



# Deep learning-based quantification of PET/CT prostate gland uptake: association with overall survival

Eirini Polymeri<sup>1,2</sup> , May Sadik<sup>3</sup>, Reza Kaboteh<sup>3</sup>, Pablo Borrelli<sup>3</sup>, Olof Enqvist<sup>4</sup>, Johannes Ulén<sup>5</sup>, Mattias Ohlsson<sup>6</sup>, Elin Trägårdh<sup>7</sup> , Mads H. Poulsen<sup>8</sup>, Jane A. Simonsen<sup>9</sup>, Poul Flemming Hoiland-Carlson<sup>9</sup>, Åse A. Johansson<sup>1,2</sup> and Lars Edenbrandt<sup>3,10</sup>

<sup>1</sup>Department of Radiology, Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg, <sup>2</sup>Department of Radiology, Region Västra Götaland, Sahlgrenska University Hospital, <sup>3</sup>Department of Clinical Physiology, Region Västra Götaland, Sahlgrenska University Hospital, <sup>4</sup>Department of Electrical Engineering, Region Västra Götaland, Chalmers University of Technology, Gothenburg, <sup>5</sup>Eigenvision AB, Malmö, <sup>6</sup>School of Information Technology, Halmstad Embedded and Intelligent Systems Research (EIS), CAISR - Centre for Applied Intelligent Systems Research, Halmstad University, Halmstad, <sup>7</sup>Department of Translational Medicine, Institute of Clinical Sciences, Lund University, Malmö, Sweden, <sup>8</sup>Department of Urology, Odense University Hospital, <sup>9</sup>Department of Nuclear Medicine, Odense University Hospital, Odense, Denmark, and <sup>10</sup>Department of Molecular and Clinical Medicine, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

## Summary

### Correspondence

Eirini Polymeri, Department of Radiology, Institute of Clinical Sciences at Sahlgrenska Academy, Sahlgrenska University Hospital, Blå Stråket 5, SE-413 45 Gothenburg, Sweden.  
E-mail: eirini.polymeri@gu.se

### Accepted for publication

Received 21 May 2019;  
accepted 22 November 2019

### Key words

artificial intelligence; convolutional neural network; objective quantification; prostatic neoplasms

**Aim** To validate a deep-learning (DL) algorithm for automated quantification of prostate cancer on positron emission tomography/computed tomography (PET/CT) and explore the potential of PET/CT measurements as prognostic biomarkers. **Material and methods** Training of the DL-algorithm regarding prostate volume was performed on manually segmented CT images in 100 patients. Validation of the DL-algorithm was carried out in 45 patients with biopsy-proven hormone-naïve prostate cancer. The automated measurements of prostate volume were compared with manual measurements made independently by two observers. PET/CT measurements of tumour burden based on volume and SUV of abnormal voxels were calculated automatically. Voxels in the co-registered <sup>18</sup>F-choline PET images above a standardized uptake value (SUV) of 2.65, and corresponding to the prostate as defined by the automated segmentation in the CT images, were defined as abnormal. Validation of abnormal voxels was performed by manual segmentation of radiotracer uptake. Agreement between algorithm and observers regarding prostate volume was analysed by Sørensen-Dice index (SDI). Associations between automatically based PET/CT biomarkers and age, prostate-specific antigen (PSA), Gleason score as well as overall survival were evaluated by a univariate Cox regression model.

**Results** The SDI between the automated and the manual volume segmentations was 0.78 and 0.79, respectively. Automated PET/CT measures reflecting total lesion uptake and the relation between volume of abnormal voxels and total prostate volume were significantly associated with overall survival ( $P = 0.02$ ), whereas age, PSA, and Gleason score were not.

**Conclusion** Automated PET/CT biomarkers showed good agreement to manual measurements and were significantly associated with overall survival.

## Introduction

High sensitivity, functionality, and quantification are hallmarks of molecular imaging, including positron emission tomography/computer tomography (PET/CT). However, these abilities are not used to the extent they deserve in cancer. Prostate-specific antigen (PSA) is non-specific (Bell

et al., 2014), the Gleason score is subjective with moderate reproducibility (Sadik et al., 2008; Ozkan et al., 2016), and bone metastases are late occurring events (Hoiland-Carlson et al., 2018). Thus, precise characterization of prostate cancer patients at the time of diagnosis is currently suboptimal and reliable markers for individualized assessment are needed.

In recent years, deep-learning (DL) has become the method of choice for automated image analysis (Lakhani et al., 2018). Radiology studies have reported on application of deep learning in different organs (Lehman et al., 2018; Nam et al., 2018; Tao et al., 2019). Still, the prognostic role of that application remains undefined and further investigation is needed.

Since automated quantitative PET/CT assessments are lacking, imaging biomarkers are still not routinely used. Deep learning in prostatic malignancies has previously been applied in magnetic resonance imaging (MRI) (Wang et al., 2017; Reda et al., 2018), whereas maximum standardized uptake value (SUV<sub>max</sub>) by <sup>18</sup>F-FDG PET/CT appears to be of relevance in predicting overall survival in metastasizing prostate cancer (Jadvar et al., 2013).

Accordingly, the primary aim of this retrospective study is to evaluate a novel three-dimensional deep learning-based technique on PET/CT images for automated assessment of cancer in the prostate gland and its agreement with manual assessment. The secondary aim was to investigate the association between automatically based estimates of prostate cancer uptake with overall survival in a group of patients with relatively high risk.

## Methods

### Patients

#### Training group

A deep learning algorithm was trained to segment the prostate gland on CT scans using manual segmentation in 100 consecutive patients, who were selected from Sahlgrenska's University Hospital, Gothenburg. These patients had a mean age of 56 years (range 21–85) and had undergone PET/CT scan examination, between October 2008 and December 2010 for colorectal (n = 22), lung (n = 17), lymphoma (n = 15), head-and-neck (n = 14) or other known or suspected tumours (n = 32). All examinations were pseudonymized before the digital processing of their examinations. Thereafter, the algorithm was applied in a validation group.

#### Validation group

The patients of the validation group were derived from a prior study on staging of prostate cancer by PET/CT (Poulsen et al., 2014). Inclusion criteria for that study were as follows: biopsy-proven prostate cancer, whole-body bone scintigraphy showing  $\geq 1$  bone metastases, ability to undergo MRI and safely await androgen deprivation until all study scans were completed. Exclusion criteria were as follows: ongoing or prior androgen deprivation, pain or suspicion of medullar compression due to bone metastases. All patients gave oral and written informed consent. Out of 50 patients in the original study, five with missing PET/CT were not included, leaving 45 patients, with median age 73 years (range 53–92), for analysis. Patients were recruited between May 2009 and March 2012. At inclusion the median PSA was 83 ng/ml (range 4–5740), Gleason score was 2–6 in six patients, seven in 17, 8–10 in 21 patients and not available in one patient. Thirty-three patients in the validation group died during follow-up. Their median survival time was 2.5 years (range 0.3–7.6) compared with 7.8 years (range 6.1–8.8) in the 12 surviving patients. Clinical information and other data were collected from the local medical records up until 12 March 2018.

Ethical approval for both groups was obtained from the Ethics Committee at Gothenburg University (295-08) and the Regional Ethics Review Boards in Sweden (2016/103) and Denmark (3-3013-1692/1).

#### Imaging

Training data were obtained using an integrated PET/CT scanner (Siemens Biograph 64 Truepoint, Gothenburg, Sweden). A low-dose CT scan (64-slice helical, 120 kV, 30 mAs) was performed from the base of the skull to mid-thigh. The CT slice thickness was 5 mm. PET images were not used in the training group.

Subjects in the validation group were scanned with PET/CT (Discovery VCT, GE Healthcare, Odense, Denmark) approximately 60 min after administration of 4 MBq/kg of

**Table 1** Comparison between the automated and the manual PET/CT measurements (n = 43).

	PET/CT measurements [Mean (95% confidence interval)]		
	DL-based	Nuclear medicine physician	Radiologist
SUV <sub>max</sub> <sup>a</sup>	8.6 (7.3–10)	7.8 (6.8–8.8)	8.8 (7.5–10)
SUV <sub>mean</sub> <sup>b</sup>	4.1 (3.6–4.5)	4.4 (4–4.8)	3.9 (3.5–4.3)
VOLUME (ml) <sup>c</sup>	31 (24.7–37.4)	22 (14.8–28.6)	40 (31–48.2)
FRACTION (%) <sup>d</sup>	43 (36–49.5)	–	59 (47.2–71.4)
TLU <sup>e</sup>	138 (106.4–170.5)	110 (73–146)	168 (125.4–211)

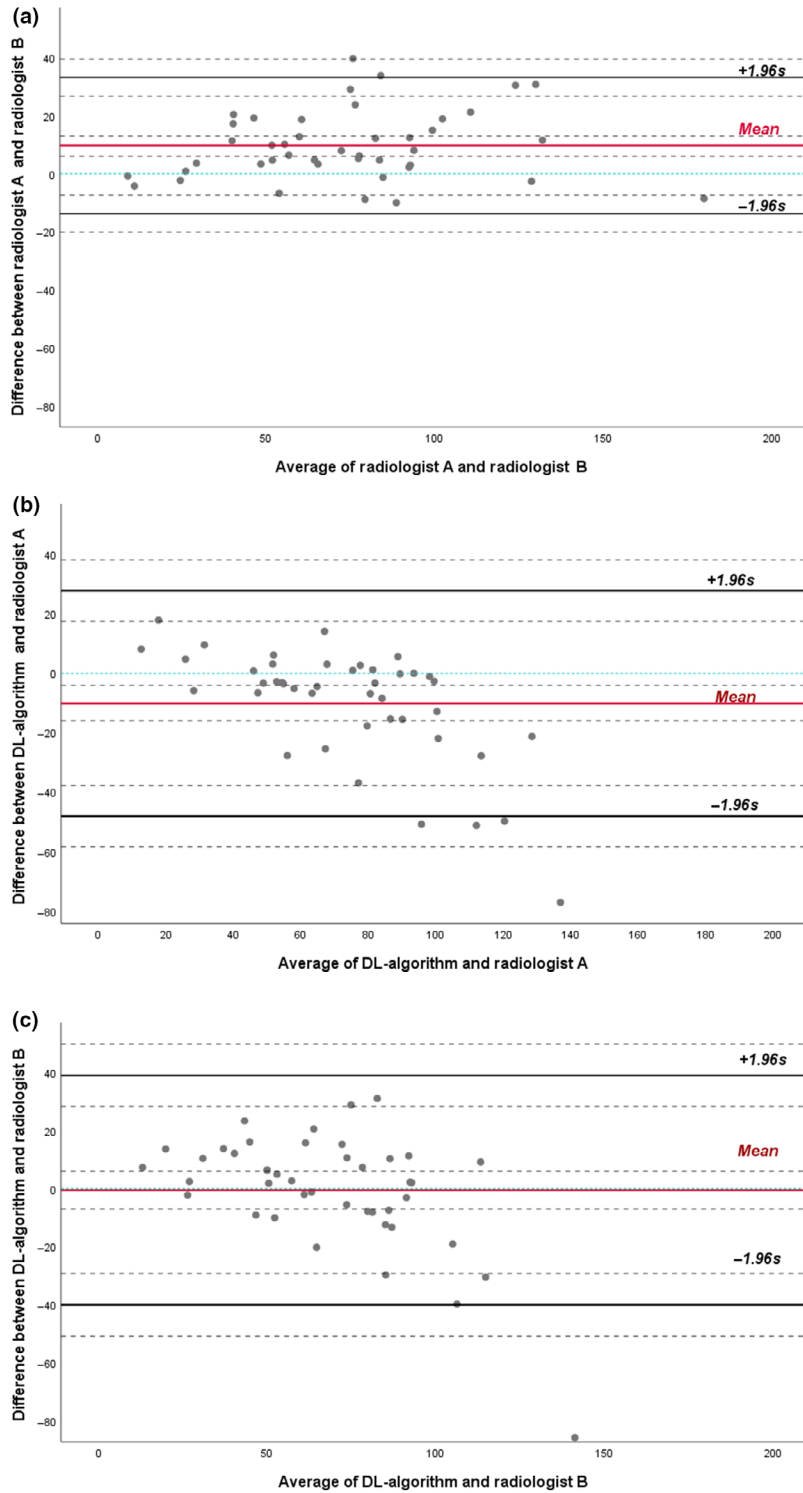
<sup>a</sup>Maximal SUV within the prostate gland.

<sup>b</sup>Average SUV for voxels with SUV > 2.65.

<sup>c</sup>Volume of prostate gland voxels with SUV > 2.65.

<sup>d</sup>Fraction of VOLUME related to the whole volume of the prostate gland.

<sup>e</sup>Product SUV<sub>mean</sub> × VOLUME reflecting the total lesion uptake.



**Figure 1** Bland–Altman plot illustrating the agreement of prostate gland volume measurements (ml) between Radiologist A and B (a), DL-algorithm and Radiologist A (b), as well as DL-algorithm and Radiologist B (c) in the validation group of 43 patients (data from Table 2). Representation of confidence interval limits for mean and agreement limits (black dotted lines) as well as the line of equality (blue dotted line).

<sup>18</sup>F-choline following a 6-h fast. A contrast-enhanced CT scan (64-slice helical, 120 kV, ‘smart mA’ maximum 400 mA) was obtained from base of the skull to mid-thigh with a slice thickness of 3.75 mm. The PET scan of the same region had an acquisition time of 2.5 min per bed position.

### Automated quantification method

The core of the automated segmentation method is a fully convolutional neural network (CNN) (Ian Goodfellow et al., 2016). For each voxel in the CT image, the network estimates the probability that this voxel belongs to the prostate gland.

**Table 2** Mean difference in prostate volume (ml) between the algorithm and the observers in the validation group of 43 patients.

Comparisons	Mean difference (95% CI <sup>d</sup> )	Upper LOA <sup>e</sup>	Lower LOA
DL <sup>a</sup> -Rad.A <sup>b</sup>	-10 (-16 to -4)	27.8	-48.2
DL-Rad.B <sup>c</sup>	-0.55 (-7 to 6)	39	-40
Rad.A-Rad.B	10 (6 to 13)	33.1	-13.8

<sup>a</sup>DL-deep learning algorithm.<sup>b</sup>Radiologist A.<sup>c</sup>Radiologist B.<sup>d</sup>95% confidence interval.<sup>e</sup>Limit of agreement.**Table 3** Sørensen-Dice index (SDI) showing the agreement between the automated and the manual segmentations of Radiologist A and Radiologist B (n = 43).

	Median	25th percentile	75th percentile
DL-based <sup>a</sup> vs Rad. A	0.83	0.76	0.84
DL-based vs Rad. B	0.80	0.74	0.84
Rad. A vs Rad. B	0.86	0.83	0.89

<sup>a</sup>DL-deep learning.

The voxel values of the CT scan were used as input to the CNN. The input was split into three separate pipes. Each pipe processes the image at a different scale, from coarse scale that is downsampled by a factor of four, to fine scale with no downsampling. Working on several scales allows for a large receptive field without excessive memory consumption, enabling efficient training of the model on a regular graphics card. Apart from the downsampling, the three pipes are identically designed with three dilated convolutional layers (Chen et al., 2018). Before merging the pipes, the two coarse-scale pipes are upsampled to the full spatial resolution. All convolutional layers use rectified linear units (Ian Goodfellow et al., 2016) as activation functions except the last one, which uses softmax to produce output probabilities that sum to one. The network contains roughly 1 million weights that have to be learnt from the training data. The goal of this process is to make the model output on the training set as similar as possible to the manual segmentations (Ian Goodfellow et al., 2016). As evaluating this loss function is extremely time-consuming, the minimization is performed using stochastic gradient descent; a random sample is selected and the gradient of the loss for this specific example is computed, then a small step in the negative direction of this gradient is taken.

### Postprocessing

The raw segmentation from the CNN is not always perfect. Sporadically, background voxels are classified as prostate gland. To address this problem only the largest connected component from the CNN prostate segmentation is kept. A similar problem

is that isolated voxels inside the prostate gland might be classified as background. This is addressed by performing morphological hole filling on the prostate gland mask.

### Training

The deep learning-based method was trained to segment the prostate gland using CT scans in which the prostate gland was segmented manually by one radiologist (Radiologist A) with six years of experience. Manual segmentation was performed using a cloud-based segmentation tool (RECOMIA, <https://www.recomia.org>) by the radiologist. The tool is voxel based allowing the observer to mark prostate gland voxels in the three-dimensional CT volume and in any of the three planes (transaxial, coronal or sagittal).

### Validation

After training, the segmentation method was applied to the validation group. The prostate glands were automatically segmented in the CT scans. Voxels in the co-registered PET images above an SUV of 2.65, as proposed by Reske et al. (2006) and corresponding to the prostate gland as defined in the CT images, were defined as abnormal. The following five automated PET-measures were calculated: maximal SUV within the prostate gland (SUVmax); average SUV for voxels with SUV > 2.65 (SUVmean); volume of voxels with SUV > 2.65 in ml (VOLUME); fraction of VOLUME related to the whole volume of the prostate gland (FRACTION); and product SUVmean × VOLUME reflecting the total lesion uptake (TLU).

### Validation methods

#### Validation of prostate volume

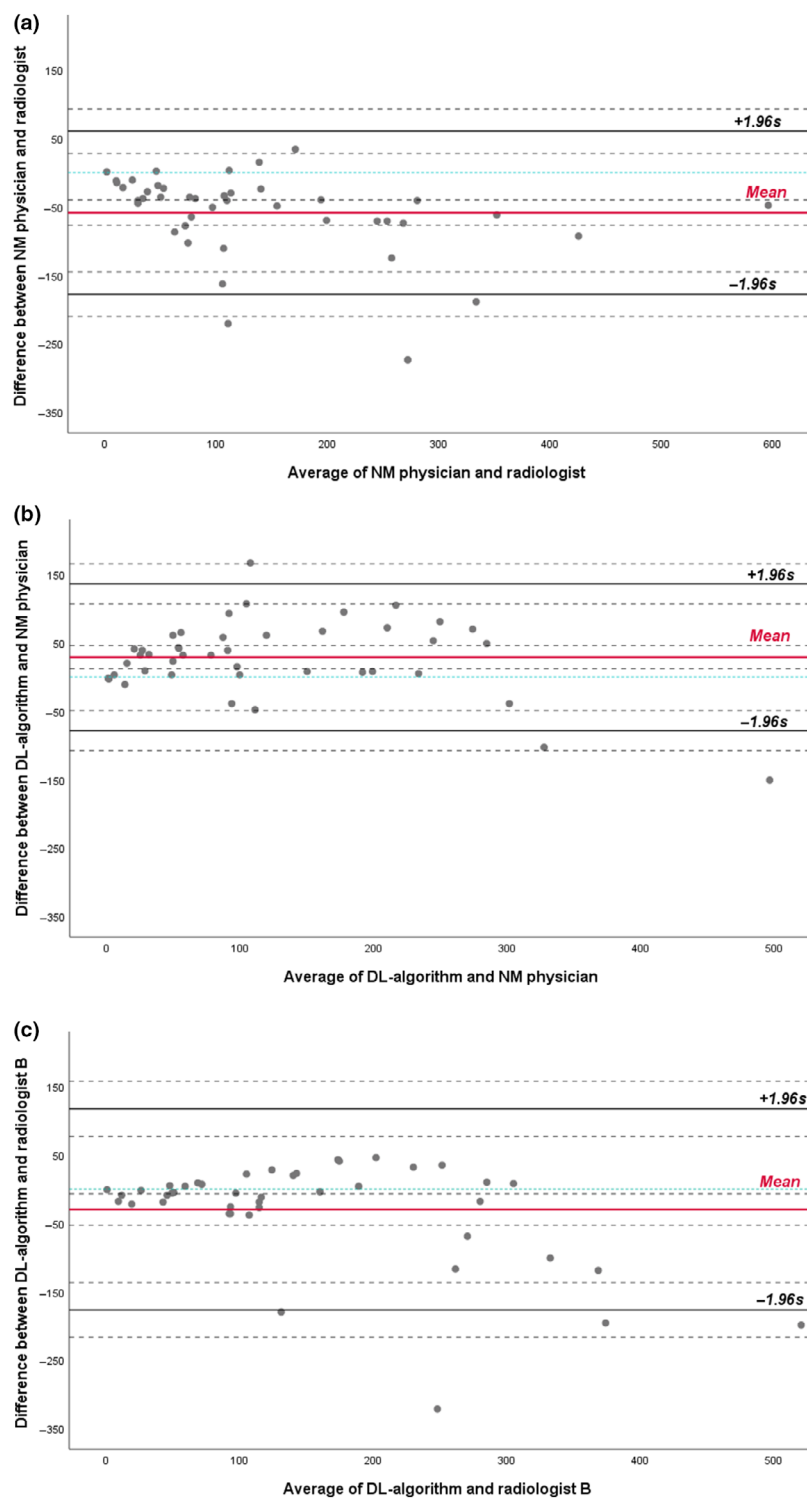
The prostate glands of the 45 patients in the validation group were automatically segmented in the CT scans by the algorithm. Further, two observers with a minimum of six years of experience performed individual manual segmentations of the prostates of the same material. Thereafter, the automated and manual segmentations were compared with each other.

#### Validation of PET uptake

An experienced radiologist and a nuclear medicine physician independently evaluated the abnormal PET uptake in the prostate gland of the validation group. The same cloud-based segmentation tool as in the training group (<https://www.recomia.org>) was used. Automated TLU measurements were compared with the manual evaluations of the physicians.

### Clinical validation

The prognostic value of automated quantification was assessed by studying the association between the PET/CT measurements



**Figure 2** Bland–Altman plot illustrating the agreement of total lesion uptake (TLU) measurements between the nuclear medicine (NM) physician and the Radiologist B (a), DL-algorithm and NM physician (b), as well as DL-algorithm and the Radiologist B (c) in the validation group of 43 patients (data from Table 4). Representation of confidence interval limits for mean and agreement limits (black dotted lines) as well as the line of equality (blue dotted line).

and the common clinical measures, that is age, Gleason score and PSA with overall survival in the validation group.

### Statistical analysis

The Sørensen-Dice index (SDI) was used to evaluate the agreement between automated and manual segmentations by

analysis of number of overlapping voxels. The SDI is defined as twice the number of overlapping voxels divided by the sum of the total amount of voxels classified as prostate in both segmentations. Coefficient values range between 0 and 1, with 1 reflecting perfect agreement (D, 1999). The agreement between the automated measurements and the observers as well as agreement between observers were further

evaluated using Bland–Altman analysis and linear regression analysis. Associations between automated PET/CT measurements, age, PSA, Gleason score and overall survival were investigated using a univariate Cox proportional hazards regression model. Overall survival was calculated from the date of PET/CT scan. Hazard ratios (HR) and 95% confidence intervals (CI) were estimated. The level of significance was set at 0.05. The R statistical computing environment (RC, 2014) as well as the SPSS statistics (version 25) were used.

## Results

### Prostate volume

The algorithm failed in two patients with hip prostheses causing beam hardening. Both examinations were excluded and results from the remaining 43 patients are shown in Table 1. The median prostate gland volume was 71 ml (range 17–118) with automated segmentation, corresponding to 65 ml (range 9–184) and 80 ml (range 9–176), with manual segmentation by the two radiologists (Fig. 1a–c). The mean difference regarding volume was 10 ml between the radiologists, and –10 ml and –0.55 ml between the algorithm and Radiologist A and B, respectively (Table 2).

The automatically derived volume was in between the two manually determined volumes in 14 of 43 (33%) patients. The overlap between automated and the two manual prostate gland segmentations showed SDIs of 0.78 and 0.79, respectively. The analysis of inter-observer agreement for the two radiologists showed an SDI of 0.84 (Table 3). Significant correlation was found between difference and mean of prostate volume of the automated and the manual volume measurements ( $P < 0.001$ ). The agreement between the algorithm and the observers was dependent on prostate volume, with the algorithm overestimating small glands and underestimating large glands.

### Lesion uptake

Median TLU was 105 (range 0–421) by the algorithm-based method. The two readers obtained a median TLU of 62 (range 0–572) and 126 (range 1–620), respectively (Fig. 2a–c). The mean difference regarding total lesion uptake was –59 between the two observers, corresponding to 29 and –30 between the algorithm and the nuclear medicine physician and radiologist, respectively (Table 4). The automated TLU values were in between the two manually obtained values in 19 of 43 (44%) patients. Regression analysis showed no significant correlation between difference and mean of the lesion uptake of the nuclear medicine physician and the algorithm ( $P = 0.07$ ), whereas the manual measurements of the Radiologist A were generally higher than their corresponding measurements made by the algorithm and the nuclear medicine physician.

### Analysis of survival

In the univariate Cox analysis, three of the automatically derived volumetric measurements (VOLUME, FRACTION and TLU), all measures by the nuclear medicine physician as well

**Table 4** Mean difference of total lesion uptake between the DL-based algorithm and the observers in the validation group of 43 patients.

Comparisons	Mean difference (95% CI) <sup>c</sup>	Upper LOA <sup>d</sup>	Lower LOA <sup>d</sup>
DL <sup>a</sup> - Radiologist B	–30 (–53 to –7)	117.5	–177
DL-NM <sup>b</sup> physician	29 (12 to 46)	136.4	–78.5
NM physician- Radiologist B	–59 (–77 to –40)	60.6	–178

<sup>a</sup>Deep learning algorithm.

<sup>b</sup>Nuclear medicine.

<sup>c</sup>95% confidence interval.

<sup>d</sup>Limit of agreement.

**Table 5** Univariate survival analysis demonstrating the association between PET/CT measurements, age, PSA, as well as Gleason score and overall survival ( $n = 43$ ).

Automatically based measures	Hazard ratio	95% CI	P-value
SUVmax <sup>b</sup>	1.03	0.95–1.11	0.51
SUVmean <sup>c</sup>	1.13	0.90–1.42	0.30
VOLUME (ml) <sup>d</sup>	1.02	1.003–1.036	0.02
FRACTION (%) <sup>e</sup>	1.02	1.003–1.038	0.02
TLU <sup>f</sup>	1.004	1.0005–1.0068	0.02
Nuclear medicine physician	Hazard ratio	95% CI	P-value
SUVmax <sup>b</sup>	1.13	1.016–1.263	0.03
SUVmean <sup>c</sup>	1.38	1.019–1.862	0.04
VOLUME (ml) <sup>d</sup>	1.01	1.001–1.024	0.03
TLU <sup>f</sup>	1.002	1.0002–1.005	0.03
Radiologist	Hazard ratio	95% CI	P-value
SUVmax <sup>b</sup>	1.04	0.973–1.122	0.23
SUVmean <sup>c</sup>	1.08	0.875–1.345	0.46
VOLUME (ml) <sup>d</sup>	1.01	1.003–1.024	0.01
FRACTION (%) <sup>e</sup>	1.0004	0.993–1.008	0.91
TLU <sup>f</sup>	1.002	1.0004–1.0046	0.02
Clinical data	Hazard ratio	95% CI	P-value
Age	1.04	0.99–1.09	0.06
PSA (log) <sup>a</sup>	1.39	0.87–2.23	0.17
Gleason score	1.15	0.88–1.52	0.30

<sup>a</sup>Prostate-specific antigen (ng/ml) logarithmic.

<sup>b</sup>Maximal SUV within the prostate gland.

<sup>c</sup>Average SUV for voxels with SUV > 2.65.

<sup>d</sup>Volume of prostate gland voxels with SUV > 2.65.

<sup>e</sup>Fraction of VOLUME related to the whole volume of the prostate gland.

<sup>f</sup>Product SUVmean × VOLUME reflecting the total lesion uptake.



as VOLUME and TLU by the radiologist were significantly associated with overall survival. The other PET/CT measurements, age, PSA (logarithmic) and Gleason score, were not. The results are shown in Table 5, where hazard ratio accounts for one unit change of each of the automated variables. Figure 3 shows the PET/CT scans from two study patients and their respective automatically based measurements.

## Discussion

Automatically based estimates of  $^{18}\text{F}$ -choline uptake in the prostate gland reflecting lesion volume and total lesion uptake were significantly associated with overall survival contrary to age, PSA and Gleason score. They were similar to values obtained manually by radiologists, but much faster. Moreover, they provide measures of cancer reflecting metabolically active cells. Thus, these measurements may be clinically more relevant than SUVmax or SUVpeak representing only minute volumes of cancerous tissue with high activity, whereas the entire volume of cancer and its overall activity is a conceptually more truthful measure of the cancerous burden (Basu et al., 2014; Boellaard et al., 2015; Ziai et al., 2016).

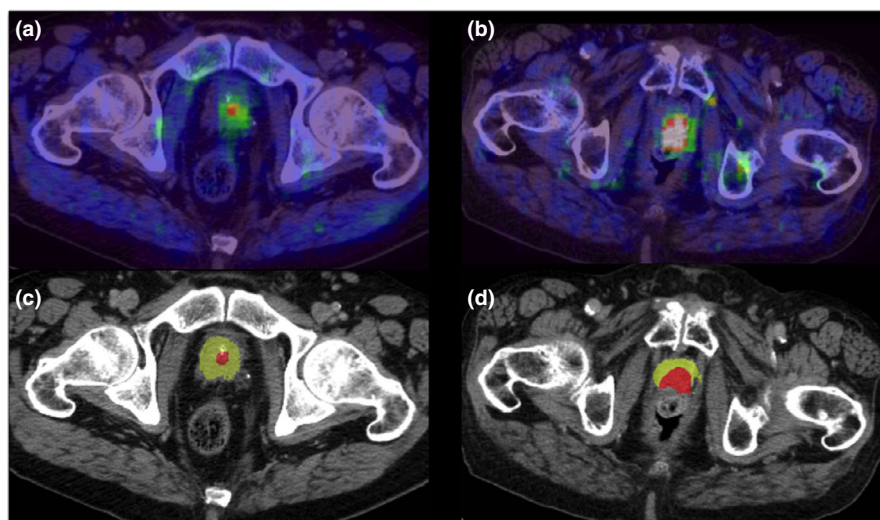
Volumetric analysis by manual processing is not standardized and is a time-consuming clinical application. The algorithm provides automated and clinically relevant measures within seconds that may substitute or significantly improve today's mainly visual assessment of CT, MRI, PET/CT and PET/MRI. Thus, it may significantly increase the diagnostic accuracy and precision of PET/CT imaging. In prostate cancer, it may become as important as PSA and Gleason score for identifying high- and low-risk patients. Our approach was trained with scans from one institution and validated by scans from another, a circumstance indicating that the algorithm is

applicable in other settings than where it was trained. Thus, it could be speculated that if properly trained, the algorithm could work with other tracers and other scanners.

## Limitations

The algorithm overestimated the volume of small glands and underestimated large glands. This could at least partly be explained by the selection of cases to the training group. Training of automated-networks depends on large data sets as well as on different anatomical conditions in each patient. Larger training groups with both normal and pathological prostate glands are likely to improve the performance of the algorithm. A fixed threshold (SUV > 2.65) to define abnormal uptake of  $^{18}\text{F}$ -choline was used. Other fixed thresholds and other ways of defining the most relevant borders of abnormal uptake should be considered in future studies (Schaefferkoetter et al., 2017).

The value of  $^{18}\text{F}$ -choline PET is still debatable. Some studies highlight its application in case of biochemical recurrence at high PSA levels (>2 ng/ml) and in monitoring of therapy response (Ceci et al., 2016; Evangelista et al., 2016). Other recent studies have analysed the higher accuracy of  $^{68}\text{Ga}$ -PSMA in detecting prostate cancer (Morigi et al., 2015; Evangelista et al., 2016; Eapen et al., 2018). Hence, the clinical application of the latter has increased over the past few years. Yet, a recent study showed a 20.5% positive detection rate of  $^{11}\text{C}$ -choline PET/CT in prostate cancer patients with biochemical failure after radical prostatectomy and PSA < 1 ng/ml (Giovacchini et al., 2019). More specific prostate cancer tracers targeting primarily prostate-specific membrane antigen and with higher positive detection rate contends for precedence in future studies. However, until such tracers become more broadly available, radiolabeled choline could still be a valuable



**Figure 3**  $^{18}\text{F}$ -choline PET/CT scans of two study patients, one aged 68 with survival time 3 years and 8 months (a, b) and another aged 72 with survival time 1 year and 3 months (c, d). Upper panels (a, c) show fused PET and CT images, lower panels (b, d) are CT images with automated segmentation of the prostate gland (yellow) and demonstration (red) of prostate gland cancer obtained from the corresponding PET scan using SUV > 2.65 as cut-off. The longer living patient (a, b) had higher PSA (1230 vs 102) and Gleason score (4 + 3 vs 3 + 4), but lower VOLUME (2 vs 38 ml), FRACTION (7% vs 76%); and TLU (7 vs 236) than the shorter living patient (c, d).

option. Further, the extensive clinical use of radiolabeled choline PET for many years has the advantage of a long-term follow-up of prostate cancer patients. Although  $^{18}\text{F}$ -choline is not an ideal prostate cancer tracer (Evangelista et al., 2016), it was used in this study because of the fund of validated scans available.

Further, the presence of skeletal metastases makes a worse prognosis itself. Yet, the main purpose of this study was to demonstrate the feasibility of the algorithm and investigate its association with overall survival in this patient category, aware of the association between cancer spread and survival.

In conclusion, automated, volume-based measures of  $^{18}\text{F}$ -choline uptake in the prostate gland could be obtained by the algorithm and showed good agreement with the manually assessed measurements. In addition, the automated PET/CT

biomarkers were significantly associated with overall survival despite suboptimal tracer characteristics. The algorithm is now open for investigation in other cancers and with other tracers.

## Acknowledgments

The study was funded by the Gothenburg University (Medical Faculty ALF Grants) Sweden and from EXINI Diagnostics AB, Lund Sweden. The funding bodies had no influence on the design of the study, data collection, data analysis or in writing of the manuscript.

## Conflict of interest

Lars Edenbrandt is employed as Scientific Director by EXINI Diagnostics AB (Lund, Sweden).

## References

- Basu S, Zaidi H, Salavati A, et al. FDG PET/CT methodology for evaluation of treatment response in lymphoma: from “graded visual analysis” and “semiquantitative SUVmax” to global disease burden assessment. *Eur J Nucl Med Mol Imaging* (2014); **41**: 2158–2160.
- Bell N, Connor Gorber S, Shane A, et al. Recommendations on screening for prostate cancer with the prostate-specific antigen test. *CMAJ* (2014); **186**: 1225–1234.
- Boellaard R, Delgado-Bolton R, Oyen WJ, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. *Eur J Nucl Med Mol Imaging* (2015); **42**: 328–354.
- Ceci F, Castellucci P, Mapelli P, et al. Evaluation of prostate cancer with  $^{11}\text{C}$ -choline PET/CT for treatment planning, response assessment, and prognosis. *J Nucl Med* (2016); **57**: 49s–54s.
- Chen LC, Papandreou G, Kokkinos I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Trans Pattern Anal Mach Intell* (2018); **40**: 834–848.
- D. G. (1999), vol. **2018**. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.9.1334&rep=rep1&type=pdf>
- Eapen RS, Nzenza TC, Murphy DG, et al. PSMA PET applications in the prostate cancer journey: from diagnosis to theranostics. *World J Urol* (2018); **37**(7): 1255–1261.
- Evangelista L, Briganti A, Fanti S, et al. New clinical indications for  $(^{18}\text{F}/(^{11}\text{C})\text{C}$ -choline, new tracers for positron emission tomography and a promising hybrid device for prostate cancer staging: a systematic review of the literature. *Eur Urol* (2016); **70**: 161–175.
- Giovacchini G, Guglielmo P, Mapelli P, et al.  $(^{11}\text{C})\text{C}$ -choline PET/CT predicts survival in prostate cancer patients with PSA < 1 NG/ml. *Eur J Nucl Med Mol Imaging* (2019); **46**: 921–929.
- Goodfellow I, Bengio Y, Courville A. *Deep Learning* (2016). MIT Press.
- Hoiland-Carlsen PF, Hess S, Werner TJ, et al. Cancer metastasizes to the bone marrow and not to the bone: time for a paradigm shift! *Eur J Nucl Med Mol Imaging* (2018); **45**: 893–897.
- Jadvar H, Desai B, Ji L, et al. Baseline  $^{18}\text{F}$ -FDG PET/CT parameters as imaging biomarkers of overall survival in castrate-resistant metastatic prostate cancer. *J Nucl Med* (2013); **54**: 1195–1201.
- Lakhani P, Prater AB, Hutson RK, et al. Machine learning in radiology: applications beyond image interpretation. *J Am College Radiol: JACR* (2018); **15**: 350–359.
- Lehman CD, Yala A, Schuster T, et al. Mammographic breast density assessment using deep learning: clinical implementation. *Radiology* (2018); **290**(1): 52–58.
- Morigi JJ, Stricker PD, van Leeuwen PJ, et al. Prospective comparison of  $^{18}\text{F}$ -fluoromethylcholine versus  $^{68}\text{Ga}$ -PSMA PET/CT in prostate cancer patients who have rising PSA after curative treatment and are being considered for targeted therapy. *J Nucl Med* (2015); **56**: 1185–1190.
- Nam JG, Park S, Hwang EJ, et al. Development and validation of deep learning-based automatic detection algorithm for malignant pulmonary nodules on chest radiographs. *Radiology* (2018); **290**(1): 218–228.
- Ozkan TA, Eruyar AT, Cebeci OO, et al. Inter-observer variability in Gleason histological grading of prostate cancer. *Scand J Urol* (2016); **50**: 420–424.
- Poulsen MH, Petersen H, Hoiland-Carlsen PF, et al. Spine metastases in prostate cancer: comparison of technetium- $^{99\text{m}}$ -MDP whole-body bone scintigraphy,  $[(^{18}\text{F})\text{C}$ -choline positron emission tomography (PET)/computed tomography (CT) and  $[(^{18}\text{F})\text{NaF}$  PET/CT. *BJU Int* (2014); **114**: 818–823.
- RC T. R. *A Language and Environment for Statistical Computing* (2014).
- Reda I, Khalil A, Elmogy M, et al. Deep learning role in early diagnosis of prostate cancer. *Technol Cancer Res Treat* (2018); **17**: 1533034618775530.
- Reske SN, Blumstein NM, Neumaier B, et al. Imaging prostate cancer with  $^{11}\text{C}$ -choline PET/CT. *J Nucl Med* (2006); **47**: 1249–1254.
- Sadik M, Suurkula M, Hoglund P, et al. Quality of planar whole-body bone scan interpretations—a nationwide survey. *Eur J Nucl Med Mol Imaging* (2008); **35**: 1464–1472.
- Schaefferkoetter JD, Wang Z, Stephenson MC, et al. Quantitative  $(^{18}\text{F})\text{F}$ -fluorocholine positron emission tomography for prostate cancer: correlation between kinetic parameters and Gleason scoring. *EJNMMI Res* (2017); **7**: 25.
- Tao Q, Yan W, Wang Y, et al. Deep learning-based method for fully automatic quantification of left ventricle function from cine MR images: a multivendor, multicenter study. *Radiology* (2019); **290**(1): 81–88.
- Wang X, Yang W, Weinreb J, et al. Searching for prostate cancer by fully automated magnetic resonance imaging classification: deep learning versus non-deep learning. *Sci Rep* (2017); **7**: 15415.
- Ziai P, Hayeri MR, Salei A, et al. Role of optimal quantification of FDG PET imaging in the clinical practice of radiology. *Radiographics* (2016); **36**: 481–496.