

# Decision Support System for Lung Cancer using PET/CT Images

David Jakobsson and Fredrik Olofsson

Supervisors:

Anders Ericsson and Johan Karlsson, Lund Institute of Technology  
Andreas Järund, WeAidU in Europe AB

Examiner:

Kalle Åström, Lund Institute of Technology

7th October 2004

## **Abstract**

Interpretation of medical images is often difficult and time consuming, even for experienced physicians. The aid of image analysis and machine learning can make this process easier. This thesis presents a fully automatic decision support system for lung cancer diagnostics with images from combined Positron Emission Tomography and Computed Tomography. Algorithms for segmentation of the lung region, localisation of suspected tumours (hot spots), detection of the heart, noise reduction and feature extraction are evaluated. A graphical user interface is also developed. A database of 99 patients with and without lung cancer diagnosis is used to train and test two learning systems, Support Vector Machines, SVM, and Artificial Neural Networks, ANN. The best result of the classification is an ROC area of 89%, obtained using ANN with seven features and two classes, lung cancer and no lung cancer.

# Acknowledgements

The authors would like to thank:

- The staff at WeAidU for giving us the opportunity to write this thesis and showing enthusiasm for our work.
- Our supervisors Anders Ericsson and Johan Karlsson, Centre for Mathematical Sciences, LTH and Andreas Järund, WeAidU.
- Examiner Kalle Åström, Centre for Mathematical Sciences, LTH.
- The staff at Clinical Physiology at Malmö University Hospital for letting us stay in their already narrow spaced office and answering all our medical questions, especially Docent Sven Valind for long and meaningful discussions and Dr. Sophia Hansson who has put up with us in her room during this period.
- Professor Lars Edenbrandt, Clinical Physiology, Malmö University Hospital. Although Edenbrandt is also part of the WeAidU staff he deserves a special mention, he has shown a lot of interest in our work and given us a lot of valuable contacts.
- PET and Cyclotron Unit at Rigshospitalet in Copenhagen for supplying the image database and answering questions regarding PET/CT and lung cancer diagnostics.

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Background</b>	<b>4</b>
2.1	Lung Cancer . . . . .	4
2.2	PET/CT . . . . .	5
2.3	Decision Support Systems . . . . .	7
2.4	Related work . . . . .	8
2.5	Objectives . . . . .	8
2.6	Material . . . . .	8
<b>3</b>	<b>Segmentation of the Lung Region</b>	<b>11</b>
3.1	Method 1: Region Growing . . . . .	12
3.2	Method 2: Global Thresholding . . . . .	15
3.3	Comparison of Method 1 and 2 . . . . .	19
3.4	Splitting of the Lung Region . . . . .	19
<b>4</b>	<b>Locating and Evaluating the Hot Spots</b>	<b>22</b>
4.1	Thresholding and Reference Value . . . . .	23
4.2	Labelling . . . . .	25
4.3	Discriminate Unwanted Regions . . . . .	25
4.4	Remaining Hot Spots . . . . .	27
<b>5</b>	<b>Feature Selection and Classification</b>	<b>29</b>
5.1	Support Vector Machines . . . . .	30
5.2	Features . . . . .	32
<b>6</b>	<b>Results</b>	<b>34</b>
6.1	SVM Classification . . . . .	34
6.2	ANN Classification . . . . .	35
<b>7</b>	<b>Conclusion and Future Work</b>	<b>41</b>
7.1	Conclusion . . . . .	41
7.2	Improvement of Methods and Algorithms . . . . .	42
7.3	Choosing Programming Language and Implementation . . . . .	42
7.4	Graphical User Interface . . . . .	42

## CONTENTS

---

<b>A</b>	<b>The User Interface</b>	<b>43</b>
A.1	Views in the Interface . . . . .	44
A.2	Set Diagnosis . . . . .	44
A.3	Functions . . . . .	44

# 1 Introduction

The medical service has been enriched with a lot of new techniques for diagnostic imaging during the last decades. The technical development has gone from static 2-D to dynamic 3-D and from analogue to digital images. Interpretation of diagnostic images is often a difficult and time consuming process, even for experienced physicians.

This master's thesis presents a decision support system that helps physicians in this matter. Image analysis and learning systems are used to make an automatic interpretation and diagnosis of patients with suspected lung cancer. The technique behind the images used in this thesis is called Positron Emission Tomography, shortened PET, which is in the area of nuclear medicine. As the name hints, a radioactive substance is used to image the body. The latest PET scanners are combined with Computed Tomography (X-ray) and are then shortened PET/CT.

The two authors are studying towards a Master of Science in Engineering Physics with image analysis and computer vision as specialisation and the thesis is a cooperation between Centre for Mathematical Sciences, Lund Institute of Technology and the company WeAidU in Europe AB. WeAidU has currently one decision support system for heart images on the market and is interested in a new product in the field. Besides WeAidU's there are few other commercial decision support systems. There is also a close connection to the PET and Cyclotron Unit at Rigshospitalet in Copenhagen that has contributed with the image database and medical knowledge and Malmö University Hospital which has been the workplace for this thesis. Physicians at these hospitals have made the necessary standpoints regarding medical issues.

The goal is to develop a fully automatic decision support system for lung cancer and a functional graphical user interface. The thesis starts off with a background of lung cancer, PET/CT, decision support systems and the material used. In the next chapters, 3 and 4, the methods and algorithms developed for the decision support system are presented with their theory and results. Chapter 5 deals with learning systems and presents the features examined. The final results of the classification can be found in Chapter 6 and the conclusion and suggestion for future work in Chapter 7.

# 2 Background

## 2.1 Lung Cancer

Cancer is a group of diseases characterised by an abnormal and unregulated growth of cells. Tissue with abnormal cell growth is called a tumour and can be malignant or benign, which is the same as cancerous or non-cancerous. The main differences are that a benign tumour grows slower, will not spread and will usually not come back if it is surgically removed.

Lung cancer is, in competition with prostate and breast cancer, the most common type of cancer and the leading cause of death by cancer for both men and women in United States and Western Europe. The majority of all cases is caused by tobacco smoking. Exposure to asbestos, radon, uranium and arsenic are other risk factors. Lung cancer is a very deadly disease and has an inclination to metastasise (spread) to other parts of the body, e.g. the brain, liver, bone and bone marrow. In most cases this occurs before it is discovered. Usually, lung cancer happens after the age of 50.

There are two major groups of lung cancer, Small Cell Lung Cancer (SCLC) and Non-Small Cell Lung Cancer (NSCLC), which together cover more than 90% of all cases. They grow and spread in different ways and are treated differently. In general, SCLC, which is less common, is more aggressive and metastasises faster. The methods for diagnosing lung cancer include CT scan (Computed Tomography), PET scan (Positron Emission Tomography), MRI (Magnetic Resonance Imaging), bronchoscopy (examination of the airways with fiber optics) and biopsy (examination of lung tissue sample). The last method can be used to decide the type of cancer depending on what the cancer cells look like in a microscope.

The staging of lung cancer is an important step for deciding the right treatment. An international staging system (TNM classification) is often used, based on three characteristics.

- Growth of the primary tumour
- Extent of lymph node involvement
- Metastases in distant part of the body

A simplified classification is used for SCLC, with the two classes extensive and limited disease depending on whether the cancer has metastasised or not. Lung cancer is treated with surgery, chemotherapy and radiation depending on the type and stage of the cancer and the patient's performance status. See [7] and [14] for more information on lung cancer.

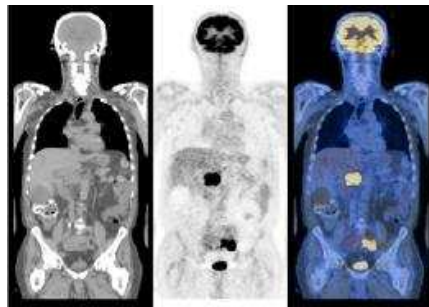


Figure 2.1: *The CT, PET and fused PET/CT images.*

## 2.2 PET/CT

### 2.2.1 Positron Emission Tomography

Positron Emission Tomography, PET, is a modern imaging technique in nuclear medicine for measuring and quantifying biochemical processes, see [2] and [20]. A radioactive isotope incorporated in a tracer substance is injected into the patient's body, the decay is measured from different directions with a detector and the image is reconstructed in a computer. This is the basic concept for nuclear imaging in general. The special characteristic of PET is that the isotopes used have a positron decay. The positron is annihilated with an electron in the tissue, causing two gamma rays in opposite direction that are registered by a ring of detectors. A coincidence in two opposite detectors is needed for a count to be registered. With this mode of detection, both absorption artifacts and random coincidences can be corrected for. This allows the distribution of isotope concentration to be measured in quantitative terms. As in all nuclear medicine images, PET images contain a level of background noise constituted by random coincidences.

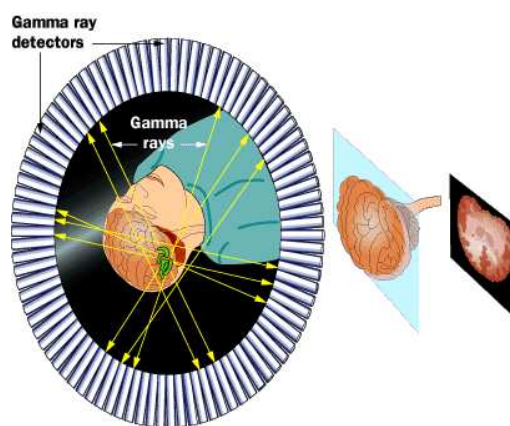


Figure 2.2: *An illustration of how coincidences are registered in the ring of detectors of the PET scanner.*

Depending on the tracer substance used, different kind of examinations are possible, for example regional and absolute blood flow, metabolism, protein synthesis, gene expression and tissue hypoxia. In this thesis, where the focus is on lung cancer, the isotope used is F-18 (with a half-life of 110 minutes) and the PET images show the metabolism of the tracer substance FDG (fluorodeoxyglucose). Tumours can, in this case, be detected since they have a higher consumption of the tracer substance compared to normal tissue.

Tomography is a technique for 3-D imaging, which makes it possible to look at a specific slice of the body. This is achieved by scanning the object in question from different angles and using iterative algorithms and mathematical transforms to reconstruct the image. Mathematically, the task is to reconstruct the function of three variables from integrals along the straight lines crossing the 3-D curve from the scanning process. The theory behind this is quite extensive and not essential for this thesis, it will therefore not be explained any further. There is a lot of literature on the subject, see for example [21].

One drawback of PET is that a scanner in operation needs to be located close to a cyclotron, because of the short half-life of the isotopes used. This circumstance makes the technology more expensive than for example Single Photon Emission Computed Tomography (SPECT). For an extensive description of the PET technique and PET diagnostics, see [21].

### 2.2.2 Computed Tomography and Combined PET/CT

In Computed Tomography, CT, X-ray photons are used to scan the patient's body from different angles and build a 3-D image. CT images give detailed anatomical information but no information about functionality. A tumour can for example be possible to detect in a CT image, but there is little information about malignancy and growth rate.

The combined PET/CT technique merges the advantages of these two methods for medical imaging in one device. It gives the possibility to make the PET and the CT scan at the same time point, when the patient lies in the same position and with a good match of the two images as a result. While the PET image gives information about the metabolism of the tumour, the CT image shows its location in the body. As a bonus, the CT images can be used for attenuation correction of the PET images which shorten the whole scan time with 20-30 minutes [20]. A typical scan time for a PET scanner without CT is 40-60 minutes. Attenuation correction is needed since not all of the decays will be registered [21]. Some photons will be scattered out of the field view or absorbed within the body. It is therefore essential to determine the probability for a coincidence event for each line that joins two detectors.

For PET scanners without CT, the attenuation correction map is created by using an external source of 511 keV photons, called a transmission scan. It basically works in the same way as CT but the image is of lower resolution and contains a higher amount of noise. The attenuation correction image is however used by the physicians in case the scanner is not combined with CT. It is probable that the same thing could be achieved in an automatic interpretation.

As seen in Figure 2.1 the PET and the CT images can be fused into a PET/CT image. Notice that PET images are usually displayed inverted, which means that regions of high uptake, hot spots, appear darker than the back-



Figure 2.3: *The GE Discovery LS PET/CT scanner at Rigshospitalet in Copenhagen.*

ground. In this thesis the original PET image is used for all analysis and processing, while the inverted image is used for displaying.

### 2.3 Decision Support Systems

To make a diagnosis of a full body FDG-PET scan takes typically one hour for two physicians at Rigshospitalet in Copenhagen. It is a difficult task, especially for new physicians that are practicing PET diagnostics. At the same time the resources for diagnosing the increasing number of medical images are shrinking and misinterpretations can lead to incorrect or no treatment of the patient.

Decision support systems using image analysis and learning systems, such as Artificial Neural Networks and Support Vector Machines, can be a great help in the medical service and an efficient educational tool. The idea is to gather the knowledge of medical experts in the area in a computer program that should be able to make an automatic diagnosis. When the physician has made an interpretation, the diagnosis can be verified with the automatic diagnosis of the decision support system. It works as a second opinion that perhaps sometimes can replace the need of two physicians. If the physician's opinion agree with the one of the decision support system, the diagnosis is secured, and if it disagrees it encourages the physician to look at the images again or ask a colleague for complementary opinion. A decision support system can also provide information that facilitates the diagnostic procedure. In this thesis image analysis is used to automatically find and highlight suspected tumours in a graphical user interface, together with the relevant information used by the system to make an automatic diagnosis.

One of the most important components of a decision support system is the database of examples used to train and test the learning system. The database should contain the same type of medical examinations as the system will be used for. During the learning phase the system is fed with these images together with diagnoses made by medical experts and thereby learns how to mimic a physician.

## 2.4 Related work

It is not easy to find earlier work about decision support systems and automatic interpretation of nuclear images. In [23], an Artificial Neural Network is used to improve nodal staging accuracy for NSLC with F-18 FDG-PET images. Holst et al. describe methods for automatic interpretation of electrocardiogram and lung scintigrams in [11]. Ericsson et al. used Support Vector Machines for automatic interpretation of diagnostic images in [9]. Other sources of inspiration are the CARE system that is developed and sold by WeAidU and an earlier master's thesis by Andreas Järund [15].

## 2.5 Objectives

The goal of this master's thesis is to develop a decision support system using PET/CT images. By suggestion of physicians at the PET and Cyclotron Unit, Rigshospitalet in Copenhagen the focus will be on lung cancer, since it is a common disease and fairly easy to detect. As the physicians, we will take advantage of the CT image to achieve a better orientation in the PET image.

As a first step, a study of relevant literature and articles on the subject has been carried out. Segmentation of CT images, feature extraction of cancer in PET images and earlier decision support systems was of special interest. The medical aspects of lung cancer and PET/CT was also of importance. In the next step, methods for image segmentation in the CT image and for locating regions of interest to extract relevant features of lung cancer in the PET image are constructed. The features should ideally be independent of individual variations of the patients. The final step is to train and test a learning system to evaluate the possible features. Above all this, a Matlab graphical user interface in which the decision support system can be run was implemented.

## 2.6 Material

The full body PET/CT scans used in this thesis were collected from the PET and Cyclotron Unit at Rigshospitalet in Copenhagen and a GE made Discovery LS PET/CT scanner. The database consists of 35 healthy patients, in question of lung cancer, and 64 patients with lung cancer diagnosis, examined between February 2002 and February 2004.

Each scan has an equal number of PET and CT slices, most often 171, 205 or 234, covering the head down to the pelvis. The image slice properties are listed in Table 2.1. Although the slice thickness has a difference of 0.75 mm, the slices correspond to each other since the CT slices have an overlap. A PET/CT

	<i>PET</i>	<i>CT</i>
<i>Image size</i>	128 · 128 pixels	512 · 512 pixels
<i>Slice thickness</i>	4.25 mm	5 mm
<i>Pixel size</i>	3.98 mm	0.98 mm

Table 2.1: *The resolutions of the PET and CT images.*

## 2. Background

---

scan is saved in DICOM format [8], which is a standardised file type for medical use with a header consisting of information tags. The pixel data of each slice has a bit depth of 15 bits, giving values between 0-32767. In order to get the tissue activity in each point, Bq/cc, as measured by the PET/CT scanner, the pixel data needs to be rescaled by the information of the tags 'Rescale slope' and 'Rescale intercept' in the DICOM header. Each image slice has different values of these tags. By using the formula

$$y = a \cdot x + b,$$

where  $x$  is the original pixel value,  $y$  the new output pixel value,  $a$  'Rescale slope' and  $b$  'Rescale intercept', for each image slice of the PET scan, the pixel values are rescaled to tissue activity, Bq/cc.

The CT pixel data is in Hounsfield units, HU, which is a standardised and accepted system for displaying CT images. The original linear attenuation coefficients are rescaled with an affine transformation into a system of unit where water is defined as 0 and air -1000.

### 2.6.1 Standardised Uptake Value

Calculation of Standardised Uptake Value, SUV, is a common way to compare tumours patient to patient. The tissue activity is then compensated for the injected dose, the time between injection and scan start and the total body weight of the patient. Using the body weight is most common in SUV calculation, though some hospitals and physicians prefer to use the body surface instead. SUV is calculated by assuming that 1 cc = 1 g and applying

$$SUV = \frac{A \cdot w}{d}$$

for every voxel, where  $A$  is the tissue activity (Bq/cc),  $w$  the patient's body weight (kg) and  $d$  the injected dose at scan start (Bq). With this definition follows that if the injected dose would be equally distributed over the whole body, then each point should have a SUV of 1. The injected dose provided in the DICOM header is, in this case, the activity in the dose at the time for injection. The time between injection and scan start varies between one and two hours and during this time the tracer substance is distributed over the whole body. The activity of the injected dose at scan start is given by

$$d = d_{inj} \cdot e^{-\lambda \cdot t_{diff}},$$

$$\lambda = \frac{\log 2}{T_{1/2}},$$

where  $d_{inj}$  is the activity of the injected dose at injection time,  $t_{diff}$  the time between injection time and scan start and  $T_{1/2}$  the half-life for F-18.

An SUV calculation is not very exact, but differences between subjects in the amount of tracer injection, delay between injection and scan start and body size are accounted for. It is closer to a comparable value than the tissue activity itself. When describing a tumour, physicians often take the maximum SUV in the area of the hot spot and it usually does not exceed 15. A SUV of 15 would

## 2. Background

---

be considered as an extremely aggressive tumour. A SUV of 7-8 is more typical. Since nuclear medicine images contain a high amount of noise, single voxels can always reach a high SUV and it is therefore important to include several voxels when determining the SUV of a region.

# 3 Segmentation of the Lung Region

The purpose of the segmentation of the lung region in the CT image is to achieve a better orientation in the PET image. Since the CT and the PET images correspond to each other, with some exception for artifacts caused by motion of the patient, we can with the knowledge of which voxels (3-D pixels) that are lung tissue say whether a suspected tumour lies within, on or outside the lung boundaries. Another reason is to define which transaxial image slices that belong to the thorax, to be able to search for hot spots only in that part of the image. For this matter, we assume that the thorax is the region between the top and the base of the lungs. The abdomen is more difficult to interpret since it is hard to separate tumours from other types of uptake and it is therefore an advantage to be able to exclude it, at least in the first step. Also, the focus is on detecting lung cancer which means that tumours should be present in the thorax and lung region.

One important question here is how large the mismatch between the PET and the CT images can be. It should be expected to be larger for the lung region compared to other parts of the body because of the breathing of the patient. In [4], the accuracy of PET and CT spatial registration of lung lesions was evaluated for 36 patients with a total of 48 lesions. The position of the centres of the lesions were estimated in the PET and the CT images independently and then compared. The overall mean distance between the centres was 7.55 mm, with a standard deviation of 4.73 mm. It tends to be a little more in the lower lungs than in the upper and also in the left lung in comparison to the right. The conclusion is that the spatial registration usually is accurate, but misregistrations can occur. The consequence of this mismatch is that it is possible that a part of a hot spot (or even the whole hot spot in an extreme case) will falsely be registered within or outside the lung boundaries. Hot spots can occur both inside and outside the lung region as we will see in the next chapter.

A lot of articles can be found regarding segmentation of the lung region in CT images. Hu et al. [12] describe a method of global thresholding for that purpose. Pohle and Toennies [18] suggest adaptive region growing for segmentation of medical images. In this thesis both region growing and thresholding are tried and discussed, although the later one, Method 2: Global Thresholding, is chosen. The aim here is to develop a fully automatic method for segmentation of the lung region.

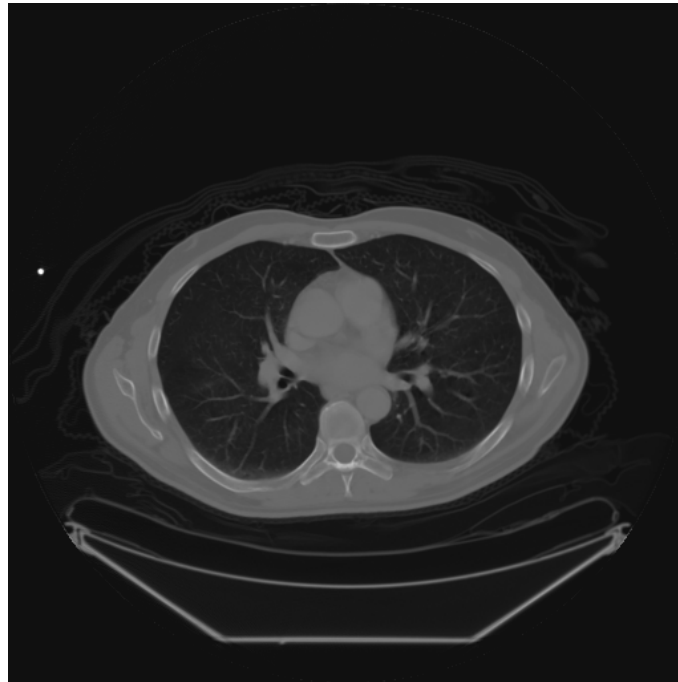


Figure 3.1: *The task of this section is to segment the lung region which can be seen as the dark region in the body. The surrounding tissue appears with a higher intensity.*

## 3.1 Method 1: Region Growing

Region growing is a recursive algorithm that groups pixels or subregions into larger regions. The grouping is based on predefined criteria, for example grey-level or colour threshold values. A region growing usually starts off with one or several seed points and then recursively add neighbouring pixels that fulfill the criteria to the region. Different connectivities can be used, in 2-D region growing the choice is usually between four and eight-connectivity. Four-connectivity can refer to both the four neighbours to the left, right, top and bottom and the four diagonal neighbours, while eight-connectivity refers to all pixels next to the seed pixel [10]. A stopping criterion needs to be set up for the growing to end. The algorithm ends by itself when no more pixels fulfill the criteria on a pixel level. Additional criteria that take into account the global perspective of the region growth are sometimes necessary to increase the power of the algorithm. For example the region growing could stop when the region reaches a predefined size. For a more extensive description of region growing, see for example [10].

The difficult task here is to find seed points, one in each lung. As mentioned earlier, the goal is to achieve a fully automatic segmentation of the lung region. It is not a lot of work for a physician to make one mouse click in each lung, but a lot of small manual steps like this will make the procedure of using the decision support system more complicated and probably less inviting. A possibility of making a manual correction in case the automatic segmentation fails is a better

option. Also when considering the preparation of a training and testing set for a learning system, which will use a large number of examples, it saves a lot of work if this step is automatic.

#### 3.1.1 1-D Spatial Signatures

One way of finding the seed points is to use the spatial signatures. The 1-D spatial signature of a 2-D image is given by summarizing the rows or the columns of the image matrix. This can be done after an operation, such as thresholding, or directly on the original grey-level intensity values. From the 1-D spatial signatures, vertical or horizontal characteristics can be found if they appear as extreme points in the spatial signature. In a 3-D case this method could start

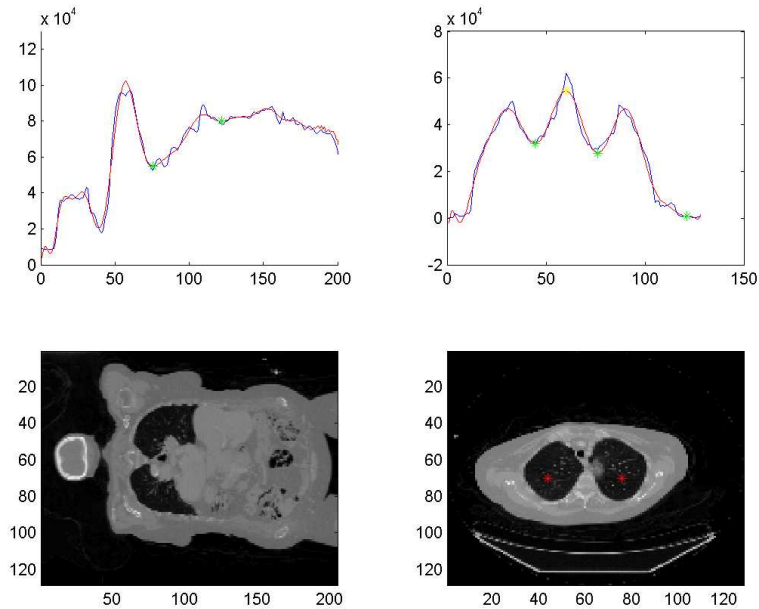


Figure 3.2: *The 1-D spatial signatures of the CT images in the frontal view (left) and transaxial view (right).*

with projecting the image in the x-, y- and z-direction and then use these three 2-D signatures to calculate the 1-D signatures. In this thesis, an initial guess of one of the three coordinates is made instead. In the CT images it is noted that the patient consistently lies in the vertical middle of the transaxial view. The bunk of the PET/CT scanner is adjusted for this purpose more or less by standard. With this in mind, the start image for calculating the 1-D signatures will be the one in the frontal view with the slice number given by the middle y-coordinate of the transaxial images, see Figure 3.1.

The x-signature of the start image, shown to the left in Figure 3.2, has a characteristic dip where the lungs are located. If a polynomial that fits the data can be found, it will be more easy to find this minimum by differentiating the

### 3. Segmentation of the Lung Region

---

polynomial. Another alternative is to use Matlab's smooth function to make the curve less harsh to facilitate the examination of the signature. When the minimum is found, we continue to the transaxial image with the slice number of the coordinate of this minimum. This transaxial image will ideally be the slice with the largest lung area. The x-signature of the transaxial image is calculated, shown to the right in Figure 3.2, giving two characteristic dips indicating the location of the lungs. Finally, the initial guess of the y-coordinate of the transaxial image is adjusted by calculating and finding the top value of the y-signature. For the two first signatures, it is important to use different criteria for finding the right minima. In this case the maxima is used to verify that the right minima have been found.

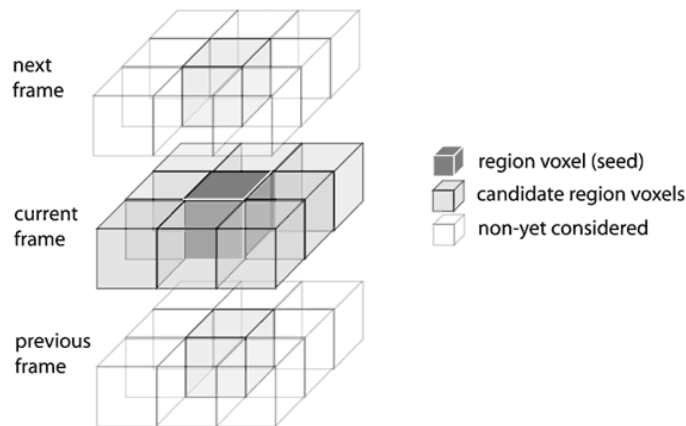


Figure 3.3: *The initial state of region growing, when only the seed voxel belongs to the region. After the growing is done on the current slice (frame), all the voxels of that region become possible seed points on the previous and next slice.*

#### 3.1.2 Region Growing of the Lung Region

In this thesis, region growing is performed on one slice at a time. The 2-D region growing is extended to 3-D by using all the pixels from the region of a slice as seed points in the next slice. This is done for all the slices below and above the initial slice until a predefined grey-level criterion is fulfilled, when the area of the segmented region in a slice is less than 40 pixels the algorithm does not include any more slices in that direction.

The region growing is only performed for a seed pixel if the pixel itself fulfill the criteria. A fixed interval of the grey-level intensity of the lung region can be used as a criterion but it is preferable to use an individual interval for each patient. An average intensity value of the lung tissue can be calculated by taking the seed points and a number of surrounding voxels. This value together with a predefined tolerance is used for both the initial seed points and all the following seeds.

The algorithm starts off with the start seed point and lets the region grow for the initial slice. When this is done, we continue by going up and down to the

neighbouring slices. All the pixels of the region for the previous slice are used as seed points, but that does not mean that we have to perform region growing for each of them which would be unnecessary and quite time consuming. First of all, as mentioned above, each seed point needs to be in the interval of the grey-level criterion. Secondly, when the region growing is done for the first seed point, all the pixels that have been covered by this step are simply deleted from the list of seed points. We can exclude them since a region growing would only result in pixels that are already covered by the previous step. This is applied on all the slices downwards until there are no more pixels left that fulfill the criteria of the region.

#### 3.1.3 Results of Method 1

This method yields a good result for most cases. However there is a problem with making the algorithm for finding seed points with 1-D spatial signatures robust. The CT images can look very different due to the size of the patient, the shape and condition of the lungs and how the patient's arms are placed. The arms should ideally be placed above the head and are done so in most cases, but for some patients this can not be done for various reasons, for example stiffness. Another difficulty is that the patient's head is sometimes not included in the scan and the same thing goes for the legs. The problem consists of making the algorithm independent of this sort of things. The signatures can look very different and to find the characteristic points in all cases is a complex problem. One solution of the problem could be to try to determine if the image of the patient has some kind of variation and in that case what kind of variation. If this can be achieved, a slightly modified algorithm for the specific type of variation could be used. There is also an obvious risk that the seed point is placed just outside the lungs or on something that has an intensity value of normal tissue, which is much higher than lung tissue.

## 3.2 Method 2: Global Thresholding

Thresholding is a simple, but yet useful, method for segmentation in the area of image analysis. It is a spatial domain method that operates directly on the pixels of the image, scanning them one by one. Suppose that the grey-level image is  $f(x, y)$ , the threshold value is  $T$  and the output image is  $g(x, y)$ , where the point  $(x, y)$  is a pixel position and the function value is the intensity of the pixel. Then an expression for the output image  $g(x, y)$  can be written

$$g(x, y) = \begin{cases} 1, & \text{if } f(x, y) > T \\ 0, & \text{if } f(x, y) \leq T \end{cases} \quad (3.1)$$

giving a binary image where pixels labelled 1 are object points and 0 are background points [10].

Shiyang et al. [12] describes a method for finding suitable threshold values in CT images through an iterative procedure. The concept is also discussed in [10] for the general case. As described in [12] the CT image data is in Hounsfield units (HU) for which air will appear with a mean intensity of approximately -1000 HU, lung tissue in the range of -910 HU to -500 HU and the chest wall, blood and bone most often above 0. A fix threshold (for example around -500

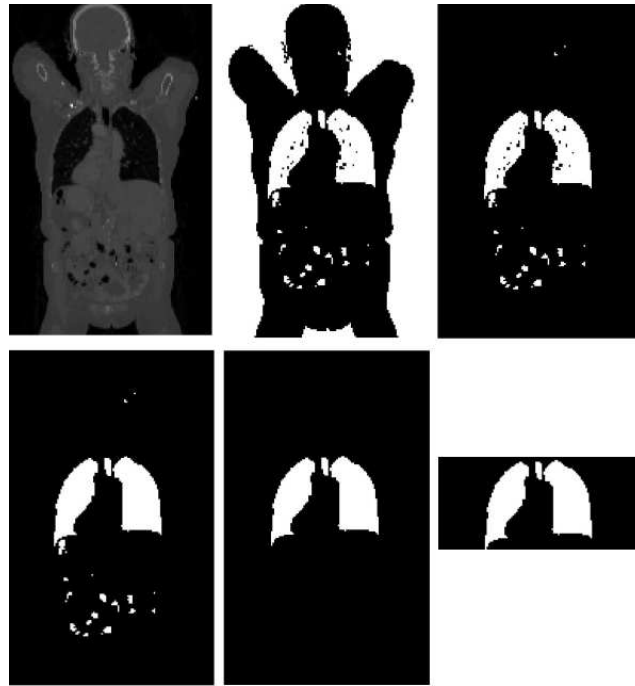


Figure 3.4: *The five steps of the thresholding algorithm.*

HU) would give a satisfying result in many cases, however finding an optimal threshold is preferable. This is achieved by defining two groups of voxels

- Body voxels (b) - chest wall, bone and blood
- Non-body voxels (n) - lung tissue and the surrounding air

and using the following algorithm

1. Select an initial threshold  $T_0$  (in this case  $T_0 = -500 HU$ ).
2. Perform thresholding as equation 3.1.
3. Use the resulting labels to compute average grey-level  $\mu_b$  and  $\mu_n$  for the voxels of the two groups.
4. Compute a new threshold value  $T_1 = 0.5 \cdot (\mu_b + \mu_n)$ .
5. Repeat step 2-4 until the difference between the threshold values in two following iterations is smaller than a predefined parameter  $T_d$  ( here  $T_d = 1$ ).

The algorithm is performed on the 3-D image yielding the lung region, the surrounding air and some small low density regions mainly caused by the airways and the intestines. An example of the resulting image is showed in Figure 3.5. The next step is to extract the lung region, which is done by identifying the 3-D

### 3. Segmentation of the Lung Region

---

connected regions in the binary image and label them with a unique number. The 26-connectivity, which refers to all the neighbouring pixels in a cube around the seed voxel, is used to determine which voxels that are connected. The algorithm used to accomplish this will be discussed in details in Section 4.2. The background, the surrounding air, can easily be removed since it is by far the largest region and also the only one connected with the borders of the image on all four sides (it is removed by switching the voxel values from 1 to 0). Once the background is removed the lung region can be identified since it is much larger compared with other cavities with high percentage of air in the body. The only difficulty here is that the lungs sometimes are connected and sometimes not in the resulting image. A size criterion can be used saying that if there are two large regions, the smaller of them has to be at least 10% of the size of the larger one. This could be a problem if the lungs are unconnected and one of the lungs are highly reduced in size because of a disease or surgery, and if there also exists another large air filled cavity in the body that could compete with it. Notice that this is only a predicted weakness and has not occurred for any of the 99 patients in the study.

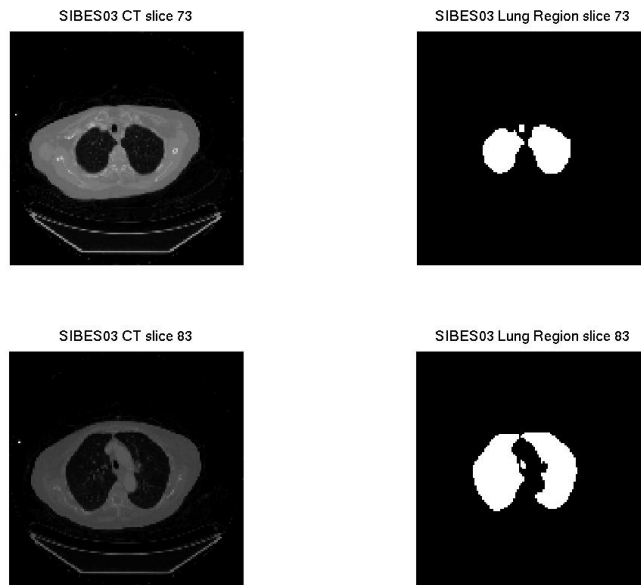


Figure 3.5: *Two CT slices and their corresponding segmented lung region.*

When the lung region is identified it is used to crop the full body image transaxial to cover only the thorax. The top and the base of the lungs are used for that matter. It is unfortunately not enough just taking the positions of the extreme points since the trachea (the windpipe) in most cases is connected to the lung region in the image, resulting in a small area of pixels in each image slice all the way up to the mouth. The same thing often happens for the base of the lungs and the intestines. Therefore, an area based criterion needs to be

### 3. Segmentation of the Lung Region

---

set up.

If we start in the image slice with the highest area of lung region pixels, go up and down checking the lung region area in each image slice and stop when the area is below a predefined parameter. Different parameter values have been tried and it was finally put to 40 pixels. The choice is a compromise between to sometimes loose some slices of the lung region and to sometimes get slices above and below the actual lung region. A parameter value of 40 pixels is quite low and favours the second case mentioned, but it is considered more important not to loose any part of the lungs since tumours can be located anywhere in that region.

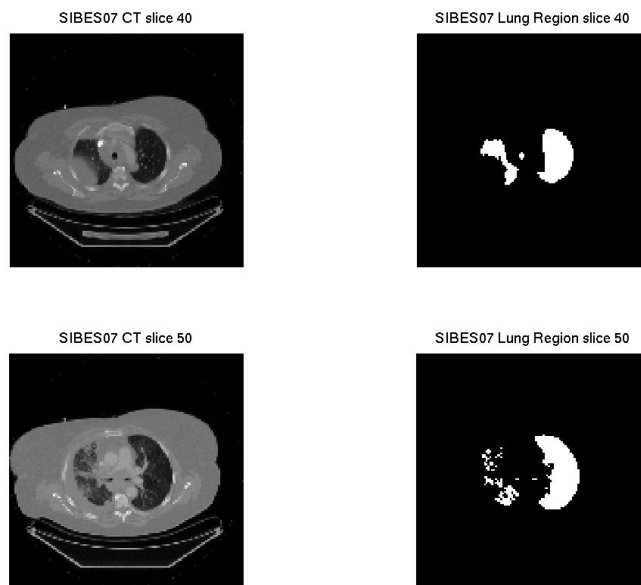


Figure 3.6: *The worst result seen of the segmentation of the lung region for the 99 cases in the database. The patient's right lung (left in image) is highly reduced in size.*

#### 3.2.1 Results of Method 2

The testing of this method on the 99 patients in the study gives very satisfying results. The lungs are never lost on any of the 99 patients, although sometimes when a lung has very high density the region becomes indistinct and consists of a lot of islands and holes. Figure 3.6 shows the worst case seen for these examples. Also, for some patients the cropping of the thorax does not succeed, resulting in a number of unwanted image slices of the neck or abdomen. Figure 3.7 shows an example of the whole sequence of slices of the thorax after segmentation of the lung region. Once the lung region is segmented, the boundaries can be overlaid on the PET image as in Figure 3.8.

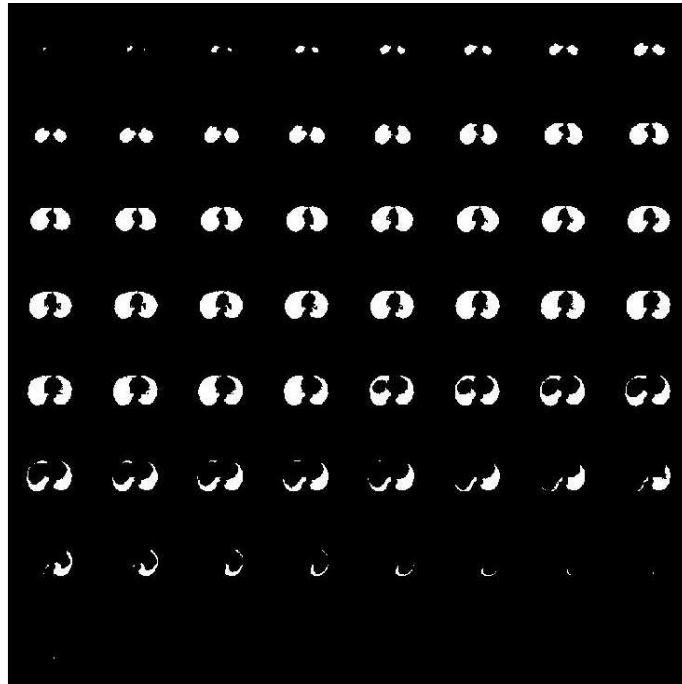


Figure 3.7: *All the image slices of the thorax after segmentation of the lung region, in this case 57 slices.*

### 3.3 Comparison of Method 1 and 2

Both method 1 and 2 give good results, though the second one is more suitable in this case. It never fails in finding the lungs for the 99 patients in the database while the first method has a couple of misses. When the lungs are found the two methods are more or less equal in finding the right area of the lung region. Thresholding is however more simple and does the two steps of method 1 in one single step. It is also less complicated to implement.

### 3.4 Splitting of the Lung Region

After the segmentation of the whole lung region a boundary between left and right lung is defined. The approach here is to find one plane that splits the 3-D lung region into two halves. Another alternative would be to define a line for every image slice.

For this purpose we calculate the centroid, i.e. centre of mass, of the 3-D lung volume. A binary image of an object is defined as

$$I(x, y, z) = \begin{cases} 1, & \text{for points on the object} \\ 0, & \text{for points off the object} \end{cases} .$$

### 3. Segmentation of the Lung Region

---

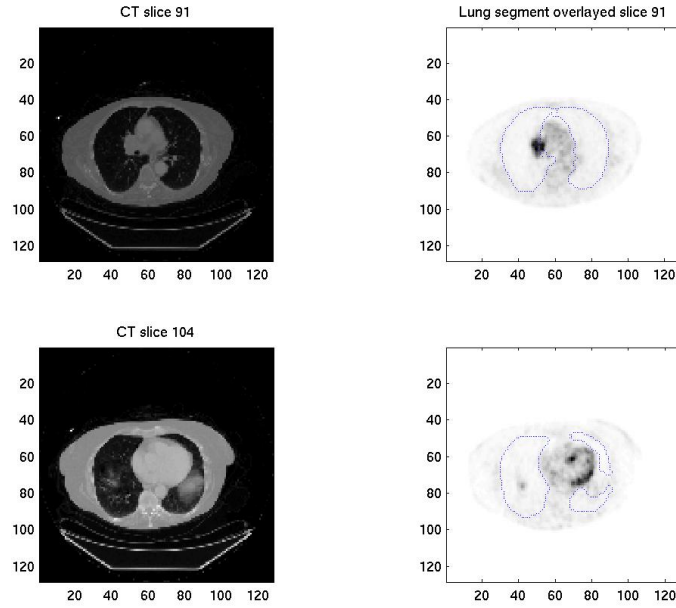


Figure 3.8: *Two CT slices of a patient and their corresponding PET slices with the lung region overlaid.*

The volume is then given by

$$V = \sum_x \sum_y \sum_z I(x, y, z)$$

and the centroid by

$$\begin{aligned} \bar{x} &= \frac{\sum_x \sum_y \sum_z x \cdot I(x, y, z)}{V}, \\ \bar{y} &= \frac{\sum_x \sum_y \sum_z y \cdot I(x, y, z)}{V}, \\ \bar{z} &= \frac{\sum_x \sum_y \sum_z z \cdot I(x, y, z)}{V}. \end{aligned}$$

The x-coordinate is used to define the plane that splits the lung region. An example of the results is shown in Figure 3.9.

### 3. Segmentation of the Lung Region

---

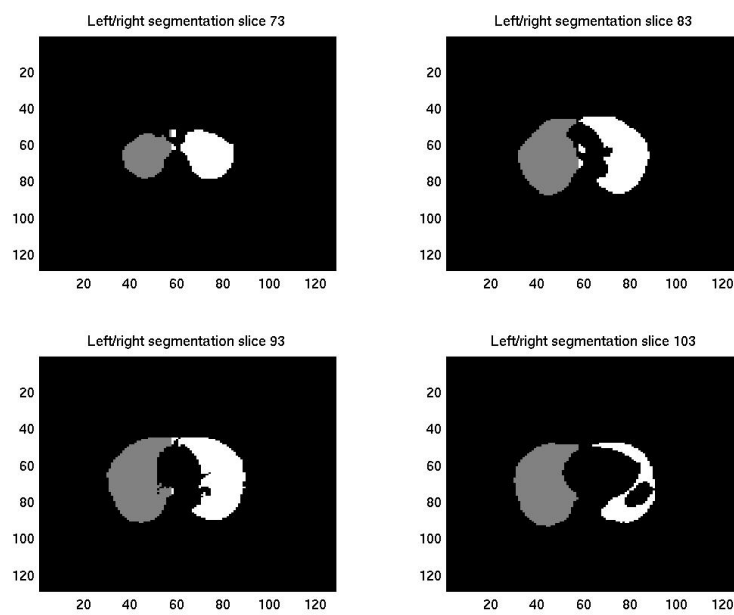


Figure 3.9: *The result of the splitting of the lung region in four slices. The grey area is the patient's right lung and the white area the left lung.*

## 4 Locating and Evaluating the Hot Spots

To make a correct diagnosis it is important to find the cause of the disease, which in case of lung cancer is tumours. In PET images tumours can be seen as groups of high intensity pixels. The intensity of a pixel shows the patients FDG uptake at that point, the higher intensity the higher uptake. A group of high intensity pixels is called a *hot spot* (see Figure 4.1). All hot spots do not come from tumours, high FDG uptake occurs in other types of tissue. Almost all metabolic activity causes FDG uptake, resulting in higher intensities in the PET image. Common sources of high intensity pixels are uptake in the heart muscle, i.e. myocardial uptake (see Section 4.3.2), inflammatory uptake and noise. Hot spots from tumours have an SUV from around three and above (see Section 2.6.1), while normal muscular activity values are from zero to two [13]. A segmentation of the hot spots and discrimination of normal uptake are made before the automatic diagnosis.

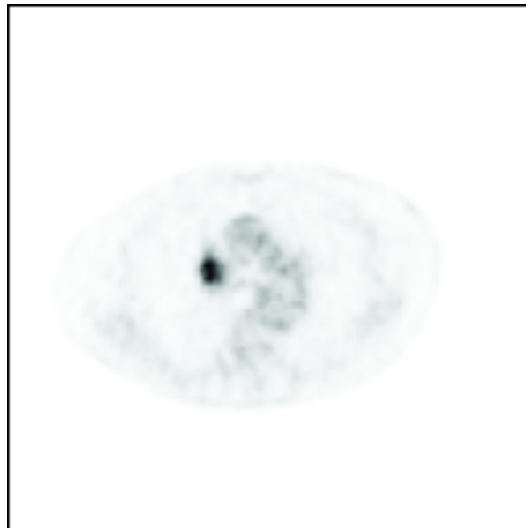


Figure 4.1: *The hot spot is the black region in this inverted image. It is common to display PET images inverted.*

## 4.1 Thresholding and Reference Value

To segment the hot spots a simple thresholding is made. Thresholding is a technique that filters pixels in an image that are above or below a threshold value. The technique was discussed in Section 3.2. Because of the variation in uptake level of the patients, there is a need to make this threshold individual. The individual variations have several origins, such as how the patient has fasted and level of muscular activity that the patient had just before the scan. To make this impact as small as possible all patients are advised to fast and rest before the scan.

One way to obtain an individual threshold is to examine the FDG uptake level in an area that can easily be compared to other patients. Such an area is located in the back. A number of pixels are segmented in the area and the mean SUV is compared to other patients with known thresholds. The area in the back contains bones and muscles but no other organs and the uptake value is often independent of lung cancer. It is therefore a suitable area when comparing patients.

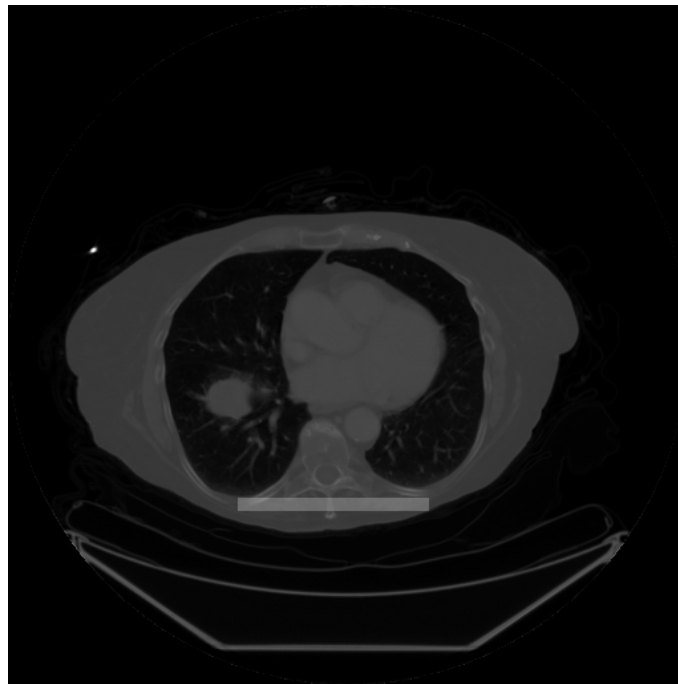


Figure 4.2: *The region from which the individual threshold is calculated is white marked.*

### Locating the Reference Area

The reference area is located in the back and the lung region is already segmented (see Chapter 3) which makes it possible to find this area. The slice where the lungs have the largest area is selected (in the CT image). This

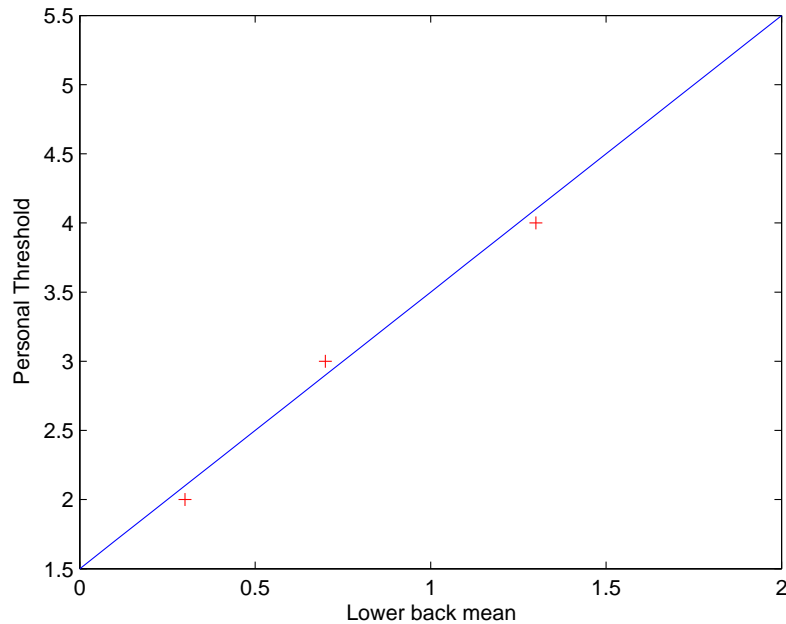


Figure 4.3: *The linear dependence of how the individual threshold is chosen.*

transaxial slice lies in the middle of the lung region and the next slices both upwards and downwards are selectable without going outside the thorax region. In both the right and left lung the coordinate for the lowest located pixel of the binary lung region is found  $(x_r, y_r), (x_l, y_l)$  in the transaxial view. The lowest value of  $y_r$  and  $y_l$  is selected to constitute the upper two corners' y-coordinate of the area. The x-coordinates,  $x_r$  and  $x_l$ , of the corners are the already selected (see Figure 4.2). The other two corners are selected 20 pixels downwards. A check is also made to see that the whole area lies within the body or else two new corners closer to the upper ones are found. This area is also located in three slices upwards and three slices downwards. To calculate the mean SUV the area found is applied to the PET image. In the PET image the mean intensity of the area is calculated. This mean value will be referred to as the *reference value*.

#### Calculation of the Threshold Equation

A translation of the reference value to a threshold is needed. The average reference value is after discussion with experienced physicians decided to correspond to a threshold of three. The highest reference value in the database is decided to correspond to four and the lowest to two. A linear dependence between these points is adapted and thereby a translation between the reference value in the back and an individual threshold is established (see Figure 4.3). After adaptation the threshold equation becomes

$$threshold = 2.0 \cdot (reference\ value) + 1.5. \quad (4.1)$$

After the personal threshold is calculated the actual thresholding is made. The output is a binary image.

### 4.2 Labelling

To find properties of a hot spot, such as volume and mean SUV, an extraction of the hot spot has to be made. In this step, which is called labelling, all pixels that belong to the same hot spot are given the same number. An overview of the algorithm follows [1].

- Given a binary image B.
- Negate the image (change all 1 value pixels to -1).
- Find a -1 pixel and change its value to a label not yet used.
- Find all its 26-neighbouring -1 pixels and give them the same label.
- Do this recursively for all -1 pixels that have a 26-connectivity.

All pixels above the threshold that have a 26-connectivity will get the same number and thereby each hot spot a unique label. The output of this algorithm is a 3-D matrix that contains only integers. The background is labelled as zero and the objects in order as they are found [17].

### 4.3 Discriminate Unwanted Regions

#### 4.3.1 Noise

Due to noise and other artifacts the labelling algorithm produces a lot of regions, especially small regions containing one to ten pixels. The pixels of the labelled regions are counted and regions with less than 27 pixels are considered as too small. These region are labelled as noise. This threshold of 27 pixels was found by testing values until no tumour was considered as noise. In the 27 pixel threshold no possible tumour was labelled as noise but instead a lot of noise was correctly removed.

#### 4.3.2 Myocardial Hot Spots

A hot spot that is very often seen in FDG-PET images is the heart, i. e. myocardial uptake (see Figure 4.4) [19]. The SUV value of the myocardial uptake can be as high (or higher) as a tumour's uptake value. The heart muscle often has a very specific shape and volume and the position of the heart is also known. To avoid labelling the heart as a tumour, a discrimination of hot spots having the specific shape, volume and position is made. Notice that cancer in the heart is very rare. Uptake in a healthy heart is almost always distributed uniformly over the muscle (as in Figure 4.4).

The discrimination algorithm is an expert system that checks if hot spots fulfill certain criteria regarding shape, volume and position. There are five criteria. If a criterion is fulfilled one or two points are given. An overview of how the algorithm works follows.

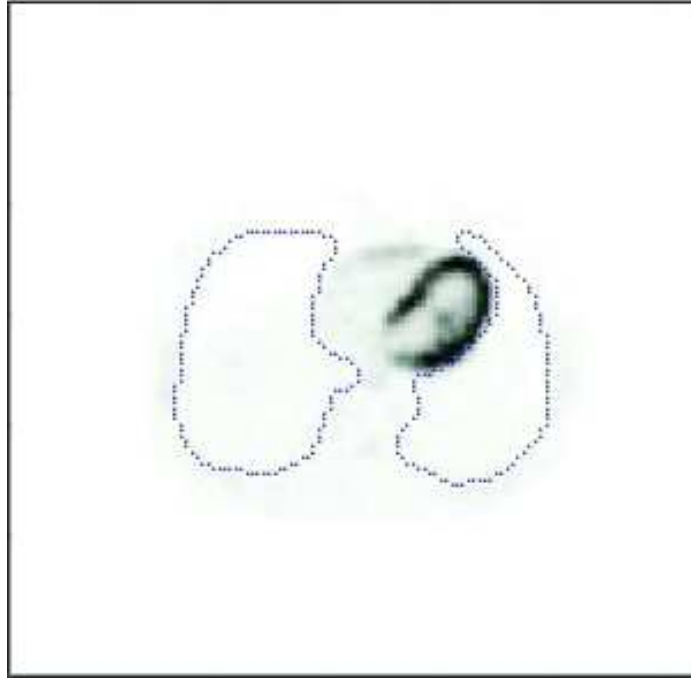


Figure 4.4: *The myocardial FDG uptake is very obvious.*

<i>Criterion</i>	<i>Points</i>
Position	2
Volume	1+1
Shape	2
Centroid Value	1
Extension	1

Table 4.1: *The criteria and how much each criterion is worth.*

The first criterion is the position criterion. A coordinate system is inserted into the image slice given by the z-coordinate of the hot spot's 3-D centroid (see Figure 4.5). The centre of mass of the body region in the image slice is used as origin of the coordinate system. Two points are given if the hot spot is located in a certain sector (see Figure 4.5). The next criterion is a two point criterion, the volume criterion. If the hot spot has a volume in the intervals 1000 to 1999 voxels (65 to 307 cc) or 3001 to 6500 voxels (195 to 420 cc) it gets one point and two points are given if the volume is in the interval 2000 to 3000 voxels (130 to 194 cc). Because of the rounded shape of the heart, the pixels in the planes given by the 3-D centroid coordinates can be summarized and compared. If one views the heart from all the three directions (frontal, sagittal and transaxial) it should roughly have the same amount of pixels. The amount of pixels in the views is compared and if they all are about the same, the hot spot is given two points. Next criterion considered is also regarding the shape. Because the

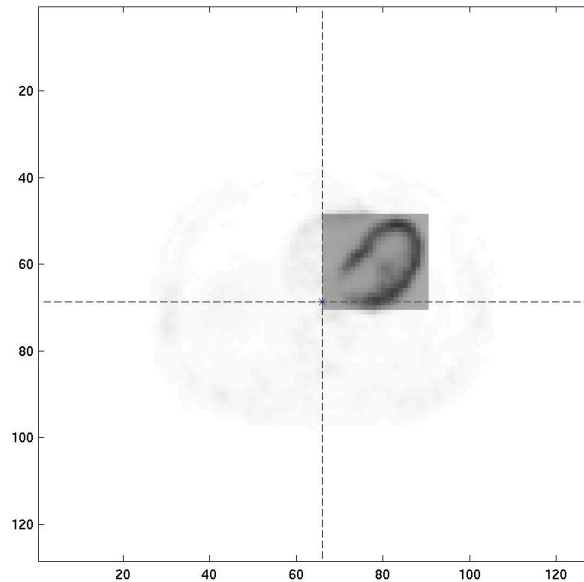


Figure 4.5: *The shaded region is the allowed position for a hot spot to be considered as a heart.*

heart has a shape looking like a horse shoe, the centroid is more likely to lie within the shoe and actually not on a high intensity pixel. If this is true another point is given. The last criterion concerns the bounding box of the hot spot. No dimension is allowed to have an extension much larger than any other. All three extensions are measured and compared.

Maximum score is a total of eight and the limit for a hot spot to be classified as heart is five. A patient could have a heart condition, it is then likely that the distribution of uptake looks different. If this is the case our system would probably not recognise the heart.

### 4.4 Remaining Hot Spots

After the heart is located (if any is visible) and noise removed a number of hot spots may be left unclassified. These hot spots are non-small, do not look like the heart and lies within the thorax region. Just because they are not classified as noise or heart, they are not necessarily tumours.

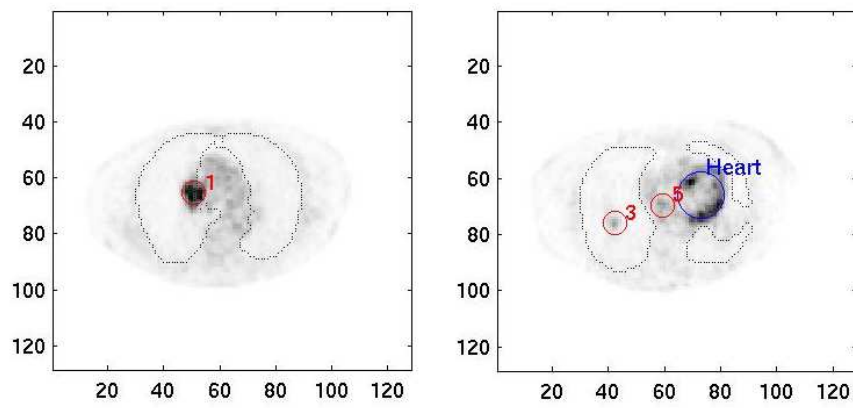


Figure 4.6: *The PET image with the heart and suspected tumours labelled.*

# 5 Feature Selection and Classification

The concept of pattern recognition is to train a system from samples, which consist of an arrangement of descriptors, often called features. A feature describes some property of the pattern and a pattern class is a collection of patterns grouped together by something they have in common. Two or more classes can be used, but in general a high number of pattern classes gives a more complex problem. It increases the number of features needed and for every feature, one more dimension is added to the mathematical problem. The final goal of pattern recognition is to build a system that automatically can assign an unknown pattern to the right pattern class.

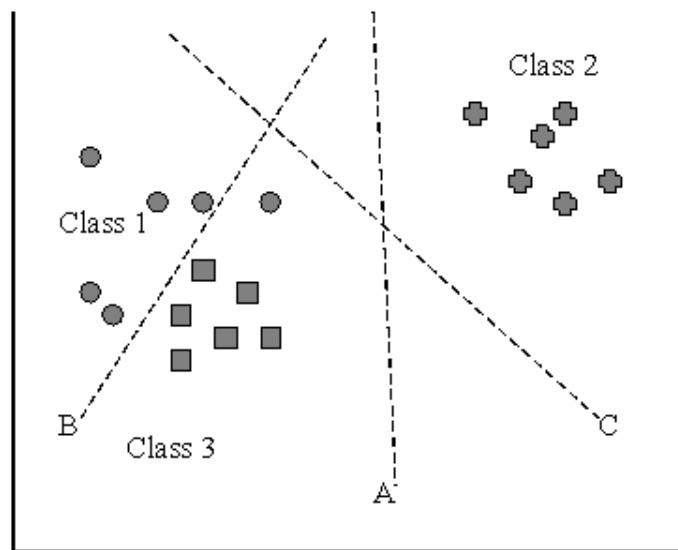


Figure 5.1: A simple classification problem with two features and three classes, rings, crosses and squares.

A 2-D problem with three classes, rings, crosses and squares, and two features can be illustrated as in Figure 5.1 with feature 1 along the x-axis and feature 2 along the y-axis. A simple approach to separate these three classes is to define

lines that divide the space into regions. The classification of unknown patterns will then be based on which side of these lines the pattern features are placed. This is the simplest form of classifiers, however using a nonlinear classifier is often preferable.

Artificial Neural Networks (ANN) [3], [6], and Support Vector Machines (SVM) [5], [22], are the two most common types of learning systems. SVM is the choice for the decision support system in this thesis since it is easy to use within Matlab. ANN is however also tried for evaluation of the features.

## 5.1 Support Vector Machines

Support Vector Machines (SVM) is a relatively new technique [5] introduced by Vapnik [22] in the 90's. SVM is a mathematical technique for solving the classification problem described above. In Figure 5.2, a classification is made for the same problem twice. A line separating two classes is drawn. The difference between the two cases is the margin width (the distance from the middle line to the dotted ones). The SVM chooses a separation line that has a large margin width and makes a small error in classification. When the feature space increases in dimensions, it is not possible to look at the problem like it is done in Figure 5.2. The line that separates the classes becomes a hyper plane. This is a plane with the dimension  $\dim(\text{feature space}) - 1$ . But the problem of maximising the margin width and minimising the classification error stays the same, even in higher dimensions. The name Support Vector Machines comes from the representation of the margin width, the dotted lines in Figure 5.2 parallel to the hyper plane are called support vectors [5].

SVM is also able to handle non-linear problems, which can not be separated with linear separation. A transform of input data maps the problem into a linear one (see figure 5.3).

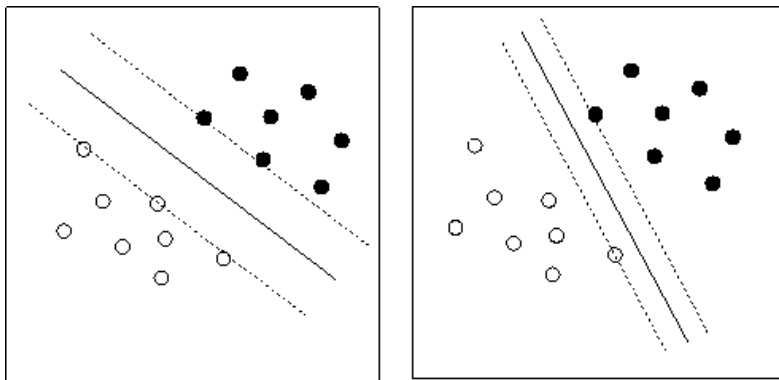


Figure 5.2: *Separation of two classes.*

Matlab does not come with an SVM toolbox, but several are available on the Internet. In this thesis OSU SVM is used [16]. This choice is made because it is very easy to get started with and the ability to set different parameters is very good.

SVM is a type of learning system, this means that you have to show your system how to classify before it can classify by itself. This is done in three steps:

- Learning
- Testing
- Classifying

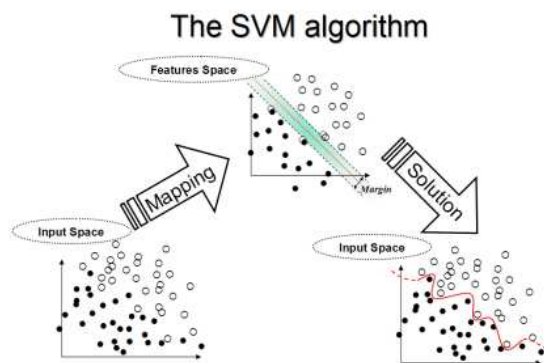


Figure 5.3: *How to classify a non-linear problem using mapping and support vector machines.*

### Learning

In the learning step you send your data to the SVM along with the correct answers. It is important that the SVM has good examples that represent your material well. As described above the SVM maximises the margin and minimises the classification error, all this is done in the learning step.

### Testing

After the SVM has been trained an important step is to test the system. It is desired to test the system with a number of examples for all the classes. Features along with the right answers are sent to the SVM. The SVM replies with a confusion matrix. This matrix contains statistics about how many of each class have been classified correctly, and if they were not classified correctly, what they were classified as. These two steps are done many times with different types of settings and features.

### Classifying

When the SVM has been validated it is ready for use. Features from unclassified objects are extracted and sent to the SVM, the SVM replies with a class for the object.

## 5.2 Features

The features examined in this thesis are:

1. Total number of hot spots found by thresholding
2. SUV of the hot spot with highest mean intensity
3. Number of hot spot voxels within the lung region
4. The volume of the largest hot spot found
5. SUV (on hot spot) / SUV (background)
6. The optimal threshold value used for the patient
7. Number of hot spot labels within the lung region
8. HU (on hot spot) / HU (background)

Some of the features are selected in consultation with physicians at Rigshospitalet in Copenhagen and some are inspired by [23]. A description of each feature follows below.

### 1. Total Number of Hot Spots Found by Thresholding

After thresholding, labelling and discrimination of small regions and myocardial uptake in the PET images, the remaining hot spots are counted. If we assume that the discrimination of unsuspected regions and the segmentation of the thorax perform well, the total number of hot spots should be a strong indicator whether the patient has lung cancer or not.

### 2. SUV of the Hot Spot with Highest Mean Intensity

With a similar discussion as above, the highest mean SUV of the hot spots are also an indicator for lung cancer. The mean SUV is chosen instead of a maximum SUV calculation to avoid high SUV caused by noise. This can happen for example in myocardial uptake. If the system would fail in the discrimination of myocardial uptake, the maximum of this region often becomes very misleading and it is therefore preferable to use the mean SUV to be able to compare different types of uptake. Tumours tend to have a more homogeneous uptake and the mean SUV can be higher though the maximum SUV is lower in comparison to the heart. Physicians normally use the maximum SUV but the situation is then very different.

### 3. Number of Hot Spot Voxels within the Lung Region

In the calculation of this feature, all the hot spot voxels that are represented in the lung region are counted. It is very rare to have a hot spot within the lung boundaries that is not a tumour. There is a risk however, that voxels of myocardial uptake will be counted if the discrimination algorithm fails. Myocardial uptake and other hot spots that lie close to the lung boundaries can have a slightly different size or location on the CT and PET image and can therefore falsely be represented in the lung region. In most cases it is only a

small part of these regions that are counted. A tumour that lies in the middle of the lungs has a strong effect on this feature.

### **4. The Volume of the Largest Hot Spot Found**

The volumes are calculated for the hot spots by counting the voxels of each label. The largest volume found is used as a feature. There is an obvious risk that myocardial uptake will have a negative influence since the system sometimes fails in detecting it. Myocardial uptake is in general much larger than uptake of tumours. There are also other types of regions with high SUV that can be very large, e.g. the liver.

### **5. SUV (On Hot Spot) / SUV (Background)**

The ratio of the SUV for the hot spot and the background of the hot spot is a way to measure the distinction or sharpness. The SUV of the hot spot is calculated as the second feature on the list and the SUV of the background is defined as the mean intensity value of the remaining voxels in the bounding box of the hot spot. Thus the background SUV is calculated individually for each hot spot. When this is done for all labels, the value of the feature is taken from the maximum ratio found. To explain the meaning of this feature, consider two hot spots with equal SUV, one located in the middle of a lung and the other in the tissue between the lungs, off the lung region. Even though they have the same SUV, the distinction of the two hot spots will be very different, as the background SUV is much lower in the lung region compared to the tissue outside the lungs.

### **6. The Optimal Threshold Value Used for the Patient**

The individual threshold value calculated for the patient as described in Section 4.1 is used as a feature. There is no medical explanation behind this choice, just an observation that the lung cancer cases in the database often have a lower mean SUV in the reference region (showed in Figure 4.2) and therefore also a lower individual threshold. The difference in age and health between the patients with lung cancer and without is a probable explanation.

### **7. Number of Hot Spot Labels within the Lung Region**

This feature is basically the same as the first one on the list and complements the third, though only hot spots that are represented within the lung boundaries are counted. The shared disadvantages of the third feature can be more critical here since the volume of the hot spots is not considered.

### **8. HU (On Hot Spot) / HU (Background)**

The discussion here follows the one of the fifth feature on the list. The only difference is that the Hounsfield Units of the CT images are used instead of SUV.

# 6 Results

In this chapter the results of the classification are presented. First of all, the test results of the Support Vector Machine used in the decision support system are discussed. Secondly, the data has been analyzed by an artificial neural network expert, Mattias Ohlsson, associate professor at Complex Systems Division, Department of Theoretical Physics, Lund University, who is connected to WeAidU. A neural network was trained and the features were individually evaluated.

## 6.1 SVM Classification

The SVM used in this thesis is a free Matlab toolbox, OSU SVM version 3.00, for which the non-linear RBF kernel is chosen. The training set consists of 20 patients without lung cancer, class 1, and 20 patients with lung cancer, class 2. The rest of the 59 patients, 15 without lung cancer and 44 with, is used for testing. The feature data is rescaled to the interval -1 to 1 for each input and the number and combination of features are varied in the test sequences. Since the number of patients in the database is small, the training and testing examples are chosen randomly through 100 iterations, where the SVM is trained and tested in each step. The mean class rate, mean confusion matrix and mean ROC area are calculated from the 100 iterations. An ROC (Receiver Operating Characteristic) curve is plotted for the one of the 100 iterations for which the ROC area (for the single step) is closest to the mean ROC area (for all 100 iterations).

In an ROC curve the sensitivity, which in this case is the share of class 2 (patients with lung cancer) that is correctly classified, is plotted against 1 - specificity, the share of class 1 that is falsely classified, for different cut values. The output values of the SVM are typically between around -1 and 1 and with a cut value of 0, the class belonging is simply decided by if the output value is positive or negative. By using another cut value than 0, it is possible to change the sensitivity and specificity of the classification to optimise the performance.

The results of the test sequences can be found in Table 6.1 - Table 6.9, which consist of one mean confusion matrix for each setting of features. A confusion matrix describes the insecurity of the classification. The first row shows how the samples of class 1 have been classified by the SVM. This means that the first column of the first row is the share of the samples of class 1 classified correctly, while the second column is the share classified falsely. In the second row, the first column is the share of class 2 classified falsely, while the second column is the share classified correctly. For example, Table 6.3 is a better result than

Table 6.4 for the classification of class 1, but worse for class 2.

The  $\gamma$ -parameter of the SVM, which determines the width of the RBF kernel, is adjusted to 0.1 in the first test sequence for all eight features and this value is kept through the rest of the testing. The default value 1 of the C-parameter of the SVM, which is the cost of constrain violation, is used. One feature at the time is excluded in the training and testing to see how it affects the classification performance. Depending on if the mean class rate and ROC area increase or decrease, we can according to this method say whether the feature is more or less important than the other ones. When this is done for all features, the least important feature (number 4) regarding the ROC area, is removed. This gives the best result so far, see Table 6.5. When moving on to excluding the second least important feature, the mean class rate and ROC area does not increase significantly and therefore the remaining features (1, 2, 3, 5, 6, 7 and 8) are kept. The ROC curves for all eight features and for feature 4 excluded are included in Figure 6.1 and Figure 6.2. The ROC areas of the test sequences are listed in Table 6.10.

### 6.1.1 Discussion

The testing indicates that none of the eight features is indispensable since the mean class rates and ROC areas stay on a high level no matter which feature that is excluded. Feature 4 seems to have little to do with this problem and tends to worsen the result when it is used. It is the least important feature seen to the ROC area. The combination of features in Table 6.5 is chosen for the decision support system since it is the best result seen to the ROC area.

## 6.2 ANN Classification

An artificial neural network is trained with the same data used for the training and testing of the SVM. An artificial neural network is another type of learning systems that is inspired by biological neural network, see for example [11]. It consists of interconnected elements, neurons, that all perform computations in parallel with a transfer function. The network between the neurons consists of virtual weights that tell how strong the connections between the nodes are. A neural network has at least two layers of neurons, the input and output. If several layers are included, those between the input and output layer are called hidden layers and in that case the classifier is non-linear. Each output node gives an output probability value between 0 and 1 and a cut value is used (usually around 0.5) for the classification.

Cross validation is used for the training and testing of the ANN, which is a method for making maximal use of a limited number of samples. The data is divided into k groups. The training is done k times, each time leaving one group out for testing. This means that, in the whole sequence, every sample will be used both for training and testing, only not in the same step.

Figure 6.3 shows the ROC curve of an ANN with five hidden nodes (non-linear classifier) and for all eight features. A sensitivity of 0.8 is a good choice here since the specificity is relatively high, 0.83, at this cut value. Figure 6.4 shows the sensitivity, specificity and total performance plotted against the cut value. A cut value of 0.58 would be suitable.

Table 6.3 - 6.4 shows the ROC area for each case when one of the eight features is excluded for both a non-linear ANN with five hidden nodes and a linear ANN.

### 6.2.1 Discussion

For the non-linear case the ANN shows a better result when feature 4 is removed. Feature 6 could be the most important feature in the linear case. Another interesting observation is that the problem seems to be more or less linear since the results of the linear ANN are as good (even slightly better for all inputs) as the non-linear. The best result, with largest ROC area, is given for the non-linear ANN when feature 4 is excluded. It seems to be the least important feature also in the ANN classification.

	<i>Class 1</i>	<i>Class 2</i>
<i>Class 1</i>	0.8373	0.1627
<i>Class 2</i>	0.2864	0.7136

Table 6.1: *The mean confusion matrix of the SVM for all eight features, mean class rate = 0.7451.*

	<i>Class 1</i>	<i>Class 2</i>
<i>Class 1</i>	0.8200	0.1800
<i>Class 2</i>	0.2720	0.7280

Table 6.2: *The mean confusion matrix of the SVM with feature 1 excluded, mean class rate = 0.7514.*

	<i>Class 1</i>	<i>Class 2</i>
<i>Class 1</i>	0.8240	0.1760
<i>Class 2</i>	0.2861	0.7139

Table 6.3: *The mean confusion matrix of the SVM with feature 2 excluded, mean class rate = 0.7419.*

## 6. Results

---

	<i>Class 1</i>	<i>Class 2</i>
<i>Class 1</i>	0.7913	0.2087
<i>Class 2</i>	0.2691	0.7309

Table 6.4: *The mean confusion matrix of the SVM with feature 3 excluded, mean class rate = 0.7463.*

	<i>Class 1</i>	<i>Class 2</i>
<i>Class 1</i>	0.8193	0.1807
<i>Class 2</i>	0.2814	0.7186

Table 6.5: *The mean confusion matrix of the SVM with feature 4 excluded, mean class rate = 0.7442.*

	<i>Class 1</i>	<i>Class 2</i>
<i>Class 1</i>	0.7913	0.2087
<i>Class 2</i>	0.2218	0.7782

Table 6.6: *The mean confusion matrix of the SVM with feature 5 excluded, mean class rate = 0.7815.*

	<i>Class 1</i>	<i>Class 2</i>
<i>Class 1</i>	0.7807	0.2193
<i>Class 2</i>	0.2541	0.7459

Table 6.7: *The mean confusion matrix of the SVM with feature 6 excluded, mean class rate = 0.7547.*

	<i>Class 1</i>	<i>Class 2</i>
<i>Class 1</i>	0.7800	0.2200
<i>Class 2</i>	0.2943	0.7057

Table 6.8: *The mean confusion matrix of the SVM with feature 7 excluded, mean class rate = 0.7246.*

	<i>Class 1</i>	<i>Class 2</i>
<i>Class 1</i>	0.8860	0.1140
<i>Class 2</i>	0.3850	0.6150

Table 6.9: *The mean confusion matrix of the SVM with feature 8 excluded, mean class rate = 0.6839.*

	<i>All</i>	<i>-1</i>	<i>-2</i>	<i>-3</i>	<i>-4</i>	<i>-5</i>	<i>-6</i>	<i>-7</i>	<i>-8</i>
<i>Area %</i>	87.3	87.1	87.8	86.0	88.1	87.7	83.8	85.0	87.1

Table 6.10: *ROC areas of the SVM.*

## 6. Results

---

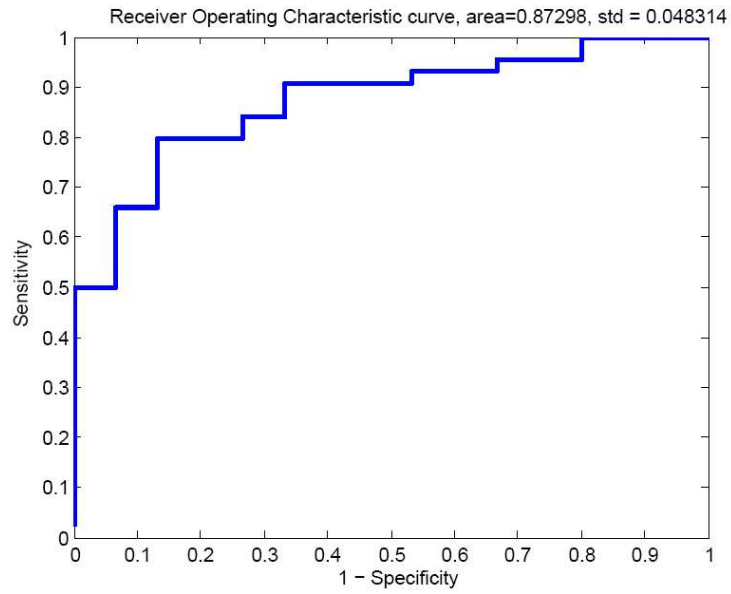


Figure 6.1: *The ROC curve of the SVM for all eight features.*

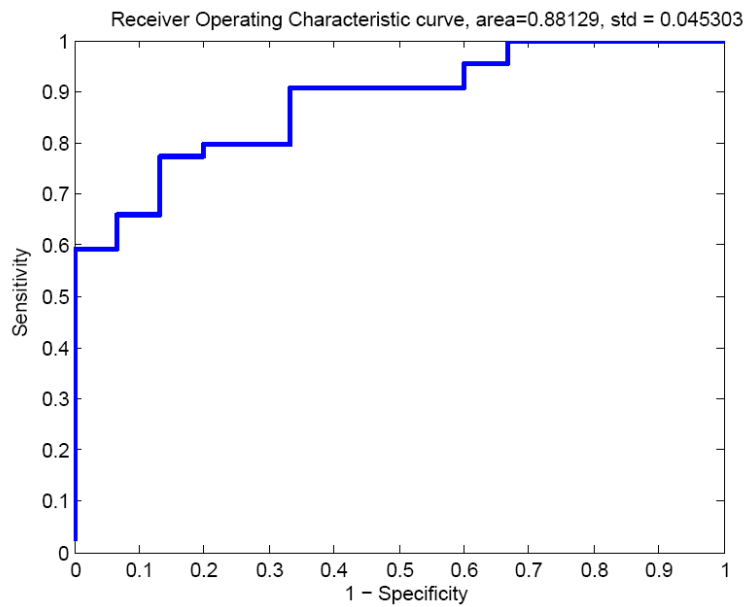


Figure 6.2: *The ROC curve of the SVM with feature 4 excluded.*

## 6. Results

---

	<i>All</i>	<i>-1</i>	<i>-2</i>	<i>-3</i>	<i>-4</i>	<i>-5</i>	<i>-6</i>	<i>-7</i>	<i>-8</i>
<i>Area %</i>	87.4	86.7	87.5	86.4	89.1	87.8	86.3	87.0	85.2

Table 6.11: *ROC areas of the ANN with five hidden nodes.*

	<i>All</i>	<i>-1</i>	<i>-2</i>	<i>-3</i>	<i>-4</i>	<i>-5</i>	<i>-6</i>	<i>-7</i>	<i>-8</i>
<i>Area %</i>	87.8	87.2	86.9	85.7	87.6	87.9	84.0	85.9	86.9

Table 6.12: *ROC areas of the linear ANN.*

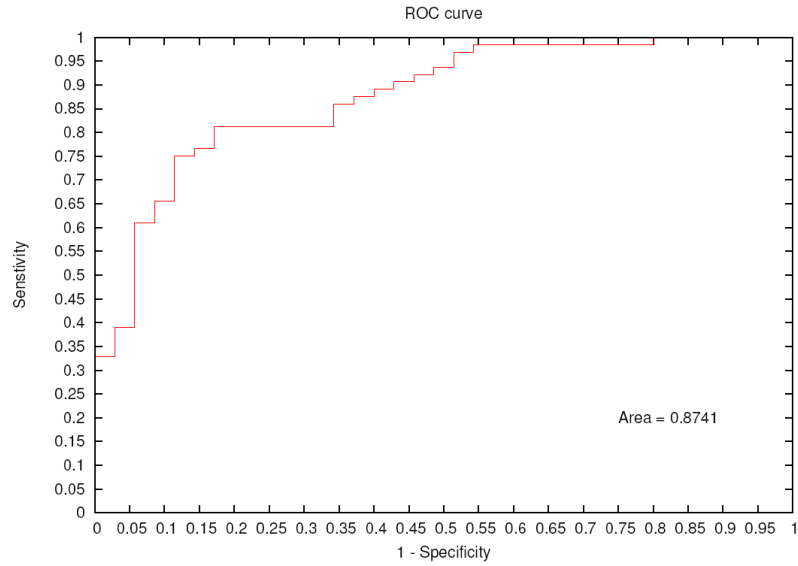


Figure 6.3: *The ROC curve of the ANN for all eight features and five hidden nodes.*

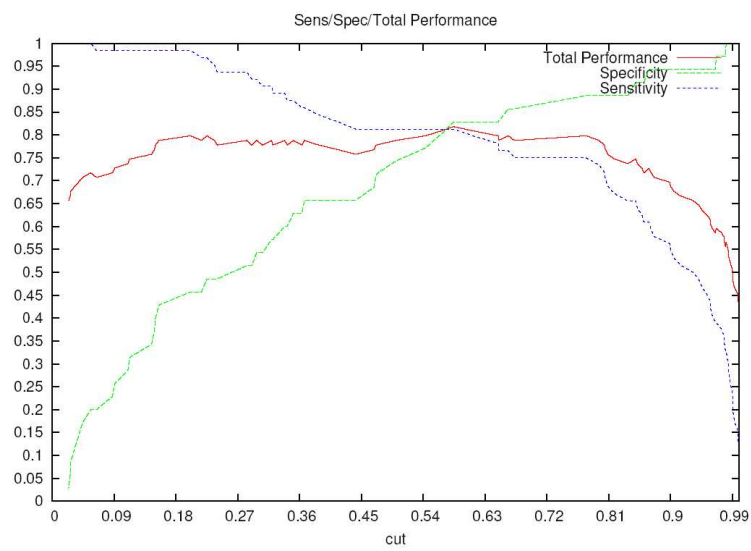


Figure 6.4: *The sensitivity, specificity and total performance versus cut value of the ANN for all eight features and five hidden nodes.*

# 7 Conclusion and Future Work

## 7.1 Conclusion

In this thesis, we have showed that it is possible to automatically detect suspected tumours in PET images and make an overall diagnosis for lung cancer with the help of machine learning. The CT image of the combined PET/CT scan was used to segment the lung region, which can be applied to the PET image with good accuracy. Uptake in the heart muscle can be recognized by its location, shape and size. Its occurrence is common in FDG-PET images and by separating it from the tumour suspected hot spots, both the automatic and manual overall diagnosis are facilitated. PET images contain an amount of noise. Single or a few neighbouring voxels can have a high intensity and it is therefore important to exclude these regions in the interpretation.

Eight features were evaluated with Support Vector Machines (SVM) and Artificial Neural Networks (ANN). The features, that were presented in Section 5.2, are based on the number, intensity, size and location of the hot spots. The segmented lung region can be used to determine whether a hot spot is located in the lungs or not. The two learning systems were trained and tested with a database of 99 patients, 64 with lung cancer and 35 without. For the SVM, the training and testing cases were chosen randomly 100 times. Cross validation were used for the ANN. By first testing all eight features and then making the same procedure with one feature removed at the time, an indication of the quality of each feature was given. This was used to improve the performance of the SVM classification by removing the least important features one by one. The best result was obtained using ANN with feature 4 excluded, which gave an ROC area of 89%. This feature combination also gave the best result of the SVM classification, an ROC area of 88%.

In comparison, the SVM and ANN show very small differences in performance. One interesting observation is that the linear ANN is equal in performance to the ANN with five hidden nodes. It is also remarkable that the ROC area in some cases increases when using fewer features than the original eight.

The decision support system was tested at Rigshospitalet in Copenhagen in August 2004 with a few new patients and the result was positive. Considering all this, the prospects of expanding the decision support system to other types of cancer besides lung cancer are very good.

## 7.2 Improvement of Methods and Algorithms

There is a lot of room for further development of the methods and algorithms in this thesis. To be able to search for tumours and metastases in the whole body and not only in the thorax, it is necessary to find ways to segment other parts of the body besides the lungs. This is a more challenging problem since the other organs, liver, kidneys, intestines and so on, have an intensity that is more equal to flesh, muscles and blood in CT images.

For the general orientation in the images, it could be a good idea to segment the bone tissue in the CT image. Bone is the type of tissue that appears with the highest intensity in CT images. With the whole skeleton segmented, relevant landmarks could be extracted and a splitting into body region would be possible. For example the region between the lower end of the thorax and upper pelvis where the skeleton basically just consists of the spine, could give the boundaries of the abdomen.

In the segmentation of the lung region, it would be useful to extract the windpipe and separate it from the lung region. This would make it possible to locate tumours in the neck.

In the segmentation of hot spots a lot of further research is necessary to find the right threshold for every case. It would probably be an advantage to make a region based thresholding of the PET image. For example a lower threshold value could be applied for the lung region, since the intensity of healthy lung tissue in PET images is much lower than other types of tissue in the body.

New features are needed if the decision support system will be designed to make a diagnosis of the whole body. There are certain places in the body where it is more likely to find tumours or metastases. If these regions could be found and segmented, the properties of the regions, e.g. maximal SUV and number of hot spots, are good features.

## 7.3 Choosing Programming Language and Implementation

A Matlab application is very slow compared to an application implemented in for example C, C++, C# or Java. Matlab is a suitable environment for research and development of new methods but a finished computer program needs to be converted to another programming language. Besides the time consumption, the possibilities of making graphical user interfaces are also quite limited in Matlab.

## 7.4 Graphical User Interface

The graphical user interface, presented in Appendix A, should ideally include a function for adding and removing hot spots manually. For example, if the automatic discrimination of myocardial uptake misses something that the user thinks is such, it should be possible to manually change the pre-classification of that hot spot and thereafter make a diagnosis with the manual changes. A 3-D visualisation of the segmented hot spots would help the physician to determine the shape and extension of tumours.

# A The User Interface

For a physician to take advantage of the results of the decision support system an interface is needed. In the interface the physician should be able to examine the patient in the same way as at an ordinary workstation. The physician is able to retrieve information about each hot spot and look at them in all three views. A short summary of what the interface contains follows. The program runs as a Matlab GUI application, and with minor adjustments it should be able to run as a stand alone program in Windows.

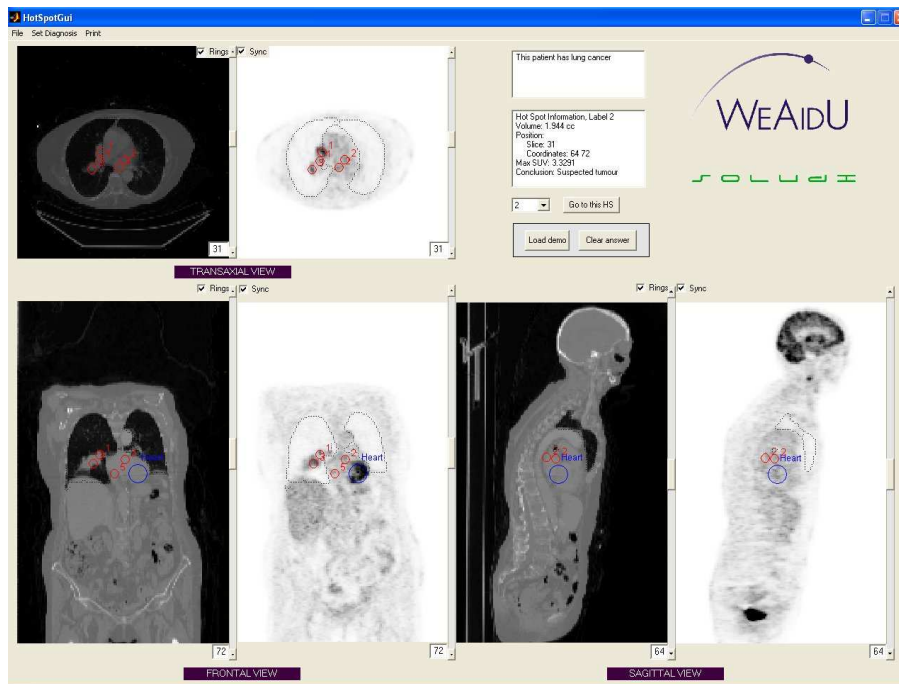


Figure A.1: *The user interface running. This patient has been diagnosed with lung cancer.*

## A.1 Views in the Interface

The interface contains PET and CT views in three dimensions, transaxial, frontal and sagittal. It is important to quickly get an overview of the current location in the body. The user is able to synchronise the movement of the PET and CT by clicking at each views sync-button in the upper left corner of the PET image. After the diagnosis is made the user is able to check the rings-button. If this button is checked the rings highlighting the hot spots are activated. In some images there can be quite a lot of rings which makes it hard to see all the details, this is why the rings are possible to switch on and off. If the rings are turned on, the lung boundaries are also displayed. In the right lower corner of all the images, the current slice number is displayed.

## A.2 Set Diagnosis

When pressing the menu button "Set diagnosis" all scripts involving localisation and classification of hot spots are executed. The process to make the diagnosis takes around ten seconds, after all scripts are finished the result is displayed in the diagnosis box above the sagittal view.

## A.3 Functions

To the right (see Figure A.1), just above the sagittal view, the user is able to retrieve all the necessary information about individual hot spots. Volume, location and SUV for the hot spots are displayed. Next to the window there is a button called "Go to this HS". If this button is pressed, the view is moved so the centre of the hot spot is viewed in all views. This functions helps overlooking all suspected areas very fast.

# Bibliography

- [1] R. Balasubramanian. *Topics in computer vision*.  
[http : //www.cis.umassd.edu/~rbalasubrama/home/index.html](http://www.cis.umassd.edu/~rbalasubrama/home/index.html), August 2004.
- [2] M. Bedford And M. N. Maisey. *Requirements for Clinical PET: Comparisons within Europe*, In *European Journal of Nuclear Medicine and Molecular Imaging*, Vol. 31, No. 2, February 2004.
- [3] CM. Bishop. *Neural Networks for pattern recognition*. Oxford University Press, 1995.
- [4] Christian Cohade, Medhat Osman, Laura T. Marshall and Richard L. Wahl. *PET-CT: accuracy of PET and CT spatial registration of lung lesions*. In *European Journal of Nuclear Medicine and Molecular Imaging*, Vol. 30, No. 5, May 2003.
- [5] N. Cristianini And J. Shawe-Taylor. *An Introduction to Support Vector Machines*. Cambridge University Press, 2000.
- [6] SS. Cross, RF. Harisson And RL. Kennedy. *Introduction to neural networks*. *Lancet* 1995;346:1075-1079.
- [7] Deborah Heart and Lung Center. *abc of lung cancer*.  
[http : //www.abc-lung-cancer.com/](http://www.abc-lung-cancer.com/), September 2004.
- [8] The DICOM Standards Committee. *DICOM homepage*.  
[http : //medical.nema.org/](http://medical.nema.org/), September 2004.
- [9] A. Ericsson. *Automatic Shape Modelling and Applications in Medical Imaging*. Mathematics LTH, Lund University, 2003.
- [10] Rafael C. Gonzalez And Richard E. Woods. *Digital Image Processing*. Prentice Hall, 2002.
- [11] Holger Holst. *New Methods for Automated Interpretation of Electrocardiograms and Lung Scintigrams*. Department of Clinical Physiology, Lund University, 2000.
- [12] S. Hu, E. A. Hoffman And J. M. Reinhardt. *Automatic Lung Segmentation for Accurate Quantitation of Volumetric X-Ray CT Images*. In *IEEE Transactions on Medical Imaging*, Vol. 20, No. 6, June 2001.

## BIBLIOGRAPHY

---

- [13] Henry Sung-Cheng Huang. *Anatomy of SUV*. In *Nuclear Medicine & Biology*, Vol. 27, 2000.
- [14] Internetmedicin i Göteborg AB. *Internetmedicin*.  
<http://www.internetmedicin.se/>, Septemer 2004.
- [15] Andreas Järund. *Lung Segments in Classification of Embolism*. Centre for Mathematical Sciences, Lund University, 2000.
- [16] Junshui Ma, Yi Zhao And Stanley Ahalt. *OSU SVM Classifier Matlab Toolbox (ver 3.00)*.  
[http://www.ece.osu.edu/~maj/osu\\_svm/](http://www.ece.osu.edu/~maj/osu_svm/), August 2004.
- [17] Mathworks *Matlab Reference*.  
<http://www.mathworks.com/support/>, August 2004.
- [18] Regina Pohle and Klaus D. Toennies. *Segmentation of Medical Images Using Adaptive Region Growing*. Department of Simulation and Graphics, Otto-von-Guericke University Magdeburg.
- [19] F. Ponzio, H. Zhuang, C. Woodfield And A. Alavi. *The Heart or a Tumor? Lung Cancer in the Lower thorax Can Mimic Myocardial FDG Uptake on PET Imaging.*, In *Clinical Nuclear Medicine*, Vol. 20, No. 2, 2003.
- [20] H. Schöder, Y. E. Erdi, S. M. Larson And H. W. D. Yeung. *PET/CT: a New Imaging Technology In Nuclear medicine*. In *European Journal of Nuclear Medicine and Molecular Imaging*, Vol. 30, No. 10, October 2003.
- [21] Peter E. Valk, Dale L. Bailey, David W. Townsend and Michael N. Maisey. *Positron Emission Tomography*. Springer-Verlag London Limited, 2003.
- [22] Vladimir N. Vapnik. *The Nature of Statistical Learning Theory*. Springer, 1995.
- [23] H. Vesselle, E. Turcotte, L. Wiens And D. Haynor. *Application of a Neural Network to Improve Nodal Staging Accuracy with 18F-FDG PET in Non-small Cell Lung Cancer*. In *The Journal of Nuclear Medicine*, Vol. 44, No. 12, December 2003.