

ORIGINAL ARTICLE

Open Access



AI-based quantification of whole-body tumour burden on somatostatin receptor PET/CT

Anni Gålne^{1,2,3*} , Olof Enqvist^{4,5}, Anna Sundlöv⁶, Kristian Valind^{1,2,3}, David Minarik^{2,3,7} and Elin Trägårdh^{2,3,8}

*Correspondence:
anni.galne@med.lu.se

¹ Department of Medical Imaging and Physiology, Skåne University Hospital, Lund, Sweden

² Department of Translational Medicine, Faculty of Medicine, Lund University, Malmö, Sweden
³ WCMM Wallenberg Centre for Molecular Medicine, Lund, Sweden

⁴ Eigenvision AB, Malmö, Sweden

⁵ Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, Sweden

⁶ Department of Clinical Sciences, Oncology and Pathology, Lund University, Lund, Sweden

⁷ Radiation Physics, Skåne University Hospital, Malmö, Sweden

⁸ Department of Medical Imaging and Physiology, Skåne University Hospital, Malmö, Sweden

Abstract

Background: Segmenting the whole-body somatostatin receptor-expressing tumour volume (SRETVwb) on positron emission tomography/computed tomography (PET/CT) images is highly time-consuming but has shown value as an independent prognostic factor for survival. An automatic method to measure SRETVwb could improve disease status assessment and provide a tool for prognostication. This study aimed to develop an artificial intelligence (AI)-based method to detect and quantify SRETVwb and total lesion somatostatin receptor expression (TLSREwb) from [⁶⁸Ga]Ga-DOTA-TOC/TATE PET/CT images.

Methods: A UNet3D convolutional neural network (CNN) was used to train an AI model with [⁶⁸Ga]Ga-DOTA-TOC/TATE PET/CT images, where all tumours were manually segmented with a semi-automatic method. The training set consisted of 148 patients, of which 108 had PET-positive tumours. The test group consisted of 30 patients, of which 25 had PET-positive tumours. Two physicians segmented tumours in the test group for comparison with the AI model.

Results: There were good correlations between the segmented SRETVwb and TLSREwb by the AI model and the physicians, with Spearman rank correlation coefficients of $r=0.78$ and $r=0.73$, respectively, for SRETVwb and $r=0.83$ and $r=0.81$, respectively, for TLSREwb. The sensitivity on a lesion detection level was 80% and 79%, and the positive predictive value was 83% and 84% when comparing the AI model with the two physicians.

Conclusion: It was possible to develop an AI model to segment SRETVwb and TLSREwb with high performance. A fully automated method makes quantification of tumour burden achievable and has the potential to be more widely used when assessing PET/CT images.

Keywords: AI, Somatostatin receptor-expressing tumour volume, [⁶⁸Ga]Ga-DOTA-TATE, [⁶⁸Ga]Ga-DOTA-TOC, PET/CT

Background

Neuroendocrine neoplasms (NENs) can be divided into often more indolent neuroendocrine tumours (NETs) and more aggressive neuroendocrine carcinomas (NECs) depending on morphology and proliferation rate (WHO 2019). Well-differentiated NETs

typically overexpress somatostatin receptors (SSTR), which can be targeted in imaging and treatment (Bozkurt et al. 2017; Kaderli et al. 2019). An important imaging method for visualizing SSTR-expressing tumours is positron-emitting ^{68}Ga -labelled somatostatin analogues such as [^{68}Ga]Ga-DOTA-TOC, [^{68}Ga]Ga-DOTA-TATE or [^{68}Ga]Ga-DOTA-NOC during hybrid imaging with positron emission tomography and computed tomography (PET/CT) (Bozkurt et al. 2017).

Clinical interpretation of PET/CT images is a visually subjective assessment, sometimes with manual measurements of maximum standardized uptake value (SUVmax) in PET images and manual measurements of the largest tumour diameter in the CT images. Although the interpretation of SSTR PET/CT has shown good inter-observer agreement, differences between readers still exist and can be clinically significant, particularly when selecting patients for peptide receptor radionuclide therapy (PRRT) (Fendler et al. 2017). High tumour uptake correlates with well-differentiated tumours, longer progression-free survival (PFS), and improved response to PRRT (Campana et al. 2010; Kratochwil et al. 2015). However, SUVmax has several disadvantages, including its representation of only one voxel within the volume of interest (VOI), which means that the whole tumour burden is not represented, and its sensitivity to noise (Foster et al. 2014). Manual or semi-automatic drawing of volumetric measurements has numerous weaknesses, particularly time constraints and inter- and intra-observer variance (Foster et al. 2014). In the clinical setting, the evaluation of PET/CT images by nuclear medicine specialists and radiologists is challenging because of the time-intensive nature of the manual analysis, lack of quantification, and limitations in reproducibility regarding evaluation and quantification.

A method for measuring somatostatin receptor-expressing tumour volume (SRETV) and total lesion somatostatin receptor expression (TLSRE) on [^{68}Ga]Ga-DOTA-TATE PET/CT has previously been suggested (Abdulrezzak et al. 2016). SRETV was measured as the tumour volume measuring more than 50% of SUVmax in a VOI, while TLSRE is calculated as the product of SRETV and SUVmean of each lesion (Abdulrezzak et al. 2016). Toriihara et al. evaluated the sum of whole-body SRETV (SRETVwb) and the sum of whole-body TLSRE (TLSREwb) concerning PFS. A larger tumour burden, measured as SRETVwb, showed significantly shorter PFS and may therefore have prognostic value (Toriihara et al. 2019). Similarly, other methods for measuring SRETVwb and TLSREwb have shown that higher tumour volume independently correlates with shorter PFS in both prospective and retrospective studies (Tirosh et al. 2017; Thuillier et al. 2022; Chen et al. 2023).

Deep learning is a subfield of machine learning, and artificial intelligence (AI) is a broader term. In the speciality of image analysis, the primary deep learning method chosen is convolutional neural networks (CNN) (Goodfellow and Courville 2016). CNNs are provided with large amounts of data, and the learning procedure imitates how information is processed through the brain, with some information processed through convolutions in numerous layers that activate different “neurons,” and in the end, there is an output, often as a set of classifiers, such as tumour lesion; yes or no (Alzubaidi et al. 2021).

A method to detect hepatic lesions on [^{68}Ga]Ga-DOTA-TATE PET/CT with AI has been developed by Wehrend et al. (2021), and more recently, Carlsen et al. developed

and implemented a CNN-based method to segment total tumour burden on [⁶⁴Cu] Cu-DOTA-TATE PET/CT (Carlsen et al. 2022).

An objective automated method to detect and quantify SSTR-expressing tumour burden could improve the assessment of disease status and response evaluation, provide a tool for prognostication and potential prediction of treatment response, and improve the evaluation of response. Therefore, this study aimed to develop a method to detect and quantify SRETVwb and TLSREwb from [⁶⁸Ga]Ga-DOTA-TOC and -TATE PET/CT images using CNNs.

Methods

Adults (≥ 18 years) with a clinical indication for [⁶⁸Ga]Ga-DOTA-TOC or -TATE PET/CT between August 2017 and December 2021, who were included in our larger study for validation of PET/CT, were eligible for this retrospective image analysis. Among these 848 patients, a subset of 200 patients were arbitrary selected for further analysis. The selection was made in a random manner without taking clinical or imaging data into account. Due to the time-intensive nature of segmenting tumours, including all patients was impossible. Exclusion criteria for further analysis in this study were inclusion in a future study planned to validate this method, incomplete examination, or larger extravasation of radiotracer outside the patient, such as urine. The study was carried out following the Declaration of Helsinki and was approved by the Swedish Ethical Review Authority (EPN LU 2016/417, 2018/753, and 2021-05734-02). All patients provided written informed consent. Data on age, sex, medical referral, and type of radiopharmaceutical were available for all patients. Clinical data on histopathological diagnosis, Ki-67, type of earlier or ongoing treatment, and TNM stage were verified by reviewing the patient's digital medical record in the test group.

PET/CT protocols

The PET/CT scans were performed at Skåne University Hospital in Lund using a Discovery MI or Discovery D690 (GE Healthcare, Milwaukee, USA) PET/CT system. In 2019, there was a shift in production from [⁶⁸Ga]Ga-DOTA-TATE to [⁶⁸Ga]Ga-DOTA-TOC, which is why both radiotracers were included in this study. The radiotracers were prepared according to established techniques (Bozkurt et al. 2017; Zhernosekov et al. 2007; Mueller et al. 2012). The patient received an intravenous injection of 2.0–2.5 MBq/kg activity (minimum administered activity 100 MBq and maximum 300 MBq), followed by a PET/CT scan approximately 60 min later. The scan covered from the base of the skull to mid-thigh. The PET acquisition time was 3.0–3 min 15 s per bed position, depending on the radiotracer and PET/CT system. Both systems used time-of-flight and point-spread function correction, and either a low-dose CT scan or a diagnostic CT with intravenous contrast (if there were no contraindications) was performed for attenuation correction and anatomic correlation.

Manual image analyses

The images had previously been analysed by an experienced nuclear medicine physician and a radiologist in a clinical setting. For the current study, retrospective segmentation of tumours was performed in consensus by a PhD student in nuclear medicine/

senior radiology resident (AG) and an experienced nuclear medicine physician (ET) using Hermes software (Hermes Medical Solutions, Stockholm, Sweden) and correlated with the clinical report. Uptake of [⁶⁸Ga]Ga-DOTA-TOC and -TATE was considered significant for tumour segmentation if it did not correspond to physiological uptake. The PET/CT examination clinical report was used as a reference for segmentation. No predefined threshold for SUVmax was used. A semi-automatic method was used to delineate the tumours on the PET/CT images, where SRETV was defined as the tumour volume with uptake higher than 50% of SUVmax within a VOI, and TLSRE was defined as the product of SRETV and mean SUV (SUVmean) of the tumour volume (Abdulrezak et al. 2016). Due to relatively high normal background uptake in the liver, manually drawn VOIs were often needed to avoid physiological uptake, as previously described (Toriihara et al. 2019). Overlap between tumour volumes was avoided. When a confluent uptake from closely related lesions was present, manually drawn VOIs around each lesion guided by the CT were needed to separate the lesions and avoid uptake from adjacent lesions. Inside the VOI, an irregular tumour volume with voxels higher than 50% of SUVmax was automatically obtained.

Whole-body tumour burden was defined as the sum of all SRETV and was described as SRETVwb. Whole-body somatostatin receptor expression was defined as the sum of all TLSRE and was described as TLSREwb. The semi-automatic tumour segmentations in the test group were separately performed by two physicians for comparison: Reader A (AG, ground truth) and Reader B (KV). Reader B was blinded to clinical parameters in correlation with the AI model.

Training the AI model

At the core of our AI model was a UNet3D CNN that takes four types of input: a CT patch, a SUV patch, an organ mask patch (Trägårdh et al. 2020), and a SUV ratio patch with the ratio of the pixel SUV to the closest local maximum. The last patch was intended to enable thresholding at 50% of the lesion SUVmax (as used for the annotations). Meyer's flooding algorithm on the SUV image determined the closest local maximum.

Before feeding them to the network, all inputs were resampled to $2.5 \times 2.5 \times 2.5$ mm. CT patches were resampled using trilinear interpolation, but the nearest neighbour was used for SUV data and manual target segmentations, as the desired segmentation might depend on the exact relationship between SUV pixel values. The neural network output has the same voxel size, and to avoid introducing artefacts, this voxel size is maintained during post-processing and statistical analysis. The latter is first resampled to the CNN resolution for pixel-wise comparison to manual segmentations. Resampling the input patches also creates a consistent receptive field size for the network, regardless of the native image resolution. A UNet3D has a receptive field size of roughly 100 pixels (in any dimension), corresponding to around 25 cm in PET/CT images.

Before training of the AI model, the PET/CT examinations were divided into groups for training, validation, and testing. Random sampling was used for splitting the dataset into these groups. The number of patients in each group was set to 17% for testing and 17% for validation and the rest for training, according to the recommended ratio of dividing the training set in approximately 80% for training and 20% for validation (Courville IJGaYBaA. 2016). Usually the same number of data is used in the test group as in

the validation group (Hastie et al. 2001). Training a neural network involves providing the network with a sequence of annotated image patches, and the way patches are chosen can affect the resulting network’s performance. Throughout the training, efforts are made to sample all images equally and preserve the balance between cancer and non-cancer images. Starting from uniform sampling within each class, the sampling of background voxels was updated in the same way as previously described in Trägårdh et al., while uniform sampling was maintained for the foreground (Trägårdh et al. 2022).

For the actual optimization, the Adam optimizer with Nesterov momentum was used. The learning rate was initialized to 10^{-4} and reduced by a factor of 0.95 for every 20,000 samples until it reaches 10^{-7} . To reduce overfitting, L2 regularization with a weight of 2×10^{-4} and a single dropout layer with a probability of 0.1 was used. The UNet3D architecture is illustrated in Fig. 1.

Connected components in the lesion mask were referred to as lesions. Post-processing was performed with the automatic removal of any lesions with a volume less than 0.05 ml, as it was found on the validation set that these were often false positives.

Statistical analysis

A lesion having partial or full overlap with the reference segmentation was defined as a true positive lesion. A segmented lesion with no overlap with the reference segmentation was defined as a false positive lesion. Lastly, a lesion in the reference segmentation not segmented by the compared segmentation was defined as a false negative lesion. True negative lesions were excluded as it is impossible to define true negative lesions that were not detected. Sensitivity was calculated on a lesion detection level as the percentage of detected lesions compared to the reference segmentation. Positive predictive value (PPV) was calculated as the percentage of true positive lesions divided by the sum of true positive and false positive lesions, compared to the reference. Specificity could not be calculated since the AI model did not detect true negative lesions.

Calculations for true positives, false positives, false negatives, sensitivity, and PPV were made by comparing the AI model segmentations with both Reader A and Reader B, and Reader B with the ground truth Reader A.

The AI model’s measurements and the Readers’ measurements of SRETVwb and TLSREwb were calculated and compared. The correlation between the AI model’s measurements and the Readers’ measurements of SRETVwb and TLSREwb was assessed using Spearman rank correlation with a 2-tailed test, and it was considered statistically

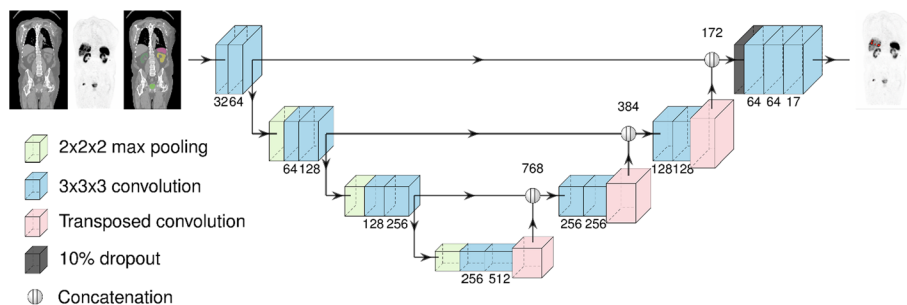


Fig. 1 Illustration of the UNet3D architecture

significant with a p -value < 0.05 . The correlation was considered very strong for Spearman’s coefficient $r > 0.8$, moderately strong for values between 0.6 and 0.8, fair for values between 0.3 and 0.5, and poor for values below 0.3 (Chan 2003). Bland–Altman analysis was used to visually assess the level of agreement between the AI model and the manual SRETVwb and TLSREwb measurements. IBM SPSS Statistics version 28 was used for all statistical analyses.

Results

Study population

The study involved 200 PET/CT scans from 200 patients who underwent a clinically necessary $[^{68}\text{Ga}]\text{Ga-DOTA-TOC}$ or $-TATE$ PET/CT at Skåne University Hospital between August 2017 and December 2021. Of these, 22 patients were excluded based on the exclusion criteria. The CONSORT diagram (Consolidated Standards of Reporting Trials) (Fig. 2) provides additional information. Ultimately, 178 patients were included and evaluated for SSTR-expressing tumour burden. The training and validation set comprised PET/CT scans from 148 patients, of whom 108 patients had suspected tumour lesions segmented, and 40 patients had no SSTR-expressing tumours. The radiotracer used in the training and validation set was $[^{68}\text{Ga}]\text{Ga-DOTA-TATE}$ in 44 examinations and $[^{68}\text{Ga}]\text{Ga-DOTA-TOC}$ in 104 examinations.

Test group

The test group comprised 30 patients whose characteristics are presented in Table 1. Sixteen patients were female, and the mean age at the time of PET/CT was 66.4 years.

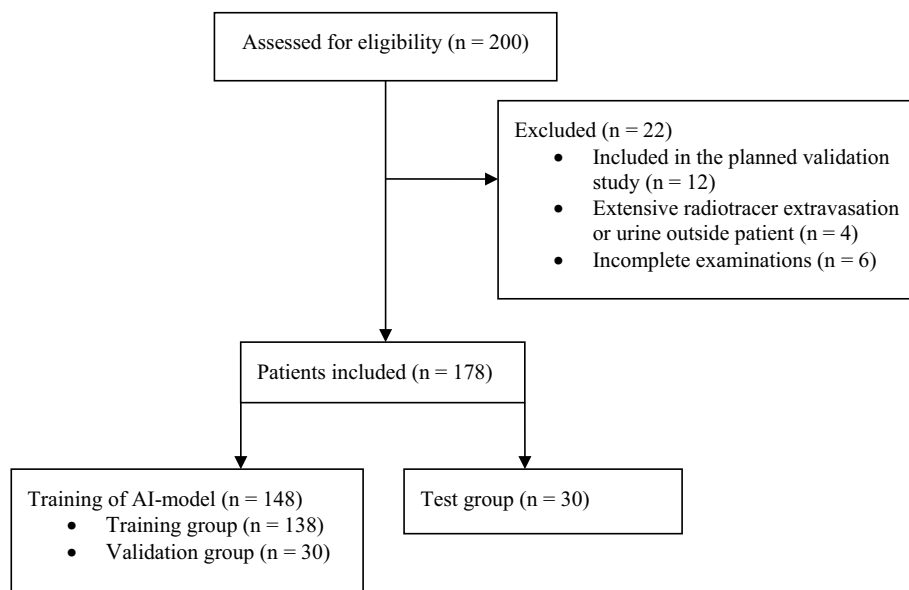


Fig. 2 CONSORT diagram of the retrospective study

Fig. 2 CONSORT diagram of the retrospective study

Table 1 Patient characteristics in test group

Characteristic	Value (n = 30)
Sex	
Female	16
Mean age at PET/CT (y)	66.4 (SD 12.8)
<i>PET with</i>	
DOTA-TATE	9
DOTA-TOC	21
Low-dose CT	7
Diagnostic CT with iv. contrast	21
Diagnostic CT without iv. contrast	2
<i>Reason for PET/CT</i>	
First examination	7
Follow-up after treatment	13
Follow-up after surgery	1
Suspected progressive disease	5
Evaluation before decision about PRRT	4
<i>Primary tumour</i>	
No verified NET and no SSTR-expressing lesion	3
Small bowel	18
Pancreas	3
Medullary thyroid cancer	2
Stomach	1
MEN 1	1
Atypical carcinoid	1
Small cell neuroendocrine carcinoma	1
<i>Ki-67 (%)</i>	
< 3	17
3 to ≤ 20	7
> 20	2
Unknown	1
<i>Earlier treatment before PET/CT</i>	
Surgery	22
PRRT	2
Chemotherapy	2
Somatostatin analog	1
Ongoing treatment with somatostatin analog	13
<i>TNM stage at PET/CT</i>	
Only primary tumour	5
Loco-regional disease	2
Metastatic	18
NET verified pathological diagnose but no remaining tumour on PET/CT	2
No NET	3
<i>Distribution of somatostatin receptor-expressing lesions*</i>	
Pathological uptake	25
Liver	7
Lung	3
Bone	9
Abdomen (except liver)	9
Lymph nodes	10
Other locations	4
No pathological uptake	5

Table 1 (continued)

*The total number of the distribution of somatostatin receptor-expressing lesions is more than 25 as patients could have uptake on several locations

Before the PET/CT, 23 patients had an established NEN diagnosis, and the mean age since diagnosis was 5.6 years (standard deviation, SD 4.0). Four patients were diagnosed after the PET/CT, bringing the total number of confirmed NEN diagnoses to 27. Of these, 15 patients had grade 1, five had grade 2 and one had grade 3. In six patients, the grade was unknown or not applicable.

Performance of the AI model

Out of the 27 patients diagnosed with NEN, 25 had SSTR-expressing tumour lesions, and the AI model classified 24 of these patients as positive and 3 out of 5 patients as negative. When Reader A was used as a reference, Reader B classified 24 out of 25 patients as true positive and 5 as true negative. Details on image classification at the patient level are summarized in Table 2. Reader A (ground truth) segmented 267 lesions on a lesion detection level, with a median of 3 lesions per patient and a range of 0–58 lesions. The median segmented tumour volume SRETVwb was 6.8 ml for Reader A, 4.7 ml for the AI model, and 6.5 ml for Reader B. Additional details on segmented lesions and tumour volume are displayed in Table 3. The sensitivity on a lesion detection level was 80% for the AI model versus Reader A and 79% versus Reader B. When comparing Reader B versus Reader A, the sensitivity was 92%. Details on sensitivity and PPV are presented in Table 4.

There was a moderately strong correlation ($r = 0.78, p < 0.001$) between SRETVwb for the AI model versus Reader A (Fig. 3a) and a very strong correlation ($r = 0.83,$

Table 2 Classification of images on patient level

	AI model versus Reader A	AI model versus Reader B	Reader B versus Reader A
Correct classification of PET-CT image	27	26	29
True positive patient	24	23	24
True negative patient	3	3	5
False positive patient	2	3	0
False negative patient	1	1	1

True positive, true negative, false positive and false negative classification of images on patient level when using manual segmentation or reference doctor as comparison. The results are presented for the entire test group ($n = 30$)

Table 3 Segmented lesions in the test group

	Reader A	AI model	Reader B
Total number of segmented lesions	267	265	269
Number of lesions per patient (IQR)	3.0 (1.0–14.8)	3.5 (1.75–14.8)	2.5 (1.0–16.3)
SRETVwb (IQR)	6.8 (1.0–35.8)	4.7 (0.5–28.8)	6.5 (0.5–42.8)
TLSREwb (IQR)	61.9 (7.6–583.4)	40.7 (3.4–562.1)	65.1 (2.1–605.0)

Total number of segmented lesions in the test group ($n = 30$) presented for Reader A, the AI model and reference Reader B. Numbers are total number of lesions, median number of lesions per patient, whole-body somatostatin receptor-expressing tumour volume (SRETVwb) in ml and whole-body total lesion somatostatin receptor expression TLSREwb, with interquartile range (IQR)

Table 4 Classification of segmented lesions

	AI model versus Reader A	AI model versus Reader B	Reader B versus Reader A
<i>True positive lesions</i>			
Total	214	210	246
Per patient (IQR)	2 (0.0–12.0)	1 (0.0–12.0)	2.5 (1.0–14.5)
<i>False positive lesions</i>			
Total	45	41	25
Per patient (IQR)	1 (0.0–2.3)	1 (0.0–2.0)	0.0 (0.0–1.3)
<i>False negative lesions</i>			
Total	53	59	21
Per patient (IQR)	1 (0.0–2.0)	1 (0.0–3.3)	0.0 (0.0–1.0)
Sensitivity (%)	80	79	92
PPV (%)	83	84	91

Number of lesions which are true positive, false positive and false negative for the AI model versus Reader A, the AI model versus Reader B and Reader B versus Reader A. Values are total number of lesions, median number of lesions per patient, median volume (ml) and median TLSREwb with interquartile range (IQR). Sensitivity and positive predictive value (PPV) are also presented for each comparison

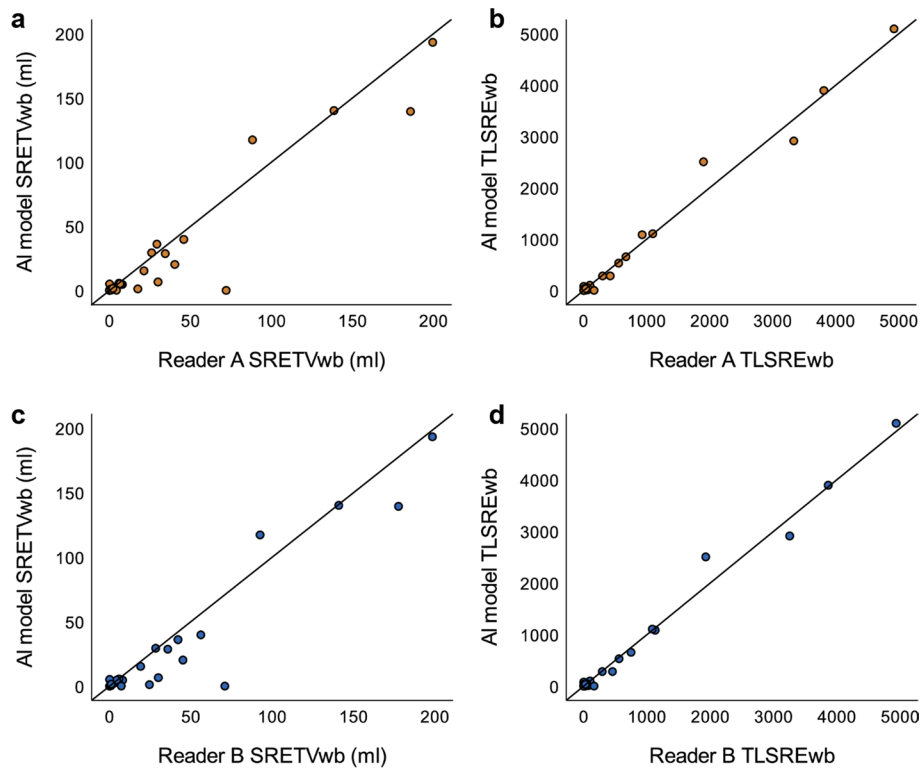


Fig. 3 Scatter plots illustrating the correlation between the AI model and Reader A for SRET Vwb, ($r=0.78$, $p<0.001$) (a), respectively, TLSREwb ($r=0.83$, $p<0.001$) (b) and between the AI model and Reader B for SRET Vwb ($r=0.73$, $p<0.001$) (c), respectively, TLSREwb ($r=0.81$, $p<0.001$) (d)

$p<0.001$) for TLSREwb (Fig. 3b). The values were similar for the AI model versus Reader B for both SRET Vwb (Fig. 3c) and TLSREwb (Fig. 3d). The correlation between Reader B versus Reader A was even higher, with $r=0.96$ ($p<0.001$) for SRET Vwb

and $r = 0.99$ ($p < 0.001$) for TLSREwb. To demonstrate the agreement between the AI model versus Reader A and the AI model versus Reader B, Bland–Altman plots were created (Fig. 4).

The AI model was less accurate in two patients, in which it segmented larger volumes of healthy organs, including part of the kidney in one patient and part of the bladder in the other. The likely reason for these mistakes was the presence of large benign cysts in the kidney of one patient, resulting in an unusual organ shape. In the other patient, a unilateral hip prosthesis caused metal artefacts over the bladder, making it difficult to distinguish on the CT images. An example of the error in the segmented bladder is illustrated in Fig. 5. Additionally, the AI model had difficulty segmenting mediastinal tumours, which are rarer than abdominal tumours and, therefore, less frequent in the training data. Similarly, tumours with very low uptake were challenging for the AI model to segment. One patient classified as negative by Reader A and B and positive by the AI model had only a tumour volume of 0.09 ml segmented by the AI model. These pixels were located in the spleen, but the CT images had significant breathing artefacts, which might have affected this segmentation. Overall, the segmentations by the AI model were consistent with those of Reader A and B, with some of the best examples of tumour segmentations demonstrated in Fig. 6.

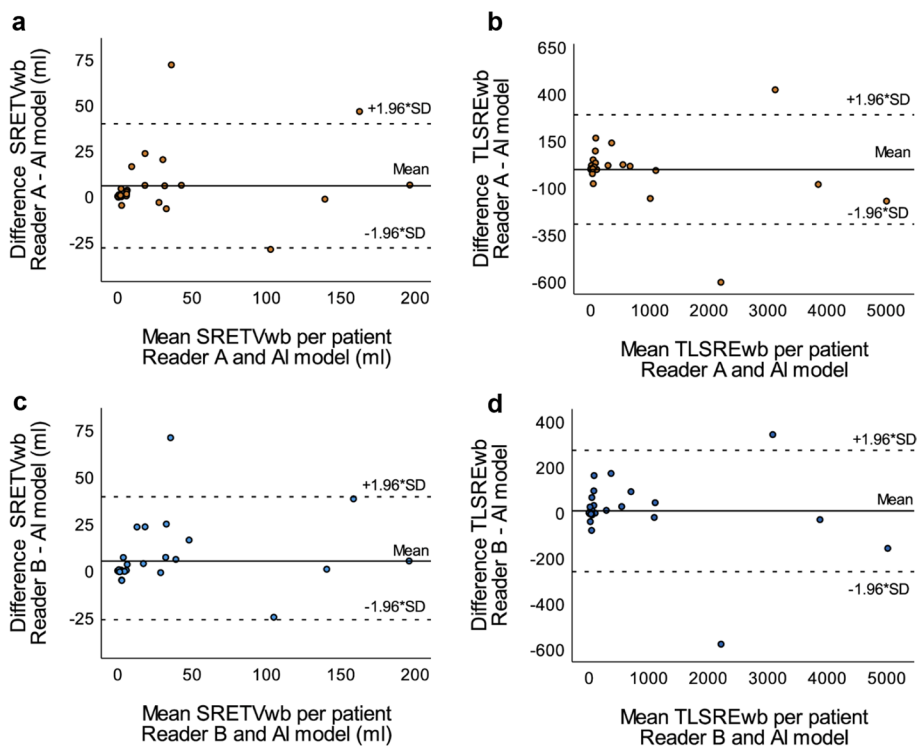


Fig. 4 Bland–Altman plots illustrating the agreement of SRET_{Vwb} between the AI model and Reader A (a) and Reader B (c), respectively, as well as the agreement of TLSRE_{wb} between the AI model and Reader A (b) and Reader B (d). Dotted lines specify the 95% upper and lower limits of agreement

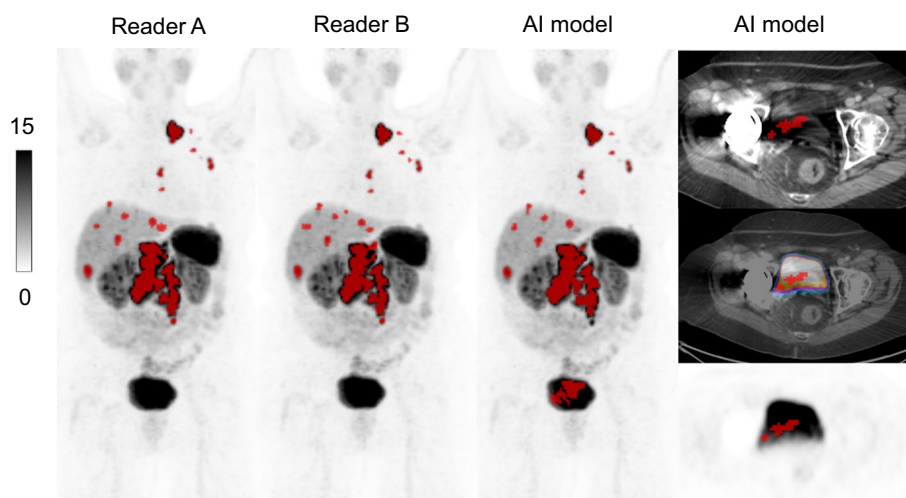


Fig. 5 MIP-images with tumour segmentation for Reader A, Reader B and the AI model, respectively. To the right axial images over bladder with the AI model segmentations. A part of the bladder was included as tumour volume by the AI model. A unilateral hip prosthesis caused metal artefacts over the pelvic area, which might have contributed to this error. SRETVwb was 88.2 ml for Reader A, 92.5 ml for Reader B and 117.2 ml for the AI model

Discussion

Segmenting the whole-body tumour burden is time-consuming but has shown value in predicting PFS (Toriishara et al. 2019; Thuillier et al. 2022). It could be an important predictive factor when selecting patients for PRRT, individualizing PRRT, for prognostication and during follow-up (Ebberts et al. 2021). This is the first study to develop a completely automatic AI model to measure the whole-body somatostatin receptor-expressing tumour volume and whole-body total lesion somatostatin receptor expression (SRETVwb and TLSREwb) on [^{68}Ga]Ga-DOTA-TOC and -TATE PET/CT images. Carlsen et al. recently developed a similar method for [^{64}Cu]Cu-DOTA-TATE PET/CT, showing a similar lesion-wise sensitivity of 84.4% compared to 80% in our model. However, they excluded negative control patients from the test group, which makes the results not entirely comparable (Carlsen et al. 2022).

Our AI model's measurements of SRETVwb and TLSREwb in the test group correlated strongly to very strongly with the segmentations made by Reader A and B. The AI model made a few obvious errors, but in those cases, reasonable explanations could be found, such as metal artefacts, large breathing artefacts, multiple cysts making the anatomy very different, or very low uptake in tumours. A large amount of training data is crucial, and a larger training set might improve these errors. Still, as this disease is quite rare (Dasari et al. 2017), it is also a limiting factor for training the AI model. As all patients are individuals, there could always be specific factors that an AI model has not encountered before, and therefore, it seems unlikely that AI tools will be able to stand alone in the near future. Approval of the segmentations by a physician is needed when implementing an AI model, as described by Carlsen et al. (2022). Additionally, post-processing using a higher cut-off volume than 0.05 ml for removal of larger lesions, such as 0.1 ml, as compared by Carlsen et al. (2022), might

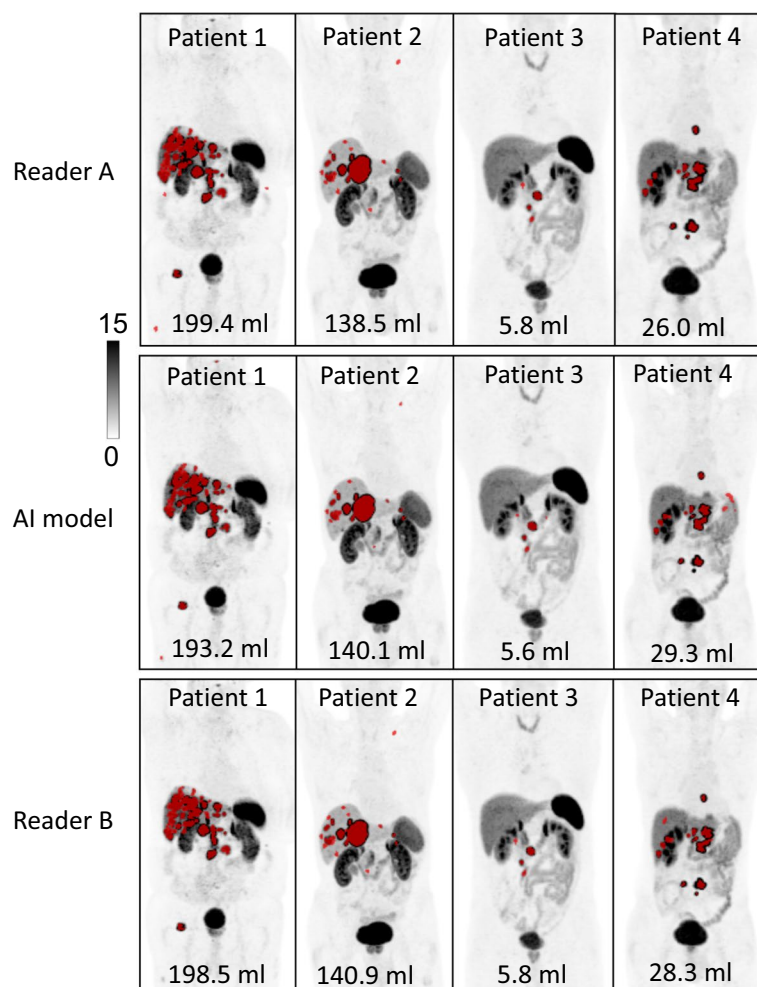


Fig. 6 Demonstration of 4 patients were the AI model segmented tumours similar as Reader A and Reader B. SRET/vwb are presented for each example

be of interest, which could have made the number of false positives lower. However, there is always a risk of leaving out too much information.

The majority of neuroendocrine tumours overexpresses the SSTR subtype 2 most, but the subtype receptor expression is associated with the type of tumour and differentiation (Reubi 2003). Both [^{68}Ga]Ga-DOTA-TOC and -TATE share the same imaging characteristics (Velikyán et al. 2014) although their binding affinity for the subtypes of SSTR is slightly different (Reubi et al. 2000). When comparing the two tracers, the differences have been found to be small (Poepfel et al. 2011). Including both tracers might comprise a larger imaging diversity when training the AI model and we believe it is more of a strength than a weakness that both radiopharmaceuticals were included.

Random sampling was used for randomisation of the patients into training, validation and test groups. This method was used to minimize the risk of introducing bias to the model. A potential weakness is that the clinical parameters in the different groups may not have been evenly distributed.

Another limiting factor of this study is that manual segmentation is very time-consuming, and a semi-automatic method was used for segmenting all tumours. This method is imperfect, and like all segmentation methods, it has weaknesses (Foster et al. 2014; Im et al. 2018). We used this semi-automatic method of delineating 50% of SUVmax because it was previously described and showed value in predicting PFS (Abdulrezzak et al. 2016; Toriihara et al. 2019). It was also a method that could be implemented in our software and was feasible compared to complete manual segmentations, especially considering the high number of metastases that occurred in some of the patients.

Detectability of tumour lesions in the liver is complicated by a high normal background uptake. Normal hepatic uptake is also influenced by other factors, such as treatment with long-acting somatostatin analogues (Gålne et al. 2019) and the tumour “sink effect” (Beauregard et al. 2012), which likely affect the detection rate of liver metastases, both for manual detection and with AI models. One difference in our study compared to other studies exploring tumour burden is that we have segmented suspected tumour lesions outside the liver with lower uptake than the liver (Thuillier et al. 2022; Wehrend et al. 2021; Carlsen et al. 2022). We included these lesions in the model because tumour burden with an uptake lower than the liver could be important in the clinical setting. Lesions with the lowest SUVmean have shown increased prognostic value compared to the highest SUVmax in [⁶⁴Cu]Cu-DOTA-TATE PET/CT (Carlsen et al. 2021). These lesions could be of a more aggressive nature with potential uptake of 2-[¹⁸F]fluoro-2-deoxy-D-glucose (2-[¹⁸F]FDG) if dual-imaging was performed (Chan et al. 2017) or could include necrosis after treatment with PRRT, thus showing treatment response. Segmentation of only lesions with higher uptake than the liver might provide a biased AI model, which would likely segment fewer false positive lesions, giving a higher PPV, potentially resulting in a model that would miss lesions of importance.

Inter-observer variability for SSTR PET/CT has shown consistent and good agreement, but still, both false positive and false negative interpretations of the overall scan result occur, as indicated in our results (Fendler et al. 2017). Some inter-observer variability is also expected because there is a difference in background information given to Reader A relative to the AI model and Reader B, who did not have access to the referral text or written clinical report.

A weakness of this study is the lack of external validation, as recently noted by Pantelis et al. in a review of AI and GEP-NEN (Pantelis et al. 2022). Although external validation was not in the scope of this study, it would have been a significant strength. For clinical validation of the method, we have planned another study to explore the value of the AI model measurements of SRET_{Vwb} and TLSRE_{wb} on [⁶⁸Ga]Ga-DOTA-TOC and -TATE PET/CT images concerning PFS and OS after treatment with PRRT.

We do not believe that AI will make radiologists or nuclear medicine specialists obsolete in the near future, but rather increase the value of PET/CT examinations by providing validated imaging biomarkers, such as tumour volume. A future AI method will still require the physician to visually inspect the segmented tumour volume and correct obviously inaccurate lesions. We believe that automated AI models, such as in the field of imaging, could help to improve patient care with better prognostication, prediction of treatment response and for evaluation of follow-up examinations.

Conclusion

To the best of our knowledge, this is the first fully automatic AI model to segment the whole-body somatostatin-expressing tumour volume, SRETVwb, and total lesion somatostatin receptor expression, TLSREwb, on [⁶⁸Ga]Ga-DOTA-TOC and -TATE PET/CT images. Our AI model showed a good correlation with ground truth with few false positive and false negative lesions per patient. Since segmenting the tumour burden is a time-intensive task, it is not feasible in clinical practice. A clinically validated AI model segmenting neuroendocrine tumour lesions could be of prognostic, as well as predictive value when selecting patients for treatment.

Abbreviations

2-[¹⁸ F]FDG	2-[¹⁸ F]fluoro-2-deoxy-D-glucose
AI	Artificial intelligence
CONSORT	Consolidated standards of reporting trials
CNN	Convolutional neural networks
CT	Computed tomography
SUVmax	Maximum standardized uptake value
SUVmean	Mean standardized uptake value
NEC	Neuroendocrine carcinoma
NEN	Neuroendocrine neoplasm
NET	Neuroendocrine tumour
PRRT	Peptide receptor radionuclide therapy
PPV	Positive predictive value
PET	Positron emission tomography
PFS	Progression-free survival
SSTR	Somatostatin receptor
SRETV	Somatostatin receptor-expressing tumour volume
SD	Standard deviation
TLSRE	Total lesion somatostatin receptor expression
VOI	Volume of interest
SRETVwb	Whole-body somatostatin receptor-expressing tumour volume
TLSREwb	Whole-body total lesion somatostatin receptor expression

Acknowledgements

Not applicable

Author contributions

AG, OE, AS, and ET contributed to conceptualization. AG, OE, KV, DM, and ET contributed to data curation. AG, OE and ET contributed to formal analysis. AG, OE, KV, DM, and ET contributed to the investigation. AG, OE, AS, and ET contributed to methodology. AG, OE, AS, and ET contributed to writing the original draft. AG, OE, KV, DM, and ET contributed to writing, reviewing, and editing the manuscript. AS, OE, and ET contributed to supervision. All authors read and approved the final manuscript.

Funding

Open access funding provided by Lund University. This study was funded through the generous contributions of the Knut and Alice Wallenberg Foundation, Region Skåne, and Lund University, as well as grants provided by the Swedish government under the ALF agreement, which is an agreement between the government and the country councils.

Availability of data and materials

The datasets generated during the current study are not publicly available due to ethical considerations. Access to individual data can only be obtained after ethical approval from the Swedish Ethical Review Authority. The AI model is available upon reasonable request.

Declarations

Ethics approval and consent to participate

The study was approved by the Swedish Ethical Review Authority (EPN LU 2016/417, 2018/753 and 2021-05734-02). All patients provided written informed consent. All procedures were performed following the Declaration of Helsinki.

Consent for publication

All patients provided consent for publication.

Competing interests

OE is a board member and stockholder of Eigenvision AB, which is a company working with research and development in automated image analysis, computer vision, and machine learning. The other authors declare no conflict of interest.

Received: 11 May 2023 Accepted: 28 June 2023

Published online: 07 August 2023

References

- Abdulrezzak U, Kurt YK, Kula M, Tutus A (2016) Combined imaging with ^{68}Ga -DOTA-TATE and ^{18}F -FDG PET/CT on the basis of volumetric parameters in neuroendocrine tumors. *Nucl Med Commun* 37(8):874
- Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O et al (2021) Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J Big Data* 8(1):53
- Beauregard JM, Hofman MS, Kong G, Hicks RJ (2012) The tumour sink effect on the biodistribution of ^{68}Ga -DOTA-octreotate: implications for peptide receptor radionuclide therapy. *Eur J Nucl Med Mol Imaging* 39(1):50–56
- Bozkurt MF, Virgolini I, Balogova S, Beheshti M, Rubello D, Decristoforo C et al (2017) Guideline for PET/CT imaging of neuroendocrine neoplasms with $(^{68}\text{Ga})\text{-DOTA-conjugated somatostatin receptor targeting peptides and } (^{18}\text{F})\text{-DOPA}$. *Eur J Nucl Med Mol Imaging* 44(9):1588–1601
- Campana D, Ambrosini V, Pezzilli R, Fanti S, Labate AM, Santini D et al (2010) Standardized uptake values of $(^{68}\text{Ga})\text{-DOTANOC PET}$: a promising prognostic tool in neuroendocrine tumors. *J Nucl Med off Publ Soc Nucl Med* 51(3):353–359
- Carlsen EA, Johnbeck CB, Loft M, Pfeifer A, Oturai P, Langer SW et al (2021) Semiautomatic tumor delineation for evaluation of $(^{64}\text{Cu})\text{-DOTATATE PET/CT}$ in patients with neuroendocrine neoplasms: prognostication based on lowest lesion uptake and total tumor volume. *J Nucl Med off Publ Soc Nucl Med* 62(11):1564–1570
- Carlsen EA, Lindholm K, Hindsholm A, Gæde M, Ladefoged CN, Loft M et al (2022) A convolutional neural network for total tumor segmentation in $[(^{64}\text{Cu})\text{Cu-DOTATATE PET/CT}]$ of patients with neuroendocrine neoplasms. *EJNMMI Res* 12(1):30
- Chan YH (2003) *Biostatistics 104: correlational analysis*. Singap Med J 44(12):614–619
- Chan DL, Pavlakis N, Schembri GP, Bernard EJ, Hsiao E, Hayes A et al (2017) Dual somatostatin receptor/FDG PET/CT imaging in metastatic neuroendocrine tumours: proposal for a novel grading scheme with prognostic significance. *Theranostics* 7(5):1149–1158
- Chen L, Jumai N, He Q, Liu M, Lin Y, Luo Y et al (2023) The role of quantitative tumor burden based on $[(^{68}\text{Ga})\text{Ga-DOTA-NOC PET/CT}]$ in well-differentiated neuroendocrine tumors: beyond prognosis. *Eur J Nucl Med Mol Imaging* 50(2):525–534
- Courville JG, YBaA (2016) *Deep learning*. MIT Press, Cambridge
- Dasari A, Shen C, Halperin D, Zhao B, Zhou S, Xu Y et al (2017) Trends in the incidence, prevalence, and survival outcomes in patients with neuroendocrine tumors in the United States. *JAMA Oncol* 3(10):1335–1342
- Ebbers SC, Heimgartner M, Barentsz MW, van Leeuwen RS, van Treijen MJC, Lam MMEG et al (2021) Gallium-68-somatostatin receptor PET/CT parameters as potential prognosticators for clinical time to progression after peptide receptor radionuclide therapy: a cohort study. *Eur J Hybrid Imaging* 5(1):22
- Fendler WP, Barrio M, Spick C, Allen-Auerbach M, Ambrosini V, Benz M et al (2017) ^{68}Ga -DOTATATE PET/CT interobserver agreement for neuroendocrine tumor assessment: results of a prospective study on 50 patients. *J Nucl Med off Publ Soc Nucl Med* 58(2):307–311
- Foster B, Bagci U, Mansoor A, Xu Z, Mollura DJ (2014) A review on segmentation of positron emission tomography images. *Comput Biol Med* 50:76–96
- Gålne A, Almquist H, Almquist M, Hindorf C, Ohlsson T, Nordenström E et al (2019) A prospective observational study to evaluate the effects of long-acting somatostatin analogs on ^{68}Ga -DOTATATE uptake in patients with neuroendocrine tumors. *J Nucl Med* 60(12):1717–1723
- Goodfellow I, Courville A (2016) *Deep learning*. MIT Press
- Hastie T, Tibshirani R, Friedman J (2001) *The elements of statistical learning*. Springer New York Inc, New York
- Im HJ, Bradshaw T, Solaiyappan M, Cho SY (2018) Current methods to define metabolic tumor volume in positron emission tomography: which one is better? *Nucl Med Mol Imaging* 52(1):5–15
- Kaderli RM, Spanjol M, Kollár A, Bütikofer L, Gloy V, Dumont RA et al (2019) Therapeutic options for neuroendocrine tumors: a systematic review and network meta-analysis. *JAMA Oncol* 5(4):480–489
- Kratochwil C, Stefanova M, Mavriopoulou E, Holland-Letz T, Dimitrakopoulou-Strauss A, Afshar-Oromieh A et al (2015) SUV of $[(^{68}\text{Ga})\text{DOTATOC-PET/CT}]$ predicts response probability of PRRT in neuroendocrine tumors. *Mol Imaging Biol* 17(3):313–318
- Mueller D, Klette I, Baum RP, Gottschaldt M, Schultz MK, Breeman WA (2012) Simplified NaCl based (^{68}Ga) concentration and labeling procedure for rapid synthesis of (^{68}Ga) radiopharmaceuticals in high radiochemical purity. *Bioconjug Chem* 23(8):1712–1717
- Pantelis AG, Panagopoulou PA, Lapatsanis DP (2022) Artificial intelligence and machine learning in the diagnosis and management of gastroenteropancreatic neuroendocrine neoplasms-A scoping review. *Diagnostics (basel)* 12(4):874
- Poeppel TD, Binse I, Petersenn S, Lahner H, Schott M, Antoch G et al (2011) ^{68}Ga -DOTATOC Versus ^{68}Ga -DOTATATE PET/CT in functional imaging of neuroendocrine tumors. *J Nucl Med* 52(12):1864–1870
- Reubi JC (2003) Peptide receptors as molecular targets for cancer diagnosis and therapy. *Endocr Rev* 24(4):389–427
- Reubi JC, Schar JC, Waser B, Wenger S, Heppeler A, Schmitt JS et al (2000) Affinity profiles for human somatostatin receptor subtypes SST1–SST5 of somatostatin radiotracers selected for scintigraphic and radiotherapeutic use. *Eur J Nucl Med* 27(3):273–282
- Thuillier P, Liberini V, Grimaldi S, Rampado O, Gallio E, Santi BD et al (2022) Prognostic value of whole-body PET volumetric parameters extracted from ^{68}Ga -DOTATOC PET/CT in well-differentiated neuroendocrine tumors. *J Nucl Med* 63(7):1014–1020

- Tirosh A, Papadakis GZ, Millo C, Sadowski SM, Herscovitch P, Pacak K et al (2017) Association between neuroendocrine tumors biomarkers and primary tumor site and disease type based on total (68)Ga-DOTATATE-Avid tumor volume measurements. *Eur J Endocrinol* 176(5):575–582
- Toriihara A, Baratto L, Nobashi T, Park S, Hatami N, Davidzon G et al (2019) Prognostic value of somatostatin receptor expressing tumor volume calculated from (68)Ga-DOTATATE PET/CT in patients with well-differentiated neuroendocrine tumors. *Eur J Nucl Med Mol Imaging* 46(11):2244–2251
- Trägårdh E, Borrelli P, Kaboteh R, Gillberg T, Ulén J, Enqvist O et al (2020) RECOMIA—a cloud-based platform for artificial intelligence research in nuclear medicine and radiology. *EJNMMI Phys* 7(1):51
- Trägårdh E, Enqvist O, Ulén J, Jögi J, Bitzén U, Hedeer F et al (2022) Freely available, fully automated AI-based analysis of primary tumour and metastases of prostate cancer in whole-body [(18)F]-PSMA-1007 PET-CT. *Diagnostics (basel)* 12(9):2101
- Velikyan I, Sundin A, Sorensen J, Lubberink M, Sandstrom M, Garske-Roman U et al (2014) Quantitative and qualitative intrapatient comparison of ⁶⁸Ga-DOTATOC and ⁶⁸Ga-DOTATATE: net uptake rate for accurate quantification. *J Nucl Med* 55(2):204–210
- Wehrend J, Silosky M, Xing F, Chin BB (2021) Automated liver lesion detection in ⁶⁸Ga DOTATATE PET/CT using a deep fully convolutional neural network. *EJNMMI Res* 11(1):98
- WHO Classification of Tumours Editorial Board (2019) Digestive system tumors. International Agency for Research on Cancer, Lyon
- Zhernosekov KP, Filosofov DV, Baum RP, Aschoff P, Bihl H, Razbash AA et al (2007) Processing of generator-produced ⁶⁸Ga for medical application. *J Nucl Med off Publ Soc Nucl Med* 48(10):1741–1748

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
