

# Segmentation of Lung Tumours in Positron Emission Tomography Scans: a Machine Learning Approach

Aliaksei Kerhet<sup>1</sup>, Cormac Small<sup>2</sup>, Harvey Quon<sup>2</sup>, Terence Riauka<sup>3</sup>,  
Russell Greiner<sup>4</sup>, Alexander McEwan<sup>1</sup>, and Wilson Roa<sup>2</sup>

<sup>1</sup> Department of Oncology, University of Alberta,  
11560 University Avenue, T6G 1Z2 Edmonton AB, Canada,  
{kerhet,mcewan}@ualberta.ca

<sup>2</sup> Department of Radiation Oncology, Cross Cancer Institute,  
11560 University Avenue, T6G 1Z2 Edmonton AB, Canada,  
{cormacsm,harveyqu,wilsonro}@cancerboard.ab.ca

<sup>3</sup> Department of Medical Physics, Cross Cancer Institute,  
11560 University Avenue, T6G 1Z2 Edmonton AB, Canada  
terencer@cancerboard.ab.ca

<sup>4</sup> Alberta Ingenuity Centre for Machine Learning,  
Department of Computing Science, University of Alberta,  
3-59 Athabasca Hall, T6G 2E8 Edmonton AB, Canada  
greiner@cs.ualberta.ca

**Abstract.** Lung cancer represents the most deadly type of malignancy. In this work we propose a machine learning approach to segmenting lung tumours in Positron Emission Tomography (PET) scans in order to provide a radiation therapist with a “second reader” opinion about the tumour location. For each PET slice, our system extracts a set of attributes, passes them to a trained Support Vector Machine (SVM), and returns the optimal threshold value for distinguishing tumour from healthy voxels in that particular slice. We use this technique to analyse four different PET/CT 3D studies. The system produced fairly accurate segmentation, with Jaccard and Dice’s similarity coefficients between 0.82 and 0.98 (the areas outlined by the returned thresholds vs. the ones outlined by the reference thresholds). Besides the high level of geometric similarity, a significant correlation between the returned and the reference thresholds also indicates that during the training phase, the learning algorithm effectively acquired the dependency between the extracted attributes and optimal thresholds.

**Key words:** Support Vector Machine (SVM), Positron Emission Tomography (PET), Radiation Treatment, Lung Cancer, Gross Tumour Volume (GTV).

## 1 Introduction

According to the Canadian Cancer Society reports [1, 2], lung cancer represents the second most common type of cancer (approximately 166,400 new cases were

expected in 2008 in Canada alone), and the one most fatal to both men and women (it is responsible for more than 1/4 of all cancer-associated deaths). Even for younger adults (aged 20-44), it is ranked first (men) and second (women) with respect to the potential number of years of life lost.

Radiation therapy involves applying beams of ionizing radiation to irradiate the tumour volume. Present-day equipment allows these beams to be directed very accurately. However, this is only effective if one can define (segment) a tumour with a similar accuracy. Unfortunately, this is not currently the case.

The conventional way to define lung tumours is based on the analysis of *computed tomography* (CT) images. The sensitivity and specificity of this imaging modality is not always high enough, which leads to significant levels of the intra- and inter-observer variability. Introduction of the *positron emission tomography* (PET) imaging modality to the process of lung tumour definition has already been shown to alter the results and decrease the variability above. However, due to some challenges related to the analysis of PET scans, the role of this modality in radiation treatment planning has not yet been well established.

This paper uses a machine learning approach to address some of these challenges. For each PET slice, our system extracts a set of attributes, passes them to a trained Support Vector Machine (SVM), returns the optimal threshold value, and applies it for distinguishing tumour from healthy voxels in that particular slice. This automatically provides a radiation therapist with a “second reader” opinion about the tumour location.

The remaining sections of this paper are organised as follows. The next section reviews the state-of-the-art for tumour segmentation in PET scans. Section 3 describes the proposed approach and provides a brief introduction into SVM for regression estimation. The experimental part and results are described in Sect. 4, followed by Sect. 5, dedicated to the discussion and conclusion.

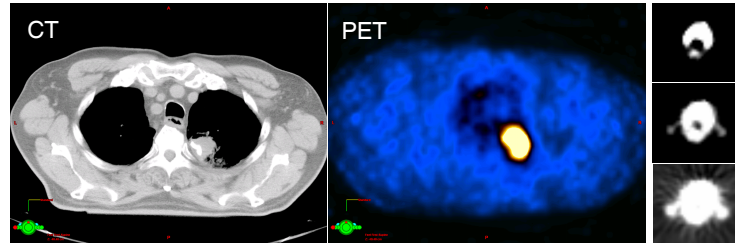
## 2 PET in Radiation Therapy: Background

Unlike CT imaging, which provides an anatomical description of a scanned body, PET imaging visualises the functionality of the body cells. Prior to a PET imaging study, the patient is injected with a radioactively marked substance, which is absorbed and metabolised differently by different types of cells. The radioactivity emitted from each region of the body is then registered, and the reconstructed images visualise quantities of the substance uptake, measured in counts of radioactive decays or some other uptake values. As cancerous cells are known to absorb more sugar than surrounding healthy tissue for many organs, most PET studies use a radioactively labelled analogue of sugar called *fluorodeoxyglucose* to visualise tumours. Today PET is primarily used in diagnostics as an indispensable technique for characterizing neoplasms and detecting distant metastases.

Besides diagnostics applications, adding PET to radiation treatment planning is also considered beneficial, as compared to using CT alone. As these two modalities are built on completely different underlying phenomena, they supply a radiation oncologist with two different and complementary perspectives on the

problem of tumour segmentation. In a recent review on lung cancer, authors of all 18 different studies (involving 661 patients) agreed that PET adds essential information, affecting the results of tumour segmentation [3].

However, the role of PET in radiation treatment planning is not well established, mainly due to the following challenges. First, PET images lack the sharpness and clearness of CT scans (Fig. 1). Second, sizes of objects in PET images strongly depend on the visualisation setup (contrast and brightness value, known as window/level setup in medical imaging; see Fig. 1, three rightmost images), and the optimal setup can vary across patients, and even across slices within a patient. Finally, using both PET and CT scans for tumour definition implies that the two scans must be co-registered, which is challenging due to PET’s blurriness.



**Fig. 1.** Corresponding CT and PET slices of a human thorax. While PET image is much more blurred than CT, the tumor area is much more evident in PET than in CT. Right: a slice of a mouse PET scan displayed using three different visualisation setups, which lead to different sizes of objects.

In the most basic case, an experienced nuclear medicine physician and/or radiation oncologist will visually interpret both PET and CT images to determine the tumour borders. Several automatic segmentation techniques have been proposed to make the interpretation of PET scans observer-independent and to cope with the challenge of choosing the right visualisation setup. They can be broadly divided into two groups. The first broad group aims to segment the tumour by searching for some inhomogeneity throughout the PET scan. Although there are some interesting examples from this group, such as gradient-based (watershed) methods [4, 5] and a multimodal generalisation of level set method [6], they are not as well established nor as frequently cited in current reviews as the methods from the second group, which aim to define the optimal threshold value of the uptake in order to segment a tumour. This second group includes approaches that define the optimal threshold as some fixed uptake value, or a fixed percentage of the maximum uptake value; other more sophisticated approaches determine the optimal threshold as the weighted sum of mean target uptake and mean background uptake, among other techniques [7–12]. Note that methods from the second group define a *single* optimal threshold value for the whole PET 3-D scan.

### 3 Proposed Approach

Our approach falls into the second group above: our system defines the optimal threshold value of the uptake, and declares each voxel to be cancerous if its uptake value is above that threshold. However, given the complexity of segmenting lung tumours in PET scans, we consider the optimal threshold as a *non-linear* function of *more* than just one or two attributes. We therefore extract a richer set of attributes from PET scans, and use a machine learning algorithm, capable of incorporating more complex dependencies based on these attributes, to find the (potentially different) optimal threshold value for *each* slice in a PET scan.

#### 3.1 PET Attributes

Several works that compared and reviewed different threshold-based lung tumour delineation algorithms suggested the use of contrast-oriented algorithms [3, 8]. Nestle *et al.* [8] defines the optimal threshold value *for the whole* PET 3-D scan as  $U_{bg} + 0.15 \times U_{70}$ , where the scalar  $U_{bg}$  is the mean uptake of a background (some homogeneous 3-D area near a tumour, e.g. mediastinum), and the scalar  $U_{70}$  is the mean uptake inside the 3-D isocontour of 70% of the maximum uptake. However, other studies suggest that even within the same 3-D scan, the optimal threshold value can vary from slice to slice with the tumour volume/cross-sectional area [11]. Our own observations also support this claim (Sect. 4.2 and Fig. 3). In line with the two considerations above, we aim to define the optimal threshold value *for each PET slice individually*, based on  $U_{bg}$  and the following 6 scalar attributes extracted from the given PET slice:

- the area and mean uptake inside the  $0.10 \times U_{70}$  contour
- the area and mean uptake inside the  $0.15 \times U_{70}$  contour
- the area and mean uptake inside the  $0.20 \times U_{70}$  contour.

#### 3.2 SVM Regressor

Support Vector Machine (SVM) [13, 14] is a very successful machine learning algorithm which has been used effectively for a wide variety of tasks, ranging from optical character recognition and electricity load prediction to biomedicine and face detection/recognition. In this work we use SVM for *regression estimation*.

During the training phase, a *training set* (available examples of the values of the seven attributes above (vector  $\mathbf{x}$ ) and the corresponding optimal threshold values (scalar  $y$ ) is analysed by the SVM. As a result of this analysis, a subset of the most important, characteristic examples (called *support vectors*) is identified and used to build the regression function, which is then used to predict optimal thresholds for new attribute vectors. In the most basic case, this is a linear function of the attributes. However, such a linear regressor is not always “expressive” enough to reflect the complexity of real-world problems. This obstacle is overcome using the so-called *kernel trick* [14], which consists in implicit mapping the vector of attributes  $\mathbf{x}$  onto a higher-dimensional space:  $\Phi : \mathbf{x} \rightarrow \Phi(\mathbf{x})$ , where

the data is more likely to allow linear approximation. Thanks to kernels, all calculations actually occur in the lower-dimensional space where vectors  $\mathbf{x}$  live, which is extremely beneficial from the computational point of view. Explicitly this leads to the following form of the regression function:

$$y(\mathbf{x}) = \sum_{i=1}^{N_{sv}} a_i K(\mathbf{x}_i, \mathbf{x}) + b, \quad (1)$$

where  $N_{sv}$  is the number of support vectors  $\mathbf{x}_i$ , and  $K(\mathbf{u}, \mathbf{v}) = \Phi(\mathbf{u}) \cdot \Phi(\mathbf{v})$  is a kernel function that implicitly calculates dot product in the higher-dimensional space where vectors  $\Phi(\mathbf{x})$  live. One of the most used and studied kernels is the so-called Gaussian kernel:  $K(\mathbf{u}, \mathbf{v}) = e^{-\gamma \|\mathbf{u} - \mathbf{v}\|^2}$ . For linear SVMs, the corresponding kernel is simply a dot product in the space where samples live:  $K(\mathbf{u}, \mathbf{v}) = \mathbf{u} \cdot \mathbf{v}$ . For a brief introduction into SVM regression, the reader is referred to [15].

A comprehensive scheme of our system is shown in Fig. 2.

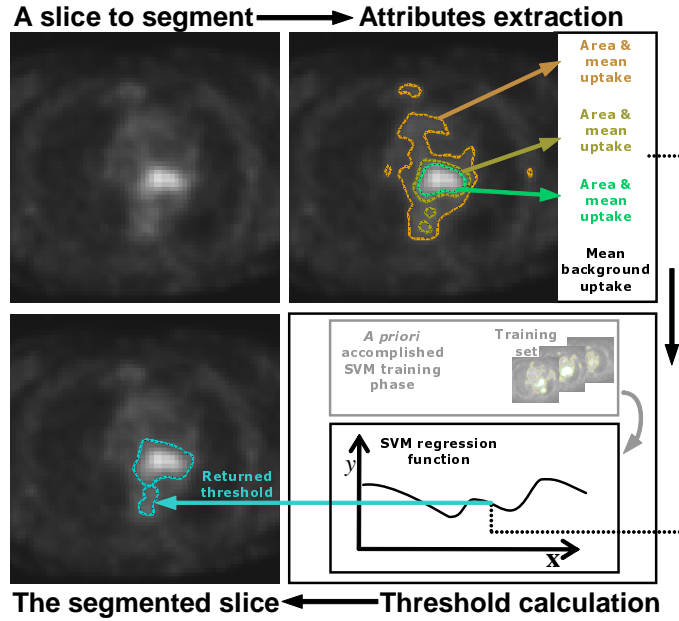


Fig. 2. A comprehensive scheme of the proposed approach.

## 4 Experiments and Results

### 4.1 Initial Data

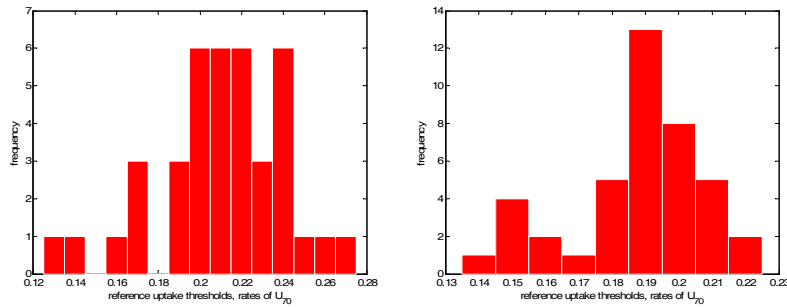
In this work we analyzed the data of two patients. Each patient underwent two same-day studies (free-breathing one and gated one [16]); and each study

comprised three scans: a diagnostic CT and a fluorodeoxyglucose PET obtained with a hybrid PET/CT scanner (Philips Gemini); and a separate treatment planning CT scan. Using a hybrid scanner ensured that the corresponding PET and CT scans were automatically and perfectly co-registered, eliminating the PET-CT co-registration challenge described in Sect. 2. This is important since it allows us to attribute the obtained results solely to the attributes extraction and the algorithm used, rather than to unwanted artifacts of the PET-CT co-registration.

For each study, a treatment planning CT scan was then manually co-registered with a diagnostic CT scan (a CT-CT co-registration is not as challenging as a PET-CT one), thus spatially linking all three scans. Two experienced radiation oncologists then used this rich and high-quality information in order to manually draw a tumour volume for the radiation treatment planning. This volume is referred to as a *gross tumour volume* (GTV). Using GTVs produced by the consensus of two radiation oncologists based on co-registered PET and CT data ensures that the obtained GTVs are of high quality.

## 4.2 Data Sets Generation

Each study of each patient was analysed separately and independently (i.e. we used a *study-specific* scenario, where both training and test sets are obtained from the same scan). Therefore,  $U_{bg}$ , which characterizes the whole scan rather than a specific slice (see Sect. 3.1), vanished from the consideration as a constant. Tumour-containing PET slices and 8 adjacent tumour-free slices were extracted. For each of these PET slices our system computed the values for 6 attributes described in Sect. 3.1. Also, for each of these slices a *reference uptake threshold* was assigned by the consensus of two radiation oncologists as the threshold that produced the segmentation most closely approximating the corresponding GTV contour. For each tumour-free slice the maximum uptake of that slice was used as the reference uptake threshold. The slice-to-slice variation of reference uptake thresholds for patient 2 is shown in Fig. 3.



**Fig. 3.** Slice-to-slice variation of reference uptake thresholds, as demonstrated by their histograms for free-breathing (left) and gated (right) studies of patient 2.

PET slices were randomly split in two groups in order to form a training set (75% of slices) and a test set (the remaining 25% of slices). This random splitting was repeated 5 times (resampling), resulting in 5 different pairs of training and test sets for the same study. The characteristics of data sets for different patients/studies are summarized in Table 1.

**Table 1.** Summary of the data sets

Patient/study	$N$	$N_+$	$N_-$	$N_{\text{train}}$	$N_{\text{test}}$
1/gated	32	24	8	24	8
1/fb	27	19	8	20	7
2/gated	41	33	8	30	11
2/fb	39	31	8	29	10

fb: free-breathing;  $N$ : total number of the slices extracted for the given patient/study;  $N_+$ : number of slices containing tumour;  $N_-$ : number tumour-free slices;  $N_{\text{train}}$ : number of slices used to form the training set (randomly selected from  $N$  slices);  $N_{\text{test}}$ : number of slices used to form the test set (remaining  $N - N_{\text{train}}$  slices)

### 4.3 SVM Training and Model Selection

We used  $\mu$ -SVM regression estimation [15] with Gaussian kernel. This variety of SVM algorithm has three parameters ( $\gamma$ ,  $C$  and  $\mu$ ) that must be set during the training phase. We approached the problem of finding their optimal values (*model selection*) by performing a five-fold cross-validation [17] on a logarithmic grid. The total training time per study was about 15 minutes (selection from 1764 different combinations of the parameters by means of 5-fold cross-validation for five different training sets).

### 4.4 Results

The following three metrics were used to evaluate the results. First, the correlation coefficient was calculated between the reference thresholds and those predicted by the algorithm. The other two measures evaluate the quality of the results in terms of geometric similarity of the regions contoured with the reference thresholds, and the regions outlined by the algorithm-predicted thresholds. To this end, Jaccard and Dice's similarity indices were calculated

$$J = |R \cap A| / |R \cup A| \quad (2)$$

$$D = 2|R \cap A| / (|R \cup A| + |R \cap A|) \quad (3)$$

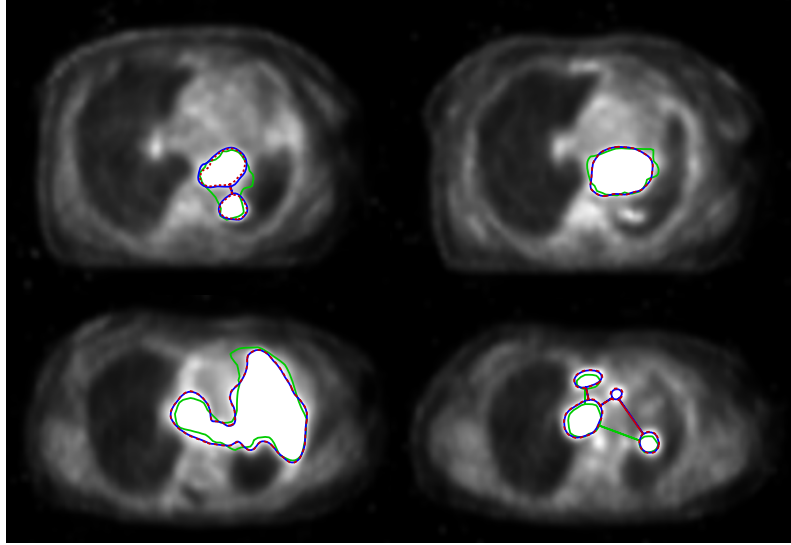
where  $R$  and  $A$  stand for the regions contoured by the reference and algorithm-predicted thresholds, respectively. Both Jaccard and Dice's indices are equal to zero when two regions have no common area, and are equal to unity when the regions match perfectly.

The results are summarized in Table 2 (since no significant difference was found between the results of the gated and free-breathing studies in the same patient, we present their averages). Several segmentation examples are presented in Fig. 4. Besides the high level of geometric similarity, a correlation between the predicted and the reference uptake thresholds also indicates that during the training phase, the learning algorithm effectively acquired the dependency between the attributes and the reference uptake threshold.

**Table 2.** Summary of the results

Patient	$C$	$J$	$D$
1	0.71	0.82 (0.61)	0.89 (0.74)
2	0.83	0.96 (0.77)	0.98 (0.87)

$C$ : correlation coefficient between the reference and algorithm-predicted thresholds;  $J$  and  $D$ : Jaccard and Dice’s similarity indices between the regions contoured by the reference and algorithm-predicted thresholds (the values in parentheses represent the results obtained with the contrast-oriented algorithm [8], see Sect. 3.1)



**Fig. 4.** Examples of GTV, reference, and predicted uptake thresholds. Grey contour (green in colour version): GTV; solid black contour (blue in colour version): region contoured by the reference uptake threshold; dashed black contour (dashed red in colour version): region contoured by the algorithm-predicted uptake threshold.



Table 2 shows that we obtained better results for the second patient. We think this is because the second patient had a bigger tumour, occupying about 30% more PET slices (see Table 1), which resulted in a bigger training set, and hence better generalisation during training.

## 5 Discussion, Conclusion and Future Work

Our approach to PET-based lung tumour definition extends standard threshold-based approaches in three important ways. First, we base the definition of the optimal thresholds on a richer set of attributes extracted from the PET scans. Second, we use an “adaptable” machine learning algorithm capable of approximating data in a complex nonlinear way. Finally, we estimate the optimal threshold for each PET slice (instead of assigning a single threshold to all slices in the study).

The two threshold contours (reference and predicted) shown in Fig. 4 look very similar. However, this does not guarantee a high similarity between them and the GTV, as in case of the upper leftmost image, where both predicted and reference regions are composed of two contours, whereas the corresponding GTV is a single contour including some additional area. This illustrates an inherent limitation of *any* approach that is based on thresholding. This is simply because the shape of GTV contour can be whatever a radiation oncologist draws. In contrast, the shape of any particular threshold contour for a given image is fixed; and therefore choosing from even infinite number of different thresholds means choosing from an infinite number of different, but FIXED shapes.

This work was performed using a *study-specific* scenario, to cope with the slice-to-slice variation of the optimal threshold values within a study, and 75% of slices should have been defined manually in order to automatically define the remaining 25%. In order to prove its fitness for the real clinical use, an *inter-study/inter-patient* scenario with a sufficiently high number of patients is necessary; and we are currently exploring the challenges of collecting the necessary cases (i.e. checking and confronting both radiation treatment and diagnostic data available at our institution). In addition, we are looking for another informative PET attributes.

This work has demonstrated the potential advantages and applicability of the machine learning methodology as a tool to help plan radiation treatment for lung cancer.

**Acknowledgments.** This project has been made possible through a grant from the Alberta Cancer Board and the Alberta Cancer Foundation. R.G. was partially funded by NSERC and the Alberta Ingenuity Centre for Machine Learning.

## References

1. Canadian cancer statistics 2008. Canadian Cancer Society (2008)
2. Parkin, D.M., Bray, F., Ferlay, J., Pisani, P.: Global cancer statistics, 2002. *CA Cancer J Clin* **55**(2) (2005) 74–108
3. Nestle, U., Kremp, S., Grosu, A.L.: Practical integration of [18F]-FDG-PET and PET-CT in the planning of radiotherapy for non-small cell lung cancer (NSCLC): the technical basis, ICRU-target volumes, problems, perspectives. *Radiother Oncol* **81**(2) (2006) 209–25
4. Drever, L.A., Roa, W., McEwan, A., Robinson, D.: Comparison of three image segmentation techniques for target volume delineation in positron emission tomography. *J Appl Clin Med Phys* **8**(2) (2007) 93–109
5. Geets, X., Lee, J.A., Bol, A., Lonnew, M., Gregoire, V.: A gradient-based method for segmenting FDG-PET images: methodology and validation. *Eur J Nucl Med Mol Imaging* **34**(9) (2007) 1427–38
6. Naqa, I.E., Yang, D., Apte, A., Khullar, D., Mutic, S., Zheng, J., Bradley, J.D., Grigsby, P., Deasy, J.O.: Concurrent multimodality image segmentation by active contours for radiotherapy treatment planning. *Med Phys* **34**(12) (2007) 4738–49
7. Black, Q.C., Grills, I.S., Kestin, L.L., Wong, C.Y.O., Wong, J.W., Martinez, A.A., Yan, D.: Defining a radiotherapy target with positron emission tomography. *Int J Radiat Oncol Biol Phys* **60**(4) (2004) 1272–82
8. Nestle, U., Kremp, S., Schaefer-Schuler, A., Sebastian-Welsch, C., Hellwig, D., Rube, C., Kirsch, C.M.: 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* **46**(8) (2005) 1342–8
9. Nestle, U., Schaefer-Schuler, A., Kremp, S., Groeschel, A., Hellwig, D., Rube, C., Kirsch, C.M.: Target volume definition for 18F-FDG PET-positive lymph nodes in radiotherapy of patients with non-small cell lung cancer. *Eur J Nucl Med Mol Imaging* **34**(4) (2007) 453–62
10. Daisne, J.F., Sibomana, M., Bol, A., Doumont, T., Lonnew, M., Gregoire, V.: Tri-dimensional automatic segmentation of PET volumes based on measured source-to-background ratios: influence of reconstruction algorithms. *Radiother Oncol* **69**(3) (2003) 247–50
11. Drever, L., Robinson, D.M., McEwan, A., Roa, W.: A local contrast based approach to threshold segmentation for PET target volume delineation. *Med Phys* **33**(6) (2006) 1583–94
12. Drever, L., Roa, W., McEwan, A., Robinson, D.: Iterative threshold segmentation for PET target volume delineation. *Med Phys* **34**(4) (2007) 1253–65
13. Vapnik, V.N.: *The Nature of Statistical Learning Theory*. second edn. Statistics for Engineering and Information Science. Springer-Verlag (1999)
14. Scholkopf, B., Smola, A.J.: *Learning with Kernels*. MIT Press, Cambridge, MA (2002)
15. Smola, A.J., Scholkopf, B.: A tutorial on support vector regression. *Neurocolt2 technical report series* (October 1998)
16. Boucher, L., Rodrigue, S., Lecomte, R., Benard, F.: Respiratory gating for 3-dimensional PET of the thorax: feasibility and initial results. *J Nucl Med* **45**(2) (2004) 214–9
17. Hastie, T., Tibshirani, R., Friedman, J.: *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. Springer (2001)