au système et au protocole d'acquisition. Les résultats sont ainsi améliorés, mais ces approches souffrent encore de nombreuses limites. L'estimation initiale du contraste est le plus souvent dépendante de l'utilisateur (18, 22), et l'optimisation nécessaire rend la méthode spécifique au système et au protocole du centre clinique. Une variabilité significative peut en effet être observée entre des centres cliniques différents utilisant pourtant le même modèle de scanner (25). En l'absence de standardisation, ces approches peuvent difficilement être utilisées dans le cadre d'études multicentriques. Ces techniques sont en outre susceptibles d'échouer dans des situations complexes, notamment de distribution hétérogène du radiotracer (6, 23, 26), comme illustré en figure 1. Il est dorénavant acquis que les seuillages fixes sont à proscrire, et qu'en absence d'alternative, une segmentation manuelle est préférable (20). Les seuillages adaptatifs sont susceptibles d'apporter une précision suffisante dans les cas simples, mais sont dépendants des utilisateurs et du système. Ils font globalement trop d'hypothèses simplificatrices pour pouvoir être considérés comme une solution d'avenir.

### 3. Méthodologies de segmentation d'images

Etant données les limites des méthodes à base de seuillage, il est pertinent de s'intéresser à des méthodologies fondées sur des approches de segmentation d'image plus avancées. Ces dernières ont en commun des outils d'analyse et de segmentation d'image ayant été développés auparavant dans d'autres domaines. Elles se distinguent par le type de méthodologie considérée, les éventuelles modifications apportées, leur niveau

d'automatisation, la nécessité de pré- et/ou post-traitement(s), et par leur niveau de validation.

Bien que les spécificités propres à chaque localisation et type de tumeurs ou au radiotracer considéré doivent être gardées à l'esprit, la définition automatique d'un VMA sur une image TEP reste avant tout une problématique de segmentation d'image. C'est dans cette optique que nous orientons cette revue critique. Certaines méthodologies ont été développées pour résoudre la problématique de façon globale, tandis que d'autres l'ont été pour une localisation particulière (ORL, poumon, rectum...) ou une problématique plus spécifique encore (imagerie dynamique, contours en radiothérapie...). Notons également que l'immense majorité des travaux ont étudié le problème en considérant le  $^{18}\text{F}$ -FDG.

### 3.2 Méthodologies considérées

Il existe littéralement des centaines de méthodes de segmentation d'images, et plusieurs des plus connues ont été considérées par différents groupes pour étudier la problématique de la segmentation d'images TEP afin de définir les VMA. Certaines cherchent à identifier les contours dans l'image, d'autres tentent d'identifier les régions ou regroupements de voxels. Ceci peut être fait suivant des critères variés comme les valeurs, les formes ou les textures, par des approches déterministes, statistiques et probabilistes, ou d'intelligence artificielle.

Une approche classique en segmentation d'images consiste en la détection de contours par analyse du gradient de l'image, afin de définir une région interne et une région externe au contour détecté. L'application directe de ce type de méthode en TEP est difficile car les contours sont flous et les gradients difficiles à identifier. Plusieurs études ont étudié l'adaptation de ces méthodes pour la définition de VMA en TEP. Par exemple la méthode du « partage des eaux » (*ang. watershed*) qui est fondée sur l'analogie avec une surface topologique se remplissant d'eau (27). Une autre approche est fondée sur la détection des pics de gradient pour identifier les contours des objets d'intérêt (28-32). Il a également été proposé d'améliorer le résultat d'un seuillage adaptatif par l'utilisation d'un contour actif (33).



Notons qu'il a été suggéré également de faire évoluer de tels contours en prenant en compte les images TEP et TDM simultanément (34).

Les méthodologies fondées sur le regroupement (*ang. clustering*) flou non supervisé de voxels (*ang. Fuzzy C-Means, FCM*) ont été utilisées par plusieurs groupes (21, 35-36). Notons que l'algorithme original est relativement simpliste et ne prend pas en compte les corrélations spatiales entre voxels, par exemple, et mène à des résultats décevants (21, 36). Il semble toutefois *a priori* assez bien adapté aux images TEP, du fait de sa modélisation floue. Le groupe ayant proposé la méthode la plus aboutie fondée sur le FCM, a utilisé une version modifiée incorporant des informations supplémentaires comme la détection automatique du nombre de classes (ou *clusters*), la corrélation spatiale des voxels et l'analyse de l'hétérogénéité du traceur (36).

Le regroupement des voxels pour définir les régions tumorales et physiologiques peut aussi se faire par la différentiation statistique des valeurs dans l'image. Ce type de méthode est très utilisé en traitement d'images et a de nombreuses applications en imagerie satellite et astronomique notamment (37). Le principe est de distinguer les voxels appartenant aux tumeurs de ceux appartenant aux tissus sains par leurs propriétés statistiques respectives. Elles sont toutefois assez peu adaptées au traitement d'images TEP à cause de la difficulté de prise en compte du flou. Une approche simple consiste à faire croître la région de l'image correspondant au VMA à partir du voxel d'intensité maximale, avec comme critère de croissance la moyenne et la variance (38). Une autre approche est appliquée aux projections MIP (*ang. Maximum Intensity Projection*) pour bénéficier du contraste ainsi augmenté, et fait correspondre le VMA défini sur le volume TEP d'origine grâce aux ensembles flous associés à un opérateur de fusion (35). L'utilisation de mélanges de gaussiennes pour opérer la classification des voxels a été proposée également. Cette méthode ne considère pas de modèle spatial et nécessite une complexe estimation du nombre total de gaussiennes à utiliser et de celles qui sont à associer au VMA (39). Dans ce contexte, il est possible de modéliser l'information spatiale dans l'image par des modèles de Markov (40). Il est aussi

possible de modifier la modélisation des données au sein de ces outils, pour prendre en compte le flou et ainsi obtenir des résultats améliorés (41). L'approche FLAB (*ang. Fuzzy Locally Adaptive Bayesian*) utilise ce principe (18, 21-22, 26).

L'analyse de texture a également fait l'objet d'adaptation pour classifier les voxels d'images TEP/TDM (42-43). Cette approche consiste à apprendre à l'algorithme, via une base de données d'images pour lesquelles la vérité est identifiée par des médecins, en quoi les voxels de tumeurs forment des textures différentes de celles des voxels physiologiques. Cet apprentissage permet la construction de classifieurs (par exemple des arbres de décision) servant à classifier les voxels d'une nouvelle image proposée en entrée de l'algorithme. Une méthode similaire dans son approche (apprentissage sur base de données pour générer un classifieur), mais utilisant des réseaux de neurones, a récemment été proposée (44).

On peut citer d'autres approches, notamment une étude qui s'est spécifiquement intéressée à la segmentation de tumeurs en détectant les fixations physiologiques situées à proximités pour éviter de les incorporer au VMA (45). Il a aussi été proposé de classifier les voxels comme tumoraux ou physiologiques en les regroupant en fonction de la courbe temps-activité (*ang. Time-activity curve*) les caractérisant (46). Cette méthode ne peut toutefois être utilisée que sur des images paramétriques issues d'acquisitions TEP dynamiques.

### 3.3 Prise en compte de la nature des images TEP

Des méthodes reconnues comme peu robustes au bruit et fondées sur des décisions de segmentation peu flexibles paraissent *a priori* mal adaptées à l'analyse d'images TEP. C'est pourtant le cas des seuillages. Certaines approches, au contraire, proposent de modéliser explicitement les caractéristiques des images TEP. C'est notamment le cas de l'approche FLAB (21). D'autres méthodes ont modifié des algorithmes connus pour obtenir une précision satisfaisante. C'est le cas de l'algorithme par partage des eaux, pour lequel le

contour d'origine a été modifié (27). De façon similaire, l'algorithme FCM a été modifié pour incorporer l'information de corrélation spatiale et celle de l'hétérogénéité du traceur (36). Les approches fondées sur des techniques issues de l'intelligence artificielle (42-44) ne proposent pas de modélisation explicite des caractéristiques des images TEP, toutefois leur nature intrinsèque doit leur permettre d'apprendre ces caractéristiques pour s'y adapter. Une alternative consiste à améliorer la qualité des images TEP préalablement à la segmentation. Ainsi, la méthodologie fondée sur la théorie des possibilités est appliquée aux projections MIP et non aux images d'origine, dans le but d'opérer la segmentation sur un contraste plus élevé (35). Les approches utilisant l'analyse du gradient sont souvent associées à une ou plusieurs étapes de pré traitement visant à permettre une détection moins complexe des contours (32, 34). Ces pré traitements atténuent le bruit et réduisent le flou. La même méthodologie proposée par l'industrie MIMVista (28-31) semble ne pas faire appel à de tels pré traitements<sup>1</sup>. Ceci pose la question de la dépendance à cette étape, car ces corrections sont délicates à étendre à l'ensemble des images TEP (47-48). D'autres approches ont par ailleurs démontré de bonnes performances sur les images d'origine sans nécessiter de tels pré traitement(s) (21, 36).

Le développement d'une approche de segmentation automatique d'images TEP nécessite en effet la prise en compte de plusieurs paramètres.

D'une part, ceux qui sont liés à la tumeur ou à l'organe à définir :

- hétérogénéité de la distribution du traceur au sein de la tumeur/organe
- hétérogénéité de la distribution du traceur au sein du fond physiologique
- contraste(s) mis en jeu entre l'objet et le fond, et au sein même de l'objet et du fond
- complexité de la forme de l'objet

D'autre part, ceux qui sont liés à la nature de l'acquisition

---

<sup>1</sup> Il est important de souligner que les détails de la méthodologie ne sont pas disponibles, seuls les résultats étant présentés dans des papiers de conférences et non des articles en journaux à comité de lecture

- niveau de bruit
- taille des voxels et échantillonnage spatial
- type et modèle du scanner
- algorithme de reconstruction et ses paramètres

Les paramètres ayant le plus d'impact sur la précision et la robustesse d'une segmentation sont l'hétérogénéité de la distribution du traceur et les contrastes mis en jeu (voir figure 1). Les VMA réels sont rarement sphériques et homogènes sur un fond homogène comme le supposent de nombreuses méthodes, et comme sont conçues les données de validation utilisées dans de nombreuses publications.

Les autres paramètres peuvent également avoir un impact important sur la qualité globale de l'image. La taille des voxels, par exemple, a un impact significatif sur la définition d'un contour. Plus les voxels utilisés pour définir la grille de reconstruction sont petits, meilleur est l'échantillonnage spatial des structures, et plus le contour peut être finement défini. Toutefois, la réduction de la taille des voxels dans la grille de reconstruction entraîne une baisse significative des statistiques disponibles en chaque voxel, et une augmentation significative du bruit (voir figure 2). Une alternative pour améliorer l'échantillonnage spatial consiste à sur échantillonner l'image, ce qui est d'ailleurs réalisé dans les stations commerciales pour la visualisation. Ceci a toutefois un impact important sur l'aspect visuel des images et peut varier en fonction de la méthode utilisée (voir figure 3). Etant donnée la grande variabilité des modèles et types de scanner en activité, le manque de standardisation des protocoles d'acquisition, et le grand nombre d'algorithmes de reconstruction existant (et la possibilité de varier leurs paramètres), les images TEP peuvent présenter des aspects très variés (voir figure 4). Se pose alors le problème de l'universalité et de la robustesse des méthodes proposées qui n'ont été validées que sur un nombre restreint de données. Enfin, les mouvements physiologiques, en particulier la respiration, ont un impact important, notamment pour les localisations thoraciques. Plusieurs études ont récemment étudié cette problématique (49-50), mais la grande majorité des méthodes ne prennent pas en compte ce

paramètre dans leur validation. Soulignons toutefois qu'il n'est pas déraisonnable de faire l'hypothèse que les effets de la respiration puissent être corrigés en amont au cours de la reconstruction ou en post reconstruction (51).

### 3.4 Validation

Afin de démontrer rigoureusement les performances d'une méthodologie, il faut évaluer :

- La précision absolue
- La robustesse
- La reproductibilité et la dépendance éventuelle à l'utilisateur

La précision désigne la capacité de l'algorithme à définir la position, la forme et le volume du VMA et pour l'évaluer, il est nécessaire de disposer d'une vérité terrain. Certains travaux ont suggéré l'utilisation des volumes segmentés sur les images TDM comme vérité terrain, ce qui est aberrant car rien ne garanti que les volumes anatomiques et fonctionnels soient parfaitement superposés. La définition manuelle par un expert ne permet pas non plus de générer une vérité terrain satisfaisante du fait de la variabilité inter comme intra utilisateurs (14, 18). Une alternative est de définir un consensus de nombreux utilisateurs. Cela reste toutefois dépendant des utilisateurs, ainsi que de la façon dont le consensus est défini (52). Dans le cas des images cliniques, c'est toutefois la seule alternative à l'analyse histopathologique qui consiste à extraire chirurgicalement l'objet et d'en réaliser des mesures macroscopiques. Cette opération comporte de nombreuses sources d'erreurs et d'approximations, l'objet pouvant subir des déformations, devant être coupé en tranches (ce qui peut provoquer des pertes de matériau), et les régions correspondant aux zones tumorales étant définies manuellement sur les coupes. Enfin, les données doivent être recalées spatialement, ce qui est complexe étant donnée la nature très différente des images. Peu de jeux de données existent (14, 53-57). Ils sont de taille réduite, de moins d'une dizaine à une trentaine de patients. Dans la majorité des cas, seul le diamètre maximal est mesuré, ce qui constitue une mesure réductrice du VMA.

Une alternative à l'utilisation de données cliniques pour évaluer la précision consiste à réaliser des acquisitions de modèles physiques réels dont les dimensions sont connues, tels les fantômes NEMA IEC contenant des sphères de taille variable. La seule mesure pertinente qu'il est possible d'évaluer sur de telles données est la robustesse par rapport à des structures de tailles et de contrastes variables. En effet, les objets en question sont des sphères dont l'activité est homogène, placées sur un fond homogène, ce qui est simpliste par rapport aux VMA réels. Des fantômes physiques permettant de générer des activités hétérogènes et/ou des formes plus complexes ont été proposées (40) mais restent loin de la complexité des images cliniques réelles et sont complexes à produire et utiliser. Une étude rigoureuse et complète de la robustesse devrait idéalement être réalisée sur des objets plus complexes en termes de formes et d'hétérogénéité, que des acquisitions de fantômes contenant des sphères. Ceci peut être réalisé grâce aux simulations numériques.

Une troisième solution permettant d'évaluer les performances d'une approche est l'utilisation de données simulées réalistes. Générer de telles données nécessite l'utilisation de simulateurs de particules comme SORTEO<sup>2</sup> (58) ou GATE<sup>3</sup> (59) combinant des modèles de fantômes anthropomorphiques numériques comme Zubal (60) ou XCAT<sup>4</sup> (61) avec des modèles de systèmes d'acquisition (62-63). Il est possible avec ces outils de générer des structures complexes et réalistes, y compris en ce qui concerne la modélisation des tumeurs (64), avec l'avantage de connaître la vérité terrain et de contrôler la plupart des paramètres. Toutefois, cette approche nécessite des compétences spécifiques et des matériels dédiés, du fait de l'exigence en termes de puissance de calcul. Des simulations analytiques plus simples peuvent également être considérées, mais les résultats doivent alors être considérés avec plus de précaution car elles impliquent des simplifications.

La mesure de mérite utilisée afin d'évaluer la précision par rapport à la vérité terrain a également son importance. Les erreurs de volume peuvent être suffisantes pour les sphères

---

<sup>2</sup> Simulation Of Realistic Tridimensional Emitting Objects

<sup>3</sup> Geant4 Application for Tomography Emission

<sup>4</sup> 4D eXtended CArdiac Torso phantom

homogènes, mais ne sont pas rigoureuses pour évaluer la précision sur des objets plus complexes, car une segmentation erronée peut fournir le même volume absolu. Des mesures de mérite comme les erreurs de classification (21) ou les coefficients de Dice (65) sont plus pertinents.

La robustesse mesure la capacité d'une méthode à fonctionner sur la grande variabilité des images TEP, sans ajustement préalable de paramètres. Il est donc nécessaire de considérer différents modèles de scanner et les algorithmes de reconstruction associés, et pour chacun, différents paramètres d'acquisition pour évaluer la robustesse au bruit, au contraste, ou encore à l'échantillonnage spatial. Il est envisageable de générer de telles données avec les outils de simulations réalistes, mais cela implique la modélisation de différents scanners et peut donc être laborieux. Une alternative consiste à exploiter des acquisitions de fantômes physiques réalisées dans différents centres, pour obtenir une large gamme de situations. Ce type de données est toutefois long à acquérir, et souffre des limites évoquées précédemment pour l'évaluation de la précision.

Evaluer la reproductibilité (aussi dénommée « répétabilité ») peut se faire par la répétition de l'exécution de la segmentation sur les mêmes données. On peut ainsi la comparer avec la variabilité observée avec la définition manuelle (18). Ainsi, l'exécution répétée d'un seuillage fixe sur une même image donne systématiquement le même résultat avec une variabilité nulle. Au contraire, l'utilisation de méthodes avancées nécessitant par exemple un processus itératif peut mener à différents résultats en fonction par exemple des paramètres d'initialisation et du critère de convergence.

L'évaluation des interactions avec l'utilisateur est plus délicate à quantifier. Il est nécessaire de déterminer si l'utilisateur doit détecter et isoler le VMA, définir des régions d'intérêt (dans le VMA et/ou le fond), ou encore ajuster des paramètres au cas par cas. La plupart des méthodes proposées font l'hypothèse préalable que le VMA a été détecté et isolé par l'utilisateur. Quelques méthodes (voir tableau 1) se sont toutefois positionnées dans le

cadre de la segmentation de l'image entière, l'interaction avec l'utilisateur intervenant une fois la segmentation effectuée, pour analyser le résultat et associer au VMA certaines des régions obtenues par la segmentation.

Par rapport à ces exigences, les méthodologies évoquées dans la section précédente fournissent des preuves de performances à des niveaux différents. Le tableau 1 résume pour chaque approche le type de segmentation utilisée, l'interaction utilisateur, les pré- ou post-traitement nécessaires, les données utilisées pour la validation de la précision (et la vérité terrain associée), et si une évaluation de la robustesse et de la reproductibilité a été réalisée.

L'analyse de ce tableau révèle de grands écarts de qualité de validation entre les publications, ainsi qu'une grande variabilité dans les données de validation considérées. On notera le faible nombre de publications ayant explicitement démontré la capacité de la méthode à segmenter avec précision des VMA réalistes dont la fixation de radiotraceur est fortement hétérogène, c'est-à-dire dont les contrastes intra-tumoraux sont suffisamment élevés pour rendre inadéquate une segmentation comme celle des seuillages (26, 29, 36). Notons le cas particulier de la méthode par gradient, qui semble capable de générer un contour sur une tumeur hétérogène contenant un cœur nécrosé, mais n'est pas capable de définir le contour interne de la nécrose en question, ou de définir des régions différentes au sein de la tumeur (voir figure 5). Pour une majorité des publications, la précision a été évaluée sur des jeux de données relativement simplistes et de taille limitée et/ou des données cliniques sans vérité terrain rigoureuse. Plusieurs approches ont été appliquées sur des jeux de données disposant de mesures histopathologiques en commun. La méthode par gradient, le FCM, et celle de la théorie des possibilités ont été testées sur 7 patients ORL (53), avec des erreurs par rapport aux volumes de  $19\pm 22\%$ ,  $9\pm 28\%$  et  $18\pm 10\%$  respectivement. Les méthodes FCM et FLAB ont été appliquées sur 18 patients avec tumeurs pulmonaires (14), avec  $\pm 6\%$  d'erreur par rapport au diamètre maximum pour FLAB (26) et  $\pm 15\%$  pour FCM (36) (voir figure 6). Dans la majorité des publications, les méthodes proposées sont comparées avec des seuillages, et non avec une méthode plus performante



déjà publiée. Ceci explique en partie la multiplication des méthodes proposées. Il est en effet plus facile de démontrer des résultats améliorés par rapport aux seuillages, dont la précision et la robustesse sont très limitées.

La plupart des études n'ont pas apporté d'informations pertinentes sur les aspects de robustesse et de répétabilité, à l'exception des méthodes par gradient de MIMvista et FLAB. Quelques résultats concernant la répétabilité sont également disponibles dans le cas du contour actif (33), des réseaux de neurones (44) et du mélange de distributions gaussiennes (39).

En conclusion, les méthodologies ayant bénéficié de la validation la plus complète (fantômes sur plusieurs scanners pour la robustesse, images simulées réalistes, images cliniques avec histopathologie, répétabilité) sont les méthodes par gradient et FLAB.

### 3.5 Démonstration de l'impact clinique

Un enjeu important d'actualité est l'évaluation de l'impact clinique d'une définition précise, fiable et reproductible des VMA pour les différentes applications de la TEP. Ceci est crucial afin de convaincre les cliniciens et les industriels de mettre en œuvre ces approches en routine. A ce jour, seuls quelques travaux sont disponibles démontrant, par exemple, un impact en dosimétrie pour la planification de radiothérapie utilisant la méthode par gradient (66). Des travaux récents démontrent également l'intérêt d'une précision accrue dans la définition des VMA, permettant d'extraire des images des paramètres tels que le VMA et le TLG (Total Lesion Glycolysis) (67) associé. Il a été démontré que ces derniers, contrairement aux mesures classiques de SUV, peuvent avoir une valeur prédictive de la survie (6) et de la réponse thérapeutique dans le cadre des lymphomes (68), des mésothéliomes (69) et des cancers localement avancés de l'œsophage (9), et ce sur l'image pré traitement uniquement. Ces paramètres nécessitent toutefois, contrairement à la mesure de  $SUV_{max}$ , une définition précise des VMA. Notons également que ceci permet d'envisager la caractérisation de l'hétérogénéité du traceur au sein du VMA (8, 70-71).

La détermination d'intervalles de confiance permettant de caractériser la reproductibilité des mesures de VMA, afin de les utiliser pour caractériser la réponse thérapeutique (2) peut se faire sur des acquisitions répétées à quelques jours d'intervalle sans traitement. Utiliser une méthode robuste permet d'atteindre le même degré de reproductibilité que le  $SUV_{max}$  ( $\pm 30\%$ ) (18), contrairement à l'utilisation de seuillages qui mènent à des niveaux de variabilité nettement plus élevés ( $\pm 35$  à  $\pm 94\%$ ) (72).

La grande majorité des méthodologies de segmentation d'images TEP qui ont été publiées ces dernières années n'ont pour l'instant pas encore été utilisées afin de démontrer l'intérêt d'une définition fiable des VMA dans les différentes applications cliniques, ce qui conduit à retarder leur adoption par l'industrie, et donc *a fortiori* par les cliniciens.

#### 4. Perspectives

Comme nous l'avons exposé, la problématique de la définition automatique (ou du moins semi-automatique) des VMA sur les images TEP a été en partie résolue par les travaux de certains auteurs, y compris pour des situations relativement complexes de formes et d'hétérogénéité, de faibles contrastes ou rapports signal sur bruit. Les difficultés résident essentiellement dans la validation, souvent délicate et controversée, et dans l'étape de transfert à l'utilisation clinique, ce que peu de groupes ont jusqu'à présent réalisé en utilisant leurs méthodes respectives. Ajoutons à cela la popularité des seuillages fixes et adaptatifs dans les publications, qui parasitent fortement la diffusion et l'acceptation au sein de la communauté clinique de méthodologies plus performantes.

Certaines difficultés pratiques restent pour l'instant non résolues. On peut citer en particulier, outre la problématique de la spécificité du radiotraceur, la différenciation automatique des fixations pathologiques et physiologiques. En ce qui concerne l'identification du VMA à segmenter, il est probable que l'intervention de l'utilisateur restera nécessaire, en particulier pour des cas complexes de tumeurs situées à proximité de zones

ou d'organes associés à une fixation physiologique élevée. La robustesse des méthodes face au manque actuel de standardisation des protocoles d'acquisitions est loin d'être démontrée, bien que certaines investigations aient déjà été menées (22). Les faibles niveaux de contraste et les hauts niveaux de bruit (ou de fixations physiologiques) associés à l'utilisation de traceurs différents du FDG sont encore, même pour les approches les plus performantes, des limites complexes à dépasser. De plus, toutes les méthodes partagent des limitations techniques en termes d'initialisation de paramètres et de dépendance à l'utilisateur et peu d'entre elles ont fait la démonstration d'une automatisation suffisante pour permettre une utilisation aisée et rapide par les cliniciens. Ceci est toutefois un obstacle surmontable car de nombreuses solutions d'interface utilisateur et d'estimation automatique existent, permettant de limiter les interventions de l'utilisateur ou de les rendre plus reproductibles. Il s'agit toutefois d'un effort à fournir essentiellement de la part des industriels pour la mise en œuvre des méthodologies développées par les équipes de recherche au sein de leurs produits destinés aux cliniciens.

Pour ce faire, les industriels doivent pouvoir identifier les méthodologies les plus prometteuses, ce qui est délicat, car la comparaison des méthodes est sujette à controverse si elle est réalisée sur la base des publications disponibles, faute de données de test communes. La mise à disposition de larges bases de données contenant des données cliniques associées à une vérité terrain rigoureuse comme les données d'histopathologie, et des données simulées réalistes couvrant une vaste gamme de situations, permettrait de mettre en place de telles études comparatives. Cet effort n'est pour le moment consenti que par quelques équipes regroupées au sein de collaborations limitées, bien que certaines initiatives soient déjà menées pour tenter de construire et mettre à disposition des bases de données plus conséquentes (73).

Les développements actuels de l'imagerie multi modale au sens large génèrent aussi de nouveaux défis que les méthodes développées jusqu'à présent ne permettent pas de prendre en compte explicitement. La tendance actuelle est en effet à l'augmentation du

nombre de modalités d'imagerie disponibles (IRM, TEP, TDM, etc.) et des modes d'acquisitions (radiotraceurs en TEP, modes en IRM, etc.). Ceci entraîne potentiellement la multiplication d'exams pour un patient donné, et les cliniciens sont confrontés à la prise en compte de multiples images, éventuellement associées, mais souvent acquises dans des configurations spatiales et temporelles différentes. La prise en compte automatique de ces données hétérogènes et multi sources (plusieurs modalités et/ou plusieurs modes d'acquisition ou traceurs pour une modalité donnée, ainsi que l'évolution temporelle à différents instants d'un traitement par exemple) pour un même patient, devra donc faire l'objet de développements appropriés, pour lesquels des outils de recalage et d'analyse d'image et de données existent mais doivent être adaptés et validés.

## 5. Conclusion

Il n'existe pour l'instant pas de consensus dans la communauté sur la méthodologie à adopter pour définir automatiquement les VMA sur les images TEP, que ce soit pour des applications de suivi thérapeutique ou la définition de nouveaux critères pronostics et prédictifs en oncologie, ou bien la définition des GTVs en radiothérapie. Malgré l'existence de méthodologies ayant démontré des performances largement supérieures aux seuillages, qui restent le standard, le manque d'études approfondies et comparatives sur des données établies comme références en est la principale raison. Ce manque s'explique principalement par une ignorance des méthodologies existantes de la part d'une majorité de cliniciens, et d'une persistance de la communauté à n'utiliser et populariser les seuillages. Il s'explique également par un manque de bases de données disponibles et ouvertes, sur lesquelles chaque groupe pourrait tester les performances d'une méthode développée et la comparer à celles proposées précédemment. Tant que de telles données et études ne sont pas disponibles, il sera difficile d'obtenir des industriels une implémentation de méthodologies performantes au sein des outils destinés au cliniciens pour leur pratique routinière.

La plupart des méthodologies existantes souffrent de défauts plus au moins fondamentaux et importants, et des travaux sont encore nécessaires, notamment en termes d'automatisation et de fiabilité. D'autre part, de nouveaux défis voient le jour avec le développement de l'imagerie TEP multi traceurs et les imageries multi modalité (TEP/TDM, TEMP/TDM, TEP/IRM, etc.), notamment pour le traitement d'informations multi dimensionnelles et multi résolution, nécessitant le développement d'approches d'analyse d'images appropriées et innovantes.

Conflits d'intérêt : aucun.

## References

1. Jerusalem G, Hustinx R, Beguin Y, Fillet G. The value of positron emission tomography (PET) imaging in disease staging and therapy assessment. *Ann Oncol*. 2002;13 Suppl 4:227-34.
2. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors. *J Nucl Med*. 2009 May;50 Suppl 1:122S-50S.
3. Jarritt PH, Carson KJ, Hounsell AR, Visvikis D. The role of PET/CT scanning in radiotherapy planning. *Br J Radiol*. 2006 Sep;79 Spec No 1:S27-35.
4. Boellaard R. Standards for PET image acquisition and quantitative data analysis. *J Nucl Med*. 2009 May;50 Suppl 1:11S-20S.
5. Lucignani G. SUV and segmentation: pressing challenges in tumour assessment and treatment. *Eur J Nucl Med Mol Imaging*. 2009 Apr;36(4):715-20.
6. Hatt M, Visvikis D, Albarghach N, Tixier F, Pradier O, Cheze-le Rest C. Prognostic value of 18F-FDG PET image-based parameters in esophageal cancer: impact of tumor delineation methodology. *European Journal of Nuclear Medicine and Molecular Imaging*. 2011.
7. Lucignani G, Larson SM. Doctor, what does my future hold? The prognostic value of FDG-PET in solid tumours. *Eur J Nucl Med Mol Imaging*. 2010 May;37(5):1032-8.
8. Tixier F, Le Rest CC, Hatt M, Albarghach N, Pradier O, Metges JP, et al. Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer. *J Nucl Med*. 2011 Mar;52(3):369-78.
9. Hatt M, Visvikis D, Pradier O, Cheze-le Rest C. Baseline (18)F-FDG PET image-derived parameters for therapy response prediction in oesophageal cancer. *Eur J Nucl Med Mol Imaging*. 2011 May 11.
10. Pan T, Mawlawi O. PET/CT in radiation oncology. *Med Phys*. 2008 Nov;35(11):4955-66.
11. Sovik A, Malinen E, Olsen DR. Strategies for biologic image-guided dose escalation: a review. *Int J Radiat Oncol Biol Phys*. 2009 Mar 1;73(3):650-8.
12. Supiot S, Lisbona A, Paris F, Azria D, Fenoglietto P. ["Dose-painting": myth or reality?]. *Cancer Radiother*. 2010 Oct;14(6-7):554-62.
13. Fox JL, Rengan R, O'Meara W, Yorke E, Erdi Y, Nehmeh S, et al. Does registration of PET and planning CT images decrease interobserver and intraobserver variation in delineating tumor volumes for non-small-cell lung cancer? *Int J Radiat Oncol Biol Phys*. 2005 May 1;62(1):70-5.
14. van Baardwijk A, Bosmans G, Boersma L, Buijsen J, Wanders S, Hochstenbag M, et al. PET-CT-based auto-contouring in non-small-cell lung cancer correlates with pathology and reduces interobserver variability in the delineation of the primary tumor and involved nodal volumes. *Int J Radiat Oncol Biol Phys*. 2007 Jul 1;68(3):771-8.
15. Ashamalla H, Rafla S, Parikh K, Mokhtar B, Goswami G, Kambam S, et al. The contribution of integrated PET/CT to the evolving definition of treatment volumes in radiation treatment planning in lung cancer. *Int J Radiat Oncol Biol Phys*. 2005 Nov 15;63(4):1016-23.
16. Schreurs LM, Busz DM, Paardekooper GM, Beukema JC, Jager PL, Van der Jagt EJ, et al. Impact of 18-fluorodeoxyglucose positron emission tomography on computed tomography defined target volumes in radiation treatment planning of esophageal cancer: reduction in geographic misses with equal inter-observer variability: PET/CT improves esophageal target definition. *Dis Esophagus*. 2010 Aug;23(6):493-501.
17. Gregoire V, Haustermans K, Geets X, Roels S, Lonneux M. PET-based treatment planning in radiotherapy: a new standard? *J Nucl Med*. 2007 Jan;48 Suppl 1:68S-77S.
18. Hatt M, Cheze-Le Rest C, Aboagye EO, Kenny LM, Rosso L, Turkheimer FE, et al. Reproducibility of 18F-FDG and 3'-deoxy-3'-18F-fluorothymidine PET tumor volume measurements. *J Nucl Med*. 2010 Sep;51(9):1368-76.
19. Dewalle-Vignon A, Abiad AE, Betrouni N, Hossein-Foucher C, Huglo D, Vermandel M. Les méthodes de seuillage en TEP : un état de l'art *Médecine Nucléaire*. 2010;34(2):119-31.

20. Hatt M, Visvikis D. Defining radiotherapy target volumes using 18F-fluoro-deoxy-glucose positron emission tomography/computed tomography: still a Pandora's box?: in regard to Devic et al. (*Int J Radiat Oncol Biol Phys* 2010). *Int J Radiat Oncol Biol Phys*. 2010 Dec 1;78(5):1605.
21. Hatt M, Cheze le Rest C, Turzo A, Roux C, Visvikis D. A fuzzy locally adaptive Bayesian segmentation approach for volume determination in PET. *IEEE Trans Med Imaging*. 2009 Jun;28(6):881-93.
22. Hatt M, Cheze Le Rest C, Albarghach N, Pradier O, Visvikis D. PET functional volume delineation: a robustness and repeatability study. *Eur J Nucl Med Mol Imaging*. 2011 Jan 12.
23. Nestle U, Kremp S, Schaefer-Schuler A, Sebastian-Welsch C, Hellwig D, Rube C, et al. Comparison of different methods for delineation of 18F-FDG PET-positive tissue for target volume definition in radiotherapy of patients with non-Small cell lung cancer. *J Nucl Med*. 2005 Aug;46(8):1342-8.
24. Biehl KJ, Kong FM, Dehdashti F, Jin JY, Mutic S, El Naqa I, et al. 18F-FDG PET definition of gross tumor volume for radiotherapy of non-small cell lung cancer: is a single standardized uptake value threshold approach appropriate? *J Nucl Med*. 2006 Nov;47(11):1808-12.
25. Ollers M, Bosmans G, van Baardwijk A, Dekker A, Lambin P, Teule J, et al. The integration of PET-CT scans from different hospitals into radiotherapy treatment planning. *Radiother Oncol*. 2008 Apr;87(1):142-6.
26. Hatt M, Cheze le Rest C, Descourt P, Dekker A, De Ruyscher D, Oellers M, et al. Accurate automatic delineation of heterogeneous functional volumes in positron emission tomography for oncology applications. *Int J Radiat Oncol Biol Phys*. 2010 May 1;77(1):301-8.
27. Tylski P, Bonniaud G, Decenciere E, Stawiaski J, Coulot J, Lefkopoulos D, et al., editors. 18F-FDG PET images segmentation using morphological watershed: a phantom study. *IEEE Nuclear Science Symposium Conference*; 2006.
28. Fogh S, Karancke J, Nelson A, McCue P, Axelrod R, Werner-Wasik W, editors. Pathologic Correlation of PET-CT Based Auto-contouring for Radiation Planning in Lung Cancer. *World Conference on Lung Cancer Meeting*; 2009.
29. Nelson A, Brockway K, Nelson A, Piper J, editors. PET Tumor Segmentation: Validation of a Gradient-based Method Using a NSCLC PET Phantom. *Society of Nuclear Medicine annual meeting*; 2009.
30. Nelson D, editor. PET Tumor Segmentation: Multi-observer Validation of a Gradient-based Method Using a NSCLC PET Phantom. *ASTRO*; 2009: I.
31. Shen G, Nelson D, Adler L, editors. PET Tumor Segmentation: Comparison of Gradient-Based Algorithm to Constant Threshold Algorithm. *AAPM*; 2007.
32. Geets X, Lee JA, Bol A, Lonneux M, Gregoire V. A gradient-based method for segmenting FDG-PET images: methodology and validation. *Eur J Nucl Med Mol Imaging*. 2007 Sep;34(9):1427-38.
33. Li H, Thorstad WL, Biehl KJ, Laforest R, Su Y, Shoghi KI, et al. A novel PET tumor delineation method based on adaptive region-growing and dual-front active contours. *Med Phys*. 2008 Aug;35(8):3711-21.
34. El Naqa I, Yang D, Apte A, Khullar D, Mutic S, Zheng J, et al. Concurrent multimodality image segmentation by active contours for radiotherapy treatment planning. *Med Phys*. 2007 Dec;34(12):4738-49.
35. Dewalle-Vignion AS, Betrouni N, Lopes R, Huglo D, Stute S, Vermandel M. A New Method for Volume Segmentation of PET Images, Based on Possibility Theory. *IEEE Trans Med Imaging*. 2011 Feb;30(2):409-23.
36. Belhassen S, Zaidi H. A novel fuzzy C-means algorithm for unsupervised heterogeneous tumor quantification in PET. *Med Phys*. 2010 Mar;37(3):1309-24.
37. Pieczynski W. Modèles de Markov en traitement d'images. *Traitement du Signal*. 2003;20(30):255-78.
38. Day E, Betler J, Parda D, Reitz B, Kirichenko A, Mohammadi S, et al. A region growing method for tumor volume segmentation on PET images for rectal and anal cancer patients. *Med Phys*. 2009 Oct;36(10):4349-58.



39. Aristophanous M, Penney BC, Martel MK, Pelizzari CA. A Gaussian mixture model for definition of lung tumor volumes in positron emission tomography. *Med Phys*. 2007 Nov;34(11):4223-35.
40. Montgomery DW, Amira A, Zaidi H. Fully automated segmentation of oncological PET volumes using a combined multiscale and statistical model. *Med Phys*. 2007 Feb;34(2):722-36.
41. Hatt M, Lamare F, Boussion N, Turzo A, Collet C, Salzenstein F, et al. Fuzzy hidden Markov chains segmentation for volume determination and quantitation in PET. *Phys Med Biol*. 2007 Jun 21;52(12):3467-91.
42. Yu H, Caldwell C, Mah K, Mozeg D. Coregistered FDG PET/CT-based textural characterization of head and neck cancer for radiation treatment planning. *IEEE Trans Med Imaging*. 2009 Mar;28(3):374-83.
43. Yu H, Caldwell C, Mah K, Poon I, Balogh J, MacKenzie R, et al. Automated radiation targeting in head-and-neck cancer using region-based texture analysis of PET and CT images. *Int J Radiat Oncol Biol Phys*. 2009 Oct 1;75(2):618-25.
44. Sharif MS, Abbod M, Amira A, Zaidi H. Artificial Neural Network-Based System for PET Volume Segmentation. *Int J Biomed Imaging*. 2010;2010.
45. Sebastian TB, Manjeshwar RM, Akhurst TJ, Miller JV. Objective PET lesion segmentation using a spherical mean shift algorithm. *Med Image Comput Comput Assist Interv*. 2006;9(Pt 2):782-9.
46. Janssen MH, Aerts HJ, Ollers MC, Bosmans G, Lee JA, Buijsen J, et al. Tumor delineation based on time-activity curve differences assessed with dynamic fluorodeoxyglucose positron emission tomography-computed tomography in rectal cancer patients. *Int J Radiat Oncol Biol Phys*. 2009 Feb 1;73(2):456-65.
47. Boussion N, Cheze Le Rest C, Hatt M, Visvikis D. Incorporation of wavelet-based denoising in iterative deconvolution for partial volume correction in whole-body PET imaging. *Eur J Nucl Med Mol Imaging*. 2009 Jul;36(7):1064-75.
48. Boussion N, Hatt M, Lamare F, Bizais Y, Turzo A, Cheze-Le Rest C, et al. A multiresolution image based approach for correction of partial volume effects in emission tomography. *Phys Med Biol*. 2006 Apr 7;51(7):1857-76.
49. Aristophanous M, Yap JT, Killoran JH, Chen AB, Berbeco RI. Four-Dimensional Positron Emission Tomography: Implications for Dose Painting of High-Uptake Regions. *Int J Radiat Oncol Biol Phys*. 2010 Oct 13.
50. Liu C, Pierce LA, 2nd, Alessio AM, Kinahan PE. The impact of respiratory motion on tumor quantification and delineation in static PET/CT imaging. *Phys Med Biol*. 2009 Dec 21;54(24):7345-62.
51. Lamare F, Cheze Le Rest C, Visvikis D. Le mouvement respiratoire en Imagerie Fonctionnelle du Cancer: une revue des effets et des méthodes de correction. *Traitement du Signal*. 2006;23.
52. Warfield SK, Zou KH, Wells WM. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans Med Imaging*. 2004 Jul;23(7):903-21.
53. Daisne JF, Duprez T, Weynand B, Lonneux M, Hamoir M, Reyckler H, et al. Tumor volume in pharyngolaryngeal squamous cell carcinoma: comparison at CT, MR imaging, and FDG PET and validation with surgical specimen. *Radiology*. 2004 Oct;233(1):93-100.
54. Yu J, Li X, Xing L, Mu D, Fu Z, Sun X, et al. Comparison of tumor volumes as determined by pathologic examination and FDG-PET/CT images of non-small-cell lung cancer: a pilot study. *Int J Radiat Oncol Biol Phys*. 2009 Dec 1;75(5):1468-74.
55. Dahele M, Hwang D, Peressotti C, Sun L, Kusano M, Okhai S, et al. Developing a methodology for three-dimensional correlation of PET-CT images and whole-mount histopathology in non-small-cell lung cancer. *Curr Oncol*. 2008 Oct;15(5):62-9.
56. Buijsen J, van den Bogaard J, Janssen MH, Bakers FC, Engelsman S, Ollers M, et al. FDG-PET provides the best correlation with the tumor specimen compared to MRI and CT in rectal cancer. *Radiother Oncol*. 2011 Feb;98(2):270-6.

57. Wu K, Ung YC, Hornby J, Freeman M, Hwang D, Tsao MS, et al. PET CT thresholds for radiotherapy target definition in non-small-cell lung cancer: how close are we to the pathologic findings? *Int J Radiat Oncol Biol Phys*. 2010 Jul 1;77(3):699-706.
58. McLennan A, Reilhac A, Brady M. SORTEO: Monte Carlo-based simulator with list-mode capabilities. *Conf Proc IEEE Eng Med Biol Soc*. 2009;2009:3751-4.
59. Jan S, Santin G, Strul D, Staelens S, Assie K, Autret D, et al. GATE: a simulation toolkit for PET and SPECT. *Phys Med Biol*. 2004 Oct 7;49(19):4543-61.
60. Zubal IG, Harrell CR, Smith EO, Rattner Z, Gindi G, Hoffer PB. Computerized three-dimensional segmented human anatomy. *Med Phys*. 1994 Feb;21(2):299-302.
61. Segars WP, Sturgeon G, Mendonca S, Grimes J, Tsui BM. 4D XCAT phantom for multimodality imaging research. *Med Phys*. 2010 Sep;37(9):4902-15.
62. Lartizen C, Kuntner C, Goertzen AL, Evans AC, Reilhac A. Validation of PET-SORTEO Monte Carlo simulations for the geometries of the MicroPET R4 and Focus 220 PET scanners. *Phys Med Biol*. 2007 Aug 21;52(16):4845-62.
63. Lamare F, Turzo A, Bizais Y, Le Rest CC, Visvikis D. Validation of a Monte Carlo simulation of the Philips Allegro/GEMINI PET systems using GATE. *Phys Med Biol*. 2006 Feb 21;51(4):943-62.
64. Le Maitre A, Segars W, Marache S, Reilhac A, Hatt M, Tomei S, et al. Incorporating Patient-Specific Variability in the Simulation of Realistic Whole-Body <sup>18</sup>F-FDG Distributions for Oncology Applications. *Proceedings of the IEEE* 2009;9(12):2026-38.
65. Dice LR. Measures of the Amount of Ecologic Association Between Species. *Ecology*. 1945;26(3):297-302.
66. Geets X, Lee JA, Castadot P, Bol A, Gregoire V. [Potential place of FDG-PET for the GTV delineation in head and neck and lung cancers]. *Cancer Radiother*. 2009 Oct;13(6-7):594-9.
67. Larson SM, Erdi Y, Akhurst T, Mazumdar M, Macapinlac HA, Finn RD, et al. Tumor Treatment Response Based on Visual and Quantitative Changes in Global Tumor Glycolysis Using PET-FDG Imaging. The Visual Response Score and the Change in Total Lesion Glycolysis. *Clin Positron Imaging*. 1999 May;2(3):159-71.
68. Cazaentre T, Morschhauser F, Vermandel M, Betrouni N, Prangere T, Steinling M, et al. Pre-therapy <sup>18</sup>F-FDG PET quantitative parameters help in predicting the response to radioimmunotherapy in non-Hodgkin lymphoma. *Eur J Nucl Med Mol Imaging*. 2010 Mar;37(3):494-504.
69. Lee HY, Hyun SH, Lee KS, Kim BT, Kim J, Shim YM, et al. Volume-based parameter of <sup>18</sup>F-FDG PET/CT in malignant pleural mesothelioma: prediction of therapeutic response and prognostic implications. *Ann Surg Oncol*. 2010 Oct;17(10):2787-94.
70. El Naqa I, Grigsby P, Apte A, Kidd E, Donnelly E, Khullar D, et al. Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. *Pattern Recognit*. 2009 Jun 1;42(6):1162-71.
71. Eary JF, O'Sullivan F, O'Sullivan J, Conrad EU. Spatial heterogeneity in sarcoma <sup>18</sup>F-FDG uptake as a predictor of patient outcome. *J Nucl Med*. 2008 Dec;49(12):1973-9.
72. Frings V, de Langen AJ, Smit EF, van Velden FH, Hoekstra OS, van Tinteren H, et al. Repeatability of metabolically active volume measurements with <sup>18</sup>F-FDG and <sup>18</sup>F-FLT PET in non-small cell lung cancer. *J Nucl Med*. 2010 Dec;51(12):1870-7.
73. Tomei S, Reilhac A, Visvikis D, Bousson N, Odet C, Giammarile F, et al. OncoPET\_DB: A Freely Distributed Database of Realistic Simulated Whole Body <sup>18</sup>F-FDG PET Images for Oncology. *IEEE Transactions on Nuclear Science*. 2010;57(1):246 - 55.

Référence(s)	Méthode	Interaction utilisateurs <sup>1</sup>	Pré et post traitement(s)	Application visée <sup>2</sup>	Données de validation et vérité terrain <sup>3</sup>	Précision sur tumeurs réalistes <sup>5</sup>	Evaluation de la Robustesse <sup>6</sup>	Evaluation de la répétabilité
27	Partage des eaux	Std + placement de plusieurs marqueurs	∅	Globale	AF(1) - Vol. et images CT	Non	Non	Non
28, 29, 30, 31	Gradient	Std + initialisation	∅	Globale	AF(5) - Vol. 25 TSMC - Vol. + TLG 18 TCH - Diam.	Oui	Oui	Oui
32	Gradient	Std + initialisation	Débruitage et déconvolution	Globale	SF et AF - Vol. + Diam. 7 TCH - Complet	Oui	Non	Non
33	SA + contour actif	Std + nécessité de fixer plusieurs paramètres	∅	Globale	AF (1 scanner) - Vol. 1 TC - ∅	Non	Non	Oui
34	Contours actifs multimodaux (TEP/TDM)	Std + initialisation de la forme du modèle déformable, sélection de poids	Normalisation et recalage des données TEP et TDM, déconvolution des images TEP	GTV TEP/TDM	AF - Vol. 2 TC - CM(1), SF	Non	Non	Non
35	Théorie des possibilités sur projections MIP	Std	∅	Globale	AF(1) - Vol. 5 TSMC – Vox. 7 TCH - Complet	Oui	Non	Non
36	Fuzzy C-Means (FCM) amélioré	Interprétation des classes de la segmentation effectuée sur l'image	Débruitage, transformations en ondelettes	Globale	3 TSA – Vox. 21 TCH - Diam. 7 TCH - Complet	Oui	Non	Non

		entière						
39	Mélange de gaussiennes	Std + initialisation du modèle, choix du nombre de classes	∅	Tumeurs pulmonaires	7 TC - ∅	Non	Non	Oui
40	Champs de Markov multi résolution	Choix du nombre de classes + interprétation des classes de la segmentation effectuée sur l'image entière	Transformations en ondelettes	Globale	AF(1) - Vol. 3 TC - ∅	Non	Non	Non
41	Chaînes de Markov floues (FHMC)	Std	∅	Globale	SF(1) et AF(2) - Vox.	Non	Non	Non
18,21,22,26	Fuzzy locally adaptive Bayesian (FLAB)	Std	∅	Globale	SF (1) et AF (4) - Vox. 20 TSMC - Vox. 18 TCH - Diam.	Oui	Oui	Oui
38	Croissance de région sur critères statistiques	Std	Optimisation sur chaque système nécessaire	Tumeurs rectales	18 TC - CM(1)	Non	Non	Non
42, 43	Arbres de décision avec apprentissage sur paramètres de texture TEP/TDM	Interprétation de la segmentation finale réalisée sur l'image entière	Apprentissage et construction des arbres de décision	GTV ORL	10 TC - CM(3)	Non	Non	Non
44	Réseau de neurones	Interprétation des classes de la segmentation	Apprentissage et construction du	Globale	AF(1) - Vol. 3 TSA - Vox.	Non	Non	Oui

		effectuée sur l'image entière	réseau de neurones		1 TCH - Diam.			
45	Algorithme "Spherical Mean shift"	Std	Rééchantillonnage dans un domaine de coordonnées différent	Globale	280 TSA - Vox.	Non	Non	Non
46	Classification des voxels à partir des courbes temps- activité	Std + initialisation et choix du nombre de classes	Imagerie dynamique uniquement, débruitage et déconvolution	Tumeurs rectales TEP et dynamique	AF - Vol. + Diam. 21 TC - CM(1)	Non	Non	Non

Tableau I

## Légende Tableau I

<sup>1</sup> std = interaction « standard » (détection et isolation du VMA)

<sup>2</sup> globale = non restreint à une application

<sup>3</sup> AF(x) = Acquisitions de fantôme sur x scanners différents ; SF(x) = Simulations de fantôme sur x scanners différents ; TSA = Tumeurs simulées de façon analytique ; TSMC = Tumeurs simulées par approche Monte Carlo ; TC = Tumeurs cliniques ; TCH = Tumeurs cliniques avec histopathologie

Vol. = volume uniquement ; Vox. = voxel à voxel ; Diam = diamètre maximum uniquement ; Complet = reconstruction histopathologique 3D; CM(x) = contours manuels (x expert(s)) ; SF = Seuillage fixe

<sup>4</sup> Fortement hétérogènes, formes complexes, faibles contrastes, etc. avec vérité terrain rigoureuse.

<sup>5</sup> Requiert de multiples acquisitions sur plusieurs systèmes et un grand nombre de paramètres.

## Table I legend

<sup>1</sup> std = « standard » interaction (detection and isolation of metabolically active volume)

<sup>2</sup> globale = not application specific

<sup>3</sup> AF(x) = Phantom acquisitions on x different scanners; SF(x) = Phantom simulations on x different scanners; TSA = Analytically simulated tumors; TSMC = Monte Carlo simulated tumors; TC = Clinical tumors; TCH = Clinical tumors with histopathology

Vol. = only volume; Vox. = voxel-by-voxel; Diam = maximum diameter; Complet = 3D histopathology reconstruction; CM(x) = manual contouring (x expert(s)); SF = fixed threshold

<sup>4</sup> Highly heterogeneous, complex shapes, low contrasts, and rigorous ground-truth.

<sup>5</sup> Requires multiple acquisitions on different systems and large number of parameters.

## Légendes des figures

Figure 1 : erreurs de classification (%) obtenues par des seuillages sur simulations GATE (59). (a) Vérité terrain simulée et (b) binarisée pour le calcul d'erreurs, (c-f) simulations avec différents niveaux de contraste et de bruit.  $T_{42}$  et  $T_{50}$  : seuils fixes à 42 et 50% du maximum.  $T_{\text{bckgd}}$  et  $T_{\text{SBR}}$  : seuillages adaptatifs de Nestle et Daisne (19).

Figure 2 : acquisition TEP corps entier au  $^{18}\text{F}$ -FDG reconstruite avec des voxels de (a)  $2 \times 2 \times 2 \text{ mm}^3$  et (b)  $4 \times 4 \times 4 \text{ mm}^3$ .

Figure 3 : image TEP d'une tumeur pulmonaire. (a) Originale ( $5.31 \times 5.31 \times 5 \text{ mm}^3$ ) et interpolée sur des voxels de  $1 \text{ mm}^3$  par approche (b) linéaire et (c) B splines bi-cubiques.

Figure 4 : (a-c) tumeurs pulmonaires sur (a) Philips Gemini, (b) Siemens Biograph et (c) GE Discovery LS, (d) tumeur œsophage et (e) rectale sur Philips Gemini, et (f) tumeur du sein avec  $^{18}\text{F}$ -FLT sur Siemens Biograph.

Figure 5 : sur des tumeurs similaires et fortement hétérogènes, (a) [issue de (31)] la méthode par gradient (rose) qui englobe toute la tumeur sans différencier le cœur du pourtour et un seuil à 37% (bleu) qui ne prend que le pourtour, et (b) FLAB avec 3 classes (26) qui définit le contour externe, le cœur nécrosé, et la fixation de plus haute activité.

Figure 6 : sur une même tumeur hétérogène, (a) FCM amélioré (36) et (b) FLAB (26). Les contours rouge, jaune et vert correspondent respectivement au FCM simple, au FCM-S (avec corrélation spatiale), et au FCM-SW (rajoutant la gestion des hétérogénéités), qui sous-évalue la partie droite de la tumeur.

## Figures legends

Figure 1: classification errors (%) obtained with thresholds on GATE simulations (59). (a) Simulated and (b) binarized (for errors computation) ground-truth, (c-f) simulations with various contrast and noise levels.  $T_{42}$  et  $T_{50}$ : fixed thresholds at 42 and 50% of the maximum.  $T_{\text{bckgd}}$  and  $T_{\text{SBR}}$ : Nestle's and Daisne' adaptive thresholds (19).



Figure 2:  $^{18}\text{F}$ -FDG PET whole-body acquisition reconstructed with (a)  $2 \times 2 \times 2 \text{ mm}^3$  and (b)  $4 \times 4 \times 4 \text{ mm}^3$  voxels.

Figure 3: PET image of a pulmonary tumor. (a) Original ( $5.31 \times 5.31 \times 5 \text{ mm}^3$ ) and resampled to  $1 \text{ mm}^3$  voxels using (b) linear and (c) bi-cubic B-splines methods with.

Figure 4: (a-c) pulmonary tumors on (a) Philips Gemini, (b) Siemens Biograph and (c) GE Discovery LS, (d) esophageal and (e) rectal tumors on Philips Gemini, and (f) breast tumor with  $^{18}\text{F}$ -FLT on Siemens Biograph.

Figure 5: on similar highly heterogeneous tumors, (a) [from (31)] gradient-based method (pink) contouring the entire tumor without differentiating the core and a 37% threshold (blue) contouring the exterior only, and (b) FLAB with 3 classes (26) defining the external contour, the necrotic core as well as the highest activity region.

Figure 6 : on the same heterogeneous tumor, (a) improved FCM [from (36)] and (b) FLAB (26). Red, yellow and green contours are those of basic FCM, FCM-S (adding spatial correlation) and FCM-SW (adding heterogeneity modeling) respectively, which underestimates the right part of the tumor.

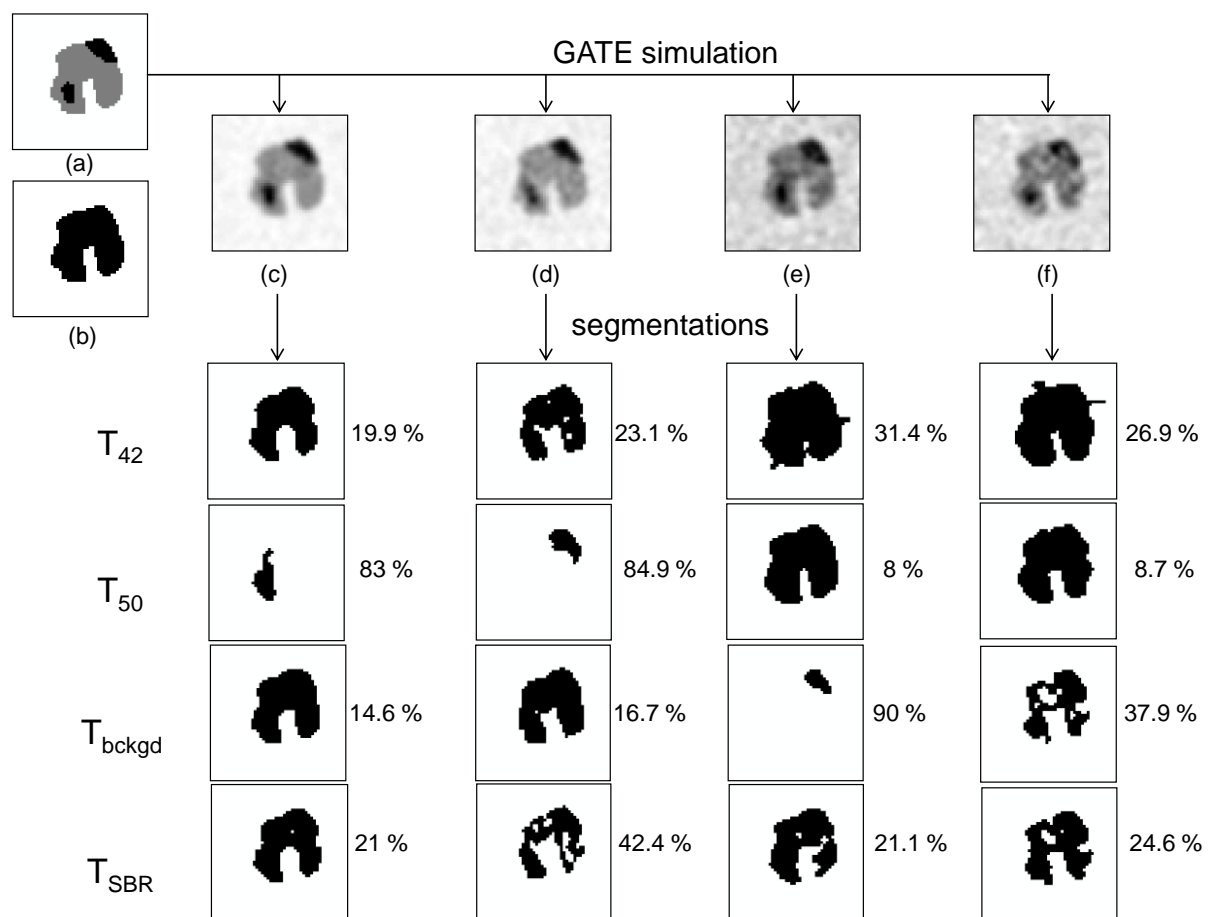


Figure 1

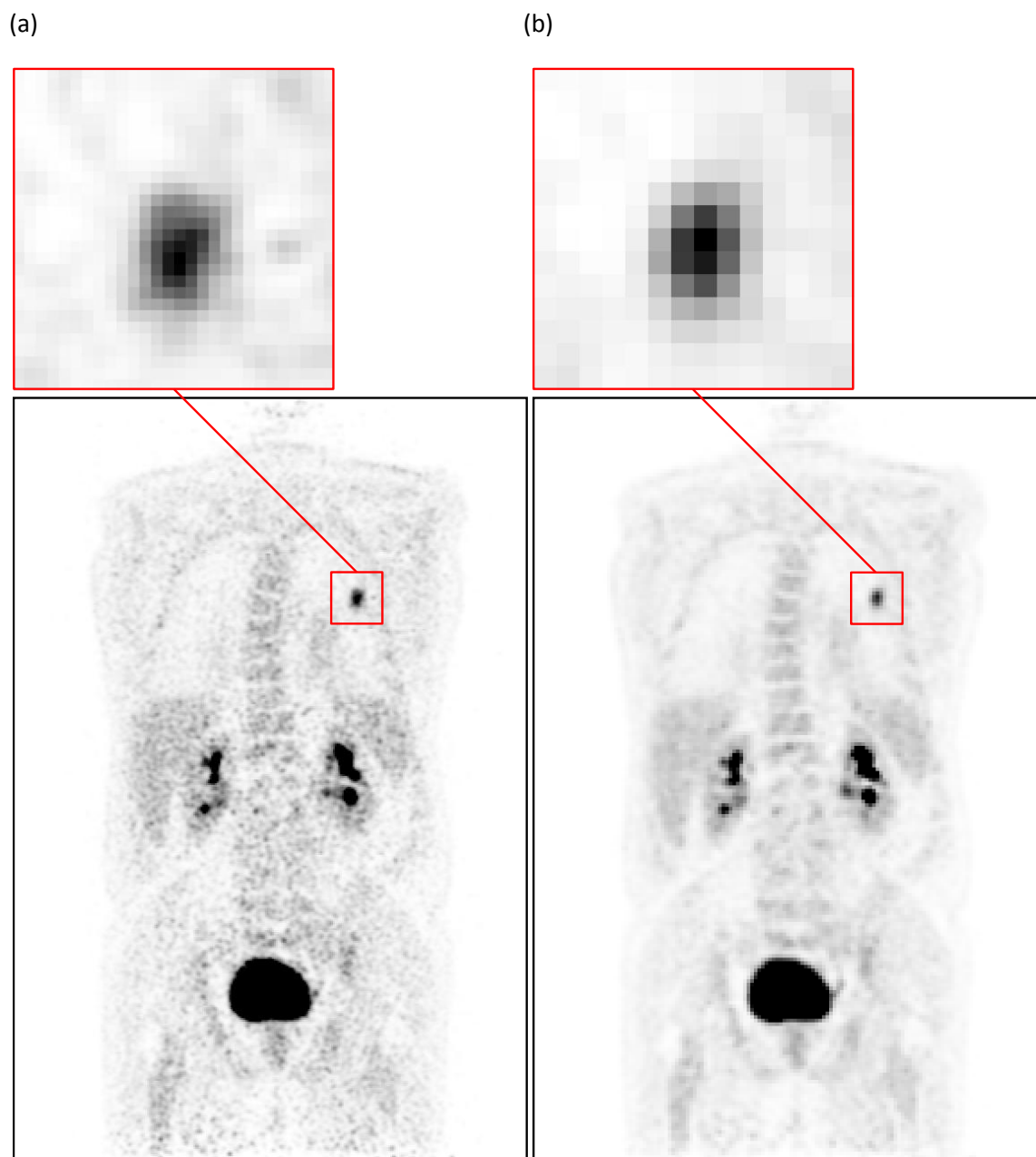


Figure 2

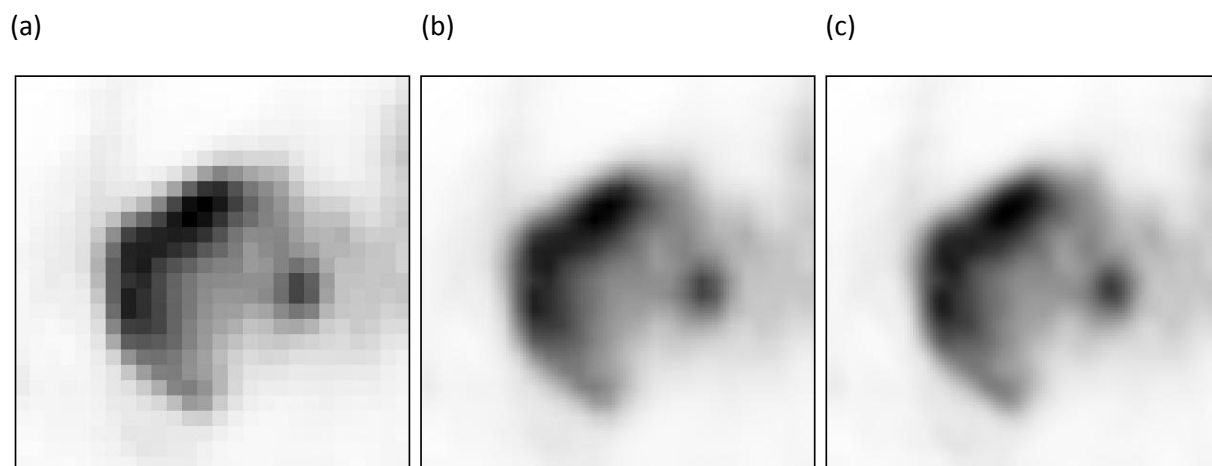


Figure 3

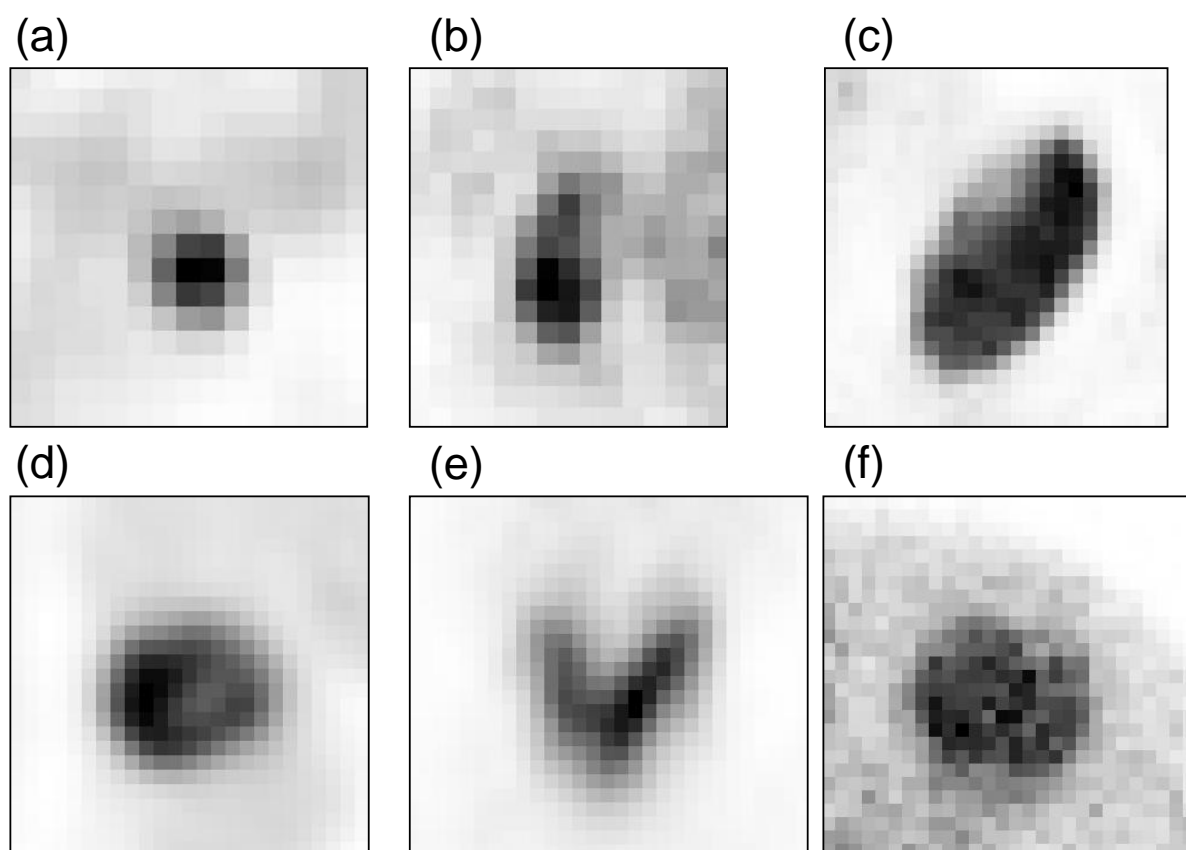
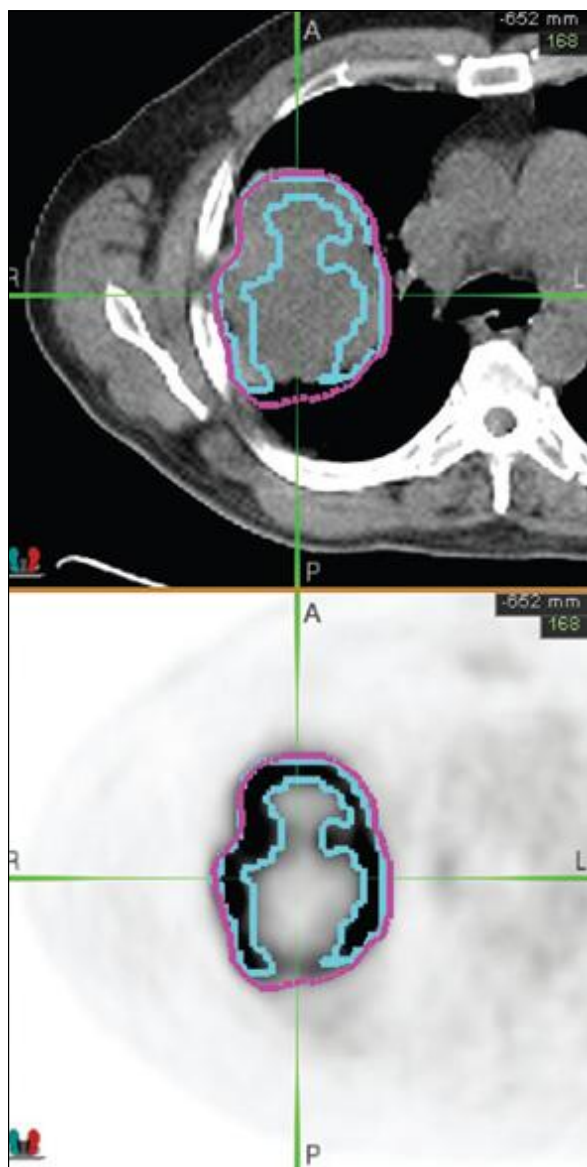
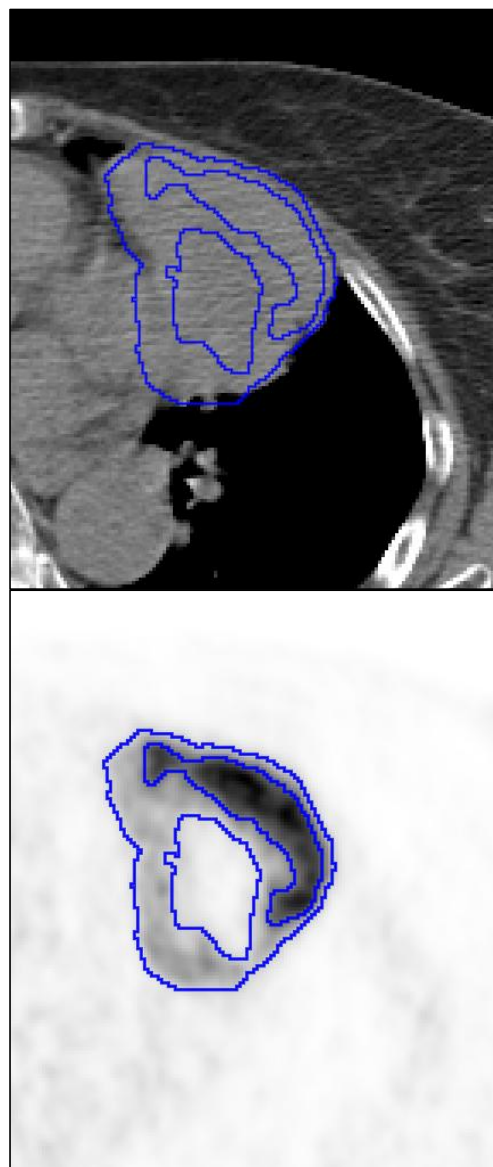


Figure 4



(a)



(b)

Figure 5

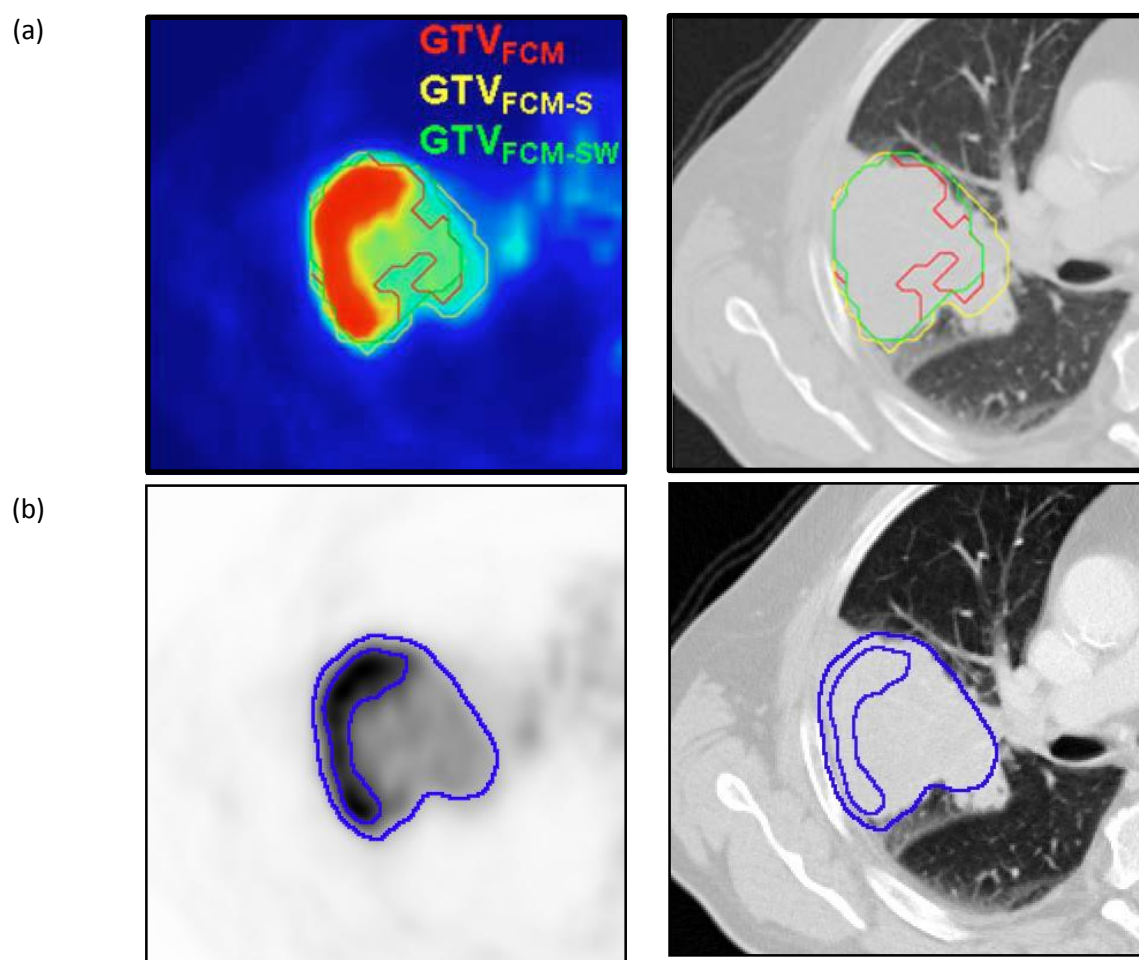


Figure 6