

VRIJE UNIVERSITEIT AMSTERDAM

RESEARCH PAPER BA

---

# Lung nodules detection in CT scans

---

*Author:*  
Chi Chun Wan

*Supervisors:*  
Dr. Mark Hoogendoorn  
Dr. Evert Haasdijk

April 4, 2017



# Contents

<b>Preface</b>	<b>2</b>
<b>Abstract</b>	<b>3</b>
<b>1 Introduction</b>	<b>4</b>
<b>2 Related Work</b>	<b>5</b>
<b>3 Data and Methodology</b>	<b>6</b>
3.1 Image Acquisition (Data) . . . . .	7
3.2 Image Pre-Processing . . . . .	8
3.3 Image Segmentation . . . . .	8
3.4 Features Extraction . . . . .	10
3.5 Classification . . . . .	15
3.5.1 Target Creation . . . . .	16
3.5.2 Models . . . . .	16
3.5.3 Unbalanced Dataset Handling . . . . .	19
<b>4 Experiments</b>	<b>20</b>
4.1 Evaluation methods . . . . .	20
4.2 Experimental setup . . . . .	21
<b>5 Results</b>	<b>22</b>
5.1 Parameter settings . . . . .	22
5.2 Features Importance . . . . .	24
5.3 Model Comparison . . . . .	26
<b>6 Discussion</b>	<b>29</b>
<b>7 Conclusion</b>	<b>30</b>

## Preface

This research paper was written as a compulsory part of the Master's program Business Analytics at the VU Amsterdam. The objective is to perform a research and write a paper that demonstrate the student's ability to describe a problem in a clear manner for the benefit of an expert manager.

I would like to thank my supervisors Dr. Mark Hoogendoorn and Dr. Evert Haasdijk for introducing me this interesting topic and for guiding me during this research.

## Abstract

Lung cancer is one of the four most common cancers in the world. Early detection and diagnosis will increase the survival rate. However, detection of early stage lung cancer in computed tomography (CT) scans is challenging and time-consuming. Radiologists will experience pressure and heavy workload considering the large number of scans they have to analyze on a daily basis. Computer aided detection (CAD) systems that automatically detect and localize lung nodules in CT scans can assist radiologists by offering a useful second opinion. This will speed up the whole process. However, a major problem in these CAD systems is the large number of false positives. In this paper two CAD systems are developed using 2D methods that reduce computational complexity and computational cost. The aim was to achieve a low number of false positives while maintaining a high sensitivity. Two different machine learning models are applied in the developed CAD systems. These models classify nodule candidates into either nodule or non-nodule. The first model is a Support Vector Machine (SVM) and the second model is a Multi-Layer Perceptron (MLP). Both models have shown good performance in classifying nodules that are not attached to the thoracic wall (non-juxtapleural nodules) but bad performance in classifying nodules that are attached to the thoracic wall (juxtapleural nodules). SVM achieved 93% sensitivity, 86% specificity and 87% precision for non-juxtapleural nodules and 45% sensitivity, 70% specificity and 60% precision for juxtapleural nodules. MLP achieved 86% sensitivity, 86% specificity and 86% precision for non-juxtapleural nodules and 8% sensitivity, 95% specificity and 41% precision for juxtapleural nodules. The developed CAD systems require further development in juxtapleural nodules detection.

# 1 Introduction

Lung cancer is one of the four most common cancers worldwide besides female breast, bowel and prostate cancer [1]. Moreover, the mortality rate has increased over the years [2]. Survival from lung cancer is directly related to detection of suspicious nodules at early stages. The lung cancer survival statistics for the Netherlands in the period 2010-2012 [3] shows that the survival rate is significantly higher if tumors are detected at an early stage.

Radiologists nowadays use chest computed tomography (CT) scans to detect lung tumors as it has a high sensitivity and low error rate. However, even with these CT scans, it requires a certain amount of time and experience to detect and label the lung tumors. This especially holds in early stages where lung tumors are still small (1-2mm). Moreover, considering the large number of cases that radiologists have to analyse on daily basis, they will experience pressure and a heavy workload.

A solution to reduce this pressure and workload is to use computer aided detection (CAD) systems that automatically detect and localize lung nodules in CT images. These systems are helpful to assist the radiologists in the process of lung tumors detection. They have many benefits such as reducing the error rate of nodule detection, reducing the operation time and detecting tumors that are overlooked by the radiologists. Several studies have shown that CAD systems offer a useful second opinion [4, 5].

However, current CAD systems still produce many false positives (predicting non-nodules as nodules) while the sensitivity is large (80%-90%) [6, 7]. Researchers have applied 3D methods that process the entire 3D nodule volume to reduce the false positives [8, 9, 10]. However, 3D methods have higher computational cost and are operational more complex than 2D methods. Therefore, the aim of this paper is to develop a CAD system for lung nodules detection using 2D methods that is able to achieve a low number of false positives while maintaining a high sensitivity.

The structure of this paper has been divided into six parts. The first part is a brief literature review. The second part provides a description of the data and methodology used in the CAD system. The third part shows the experimental setup applied to the CAD system. The fourth part presents the results of the experimental setup. The fifth part is the discussion on the

results and the final part is the conclusion for the developed CAD system.

## 2 Related Work

Computer Aided Detection (CAD) systems used in lung nodules detection generally consists of four main stages [11]: preprocessing, segmentation, structure/ROI (Region of Interest) analysis and classification.

The preprocessing stage is the process of improving the quality of the lung image. The most common processes are reduction of noise and artifact (bugs in image). Kim et al. [12] implemented smoothing to reduce noise through median filtering. S.Sivakumar et al. [13] and Jaesung et al. [14] also used this method to remove the noise from the image. Pu et al. [15], Gori et al. [16] and Wei et al. [17] implemented Gaussian smoothing to eliminate the image artifacts. Artifacts such as removing contour along the lung boundary that is likely not the lung boundary.

After the preprocessing stage, lung segmentation is applied. This process separates the lung lobe region from other tissues in the image by keeping the lung lobe region and removing the rest. Applying this process is important as it increases the accuracy and precision of nodule detection. Moreover, it decreases the computational cost of detection. A simple segmentation technique is thresholding. This technique converts a gray-scale image into a binary image using a threshold where pixels greater than the threshold are considered to be foreground and all other pixels are considered to be background [18, 19]. Another segmentation technique is region growing used by Aggarwal et al. [20] and Taher et al. [21] for lung tissue segmentation. This technique determines object boundaries based on the homogeneity of the image. It is an iterative process where neighboring pixels of initial seed points are explored and determined whether they should be added to the region.

Next, the stage of structure/ROI (Region of Interest) analysis is performed. In this stage the characteristics of the regions are extracted. Characteristics such as intensity values of pixels and morphological and texture analysis are used for detection of lung nodules. Kim et al. [22] extracted features such as shape, size, average and standard deviation of the nodule candidates. Ozekes,

S. and Osman, O. [23] implemented 3D feature extraction. These 3D features consist of 3D connected component labeling, straightness calculation, thickness calculation, determining the middle slice, vertical and horizontal widths calculation, regularity calculation and calculation of vertical and horizontal black pixel ratios. All the characteristics of nodule candidates are then used in the classification stage.

In the classification stage the nodule candidates are classified into nodules or non-nodules using a classifier. Different classifiers have been used: an automated rule-based classifier proposed by Li et al. [24], artificial neural networks (ANN) used by Arimura et al. [25], Bayesian classifier used by McCulloch et al. [26] and Support Vector Machine (SVM) used by Boroczky et al. [27].

However, these four main stages are not always applied. Li et al. [28] proposed a deep convolutional neural network for pulmonary nodule classification. This method replaces the two components of feature extraction and classification because the input of deep convolutional neural networks is ROI pixel data directly without feature extraction and selection. However, in this paper these four main stages are applied in the developed CAD system.

### **3 Data and Methodology**

In this section, the data and methodology for the developed CAD systems are described. The architecture of the stages of these systems are illustrated in Figure 1. Details of each of these stages will be explained in the following subsections.

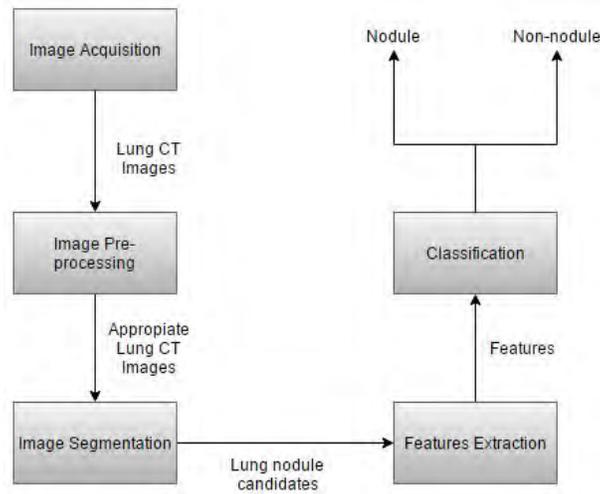


Figure 1: The architecture of the stages of the system.

### 3.1 Image Acquisition (Data)

Image acquisition refers to the process of acquiring lung CT images. These images can be found in public and private databases. However, private databases are not accessible for everyone which made the reuse of data for research impossible. Therefore, public databases are more preferred for research.

The Lung Image Database Consortium image collection (LIDC-IDRI) public database [29] is used to obtain lung CTs. This database contains diagnostic and lung cancer screening thoracic CT scans with marked-up annotated lesions of 1018 patients. Each folder includes DICOM images from a clinical thoracic CT scan and an associated XML file. The XML file records the results of a two-phase image annotation process performed by four experienced thoracic radiologists. In the first phase each radiologists independently analyzed each CT scan and marked lesions to one of three categories ("nodule  $\geq$  3 mm," "nodule  $<$  3 mm," and "non-nodule  $\geq$  3 mm"). In the second phase each radiologist independently analyzed their own marks along with the anonymized marks of the three other radiologists to render a final opinion.

### 3.2 Image Pre-Processing

Image pre-processing has been performed to construct appropriate images that are used for developing the CAD system.

The pre-process consists of two processes. The first process is the process of converting the DICOM images into TIF format for easier processing. The second process is the process of converting the results of the individual annotations in the XML files to binary images. These binary images represent the marked-up annotated lesions of each of the four radiologists for each slice. Since we want to obtain a good overview of the location of the annotation without losing any information of each radiologist, the four individual annotations are combined into one single annotation. This is accomplished by taking the union of these individual annotations. An example of this process is illustrated in Figure 2.

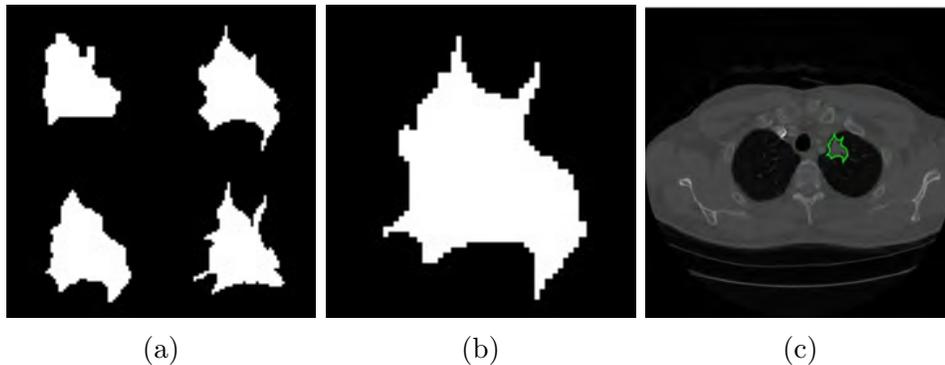


Figure 2: Image Pre-processing: (a) shows the the four marked-up annotated lesions from the four radiologists, (b) shows the united annotation and (c) shows the united annotation presented in the CT scan.

### 3.3 Image Segmentation

In this stage, the pre-processed CT scans are segmented in order to extract nodule candidates. The image segmentation method used is thresholding. This method uses a threshold to partition an image into a foreground and background, where pixels below this threshold are transformed into black pixels (background) and pixels above or equal to this threshold are transformed

into white pixels (foreground). Thus, it isolates objects by converting gray-scale images into binary images. The threshold is computed using the Otsu's method [30]. This method chooses the threshold value that minimizes the intra-class variance (the variance within the class). The intra-class variance is defined as a weighted sum of variances of the two classes (foreground and background):

$$\sigma_w^2(t) = \omega_b(t)\sigma_b^2(t) + \omega_f(t)\sigma_f^2(t) \quad (1)$$

Weights  $\omega_b$  and  $\omega_f$  are the class probabilities for a threshold  $t$  and  $\sigma_b$  and  $\sigma_f$  are the variance of the two classes. Denote the range of intensity levels as  $[0, L-1]$ , then the weight probabilities for the two classes are given by

$$\begin{aligned} \omega_b(t) &= \sum_{i=0}^{t-1} p(i) \\ \omega_f(t) &= \sum_{i=t}^{L-1} p(i) \end{aligned} \quad (2)$$

After the segmentation, a segmented image of the CT scan is obtained where the white pixels represent objects (foreground) such as nodules, lung tissues, blood vessels and thoracic wall and the black pixels represent air (background). An example is illustrated in Figure 3.

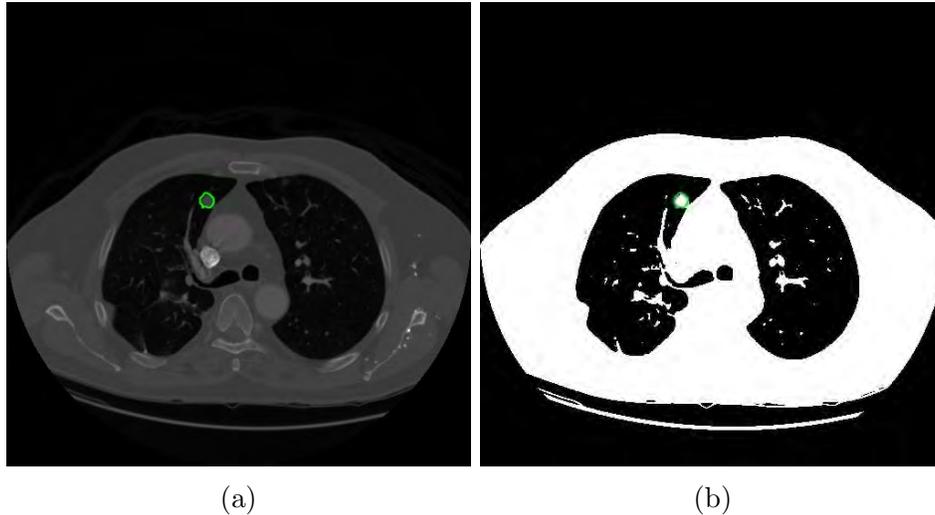


Figure 3: Image Segmentation: (a) shows the CT scan before segmentation with the annotated nodule in green and (b) shows the CT scan after segmentation where nodule candidates are presented in white pixels with the annotated nodule in green.

### 3.4 Features Extraction

In this stage, features for each connected component (nodule candidate) in the segmented CT scan image are extracted. These features give information about the nodule candidate and they are divided into three categories: 2D-shape based features, 3D-shape based features and texture based features. 2D-shape based features and texture based features are extracted from a single slice in a CT scan whereas features in 3D are extracted from multiple slices in a CT scan. Since nodule candidates are 3D objects, it is relevant to use 3D features.

#### 2D-shape based features

The shape based features are physical dimensional measures that characterize the presence of a nodule candidate. The extracted 2D-shape based features are: area, centroid, eccentricity, equivalent diameter, major axis length, minor axis length and perimeter. These features excluding centroid are the

basic characteristics of geometric features and essential to recognize objects in medical diagnosis [31]. The centroid is only used for creating the 3D-shape based features and the target. The creation of 3D-shape based features will be explained in detail in the next section and the creation of the target will be explained in detail in section 3.5.1.

The extracted 2D-shape based features [32] are defined as follows:

1. *Area*: The actual number of pixels of the nodule candidate. It is obtained by the summation of areas of white pixels in the binary image.
2. *Centroid*: The center of mass of the nodule candidate specified in coordinate (x,y).
3. *Eccentricity*: The ratio of the distance between the foci of the ellipse and its major axis length. The value is between 0 and 1, where an object with eccentricity 0 represents a circle and 1 represents a line segment.
4. *Equivalent Diameter*: The diameter of the nodule candidate computed as  $\sqrt{\frac{4*Area}{\pi}}$ .
5. *Major Axis Length*: The length (in pixels) of the major axis of the nodule candidate.
6. *Minor Axis Length*: The length (in pixels) of the minor axis of the nodule candidate.
7. *Perimeter*: The distance around the boundary of the nodule candidate. The perimeter P is measured as the sum of the distances between every consecutive boundary points. In mathematical form,

$$P = \sum_{i=1}^{n-1} |B_i B_{i+1}| + |B_n B_1| \quad (3)$$

where,  $B = B_1, \dots, B_n$  is the set of boundary points.

### 3D-shape based features

After the 2D-shape based features are derived, the 3D-shape based features from a single CT scan are extracted. First, the nodule candidates that corre-

spond with each other in each slice of this scan have to be matched. Hereby the 2D-features area for the thoracic wall and centroid for the other candidates are used. Note that the thoracic wall is a possible nodule candidate because nodules can be attached to this wall. These nodules are called juxtapleural nodules. An example of a juxtapleural nodule is illustrated in Figure 4.

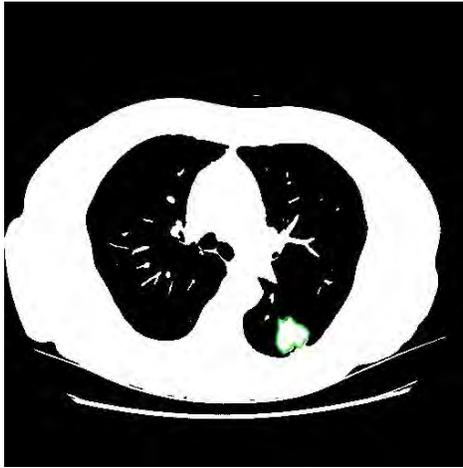


Figure 4: Example of a juxtapleural nodule annotated in green.

The thoracic wall is a nodule candidate with the largest area, so the area can be used as measurement to obtain the thoracic wall in each slice. However, this measurement cannot be used for the other nodule candidates that are not attached to the thoracic wall (non-juxtapleural nodules) as some of these candidates have the same area. Therefore, the centroid is used as measurement because the same nodule candidates in a single CT scan are likely to have centroids close to each other in the slices. Thus, the euclidean distances between the centroid of a nodule candidate  $X$  in a slice and the other nodule candidates in each of the other slices are computed. Then the nodule candidate in each of the other slices with the smallest euclidean distance is high likely representing the same nodule candidate  $X$ . However, this method is not foolproof. Problem occurs when the same nodule candidate  $X$  is not visible in a slice which results in a false match. A solution is to use a threshold  $\alpha$ . When the smallest euclidean distance is smaller or equal to this threshold  $\alpha$ , then the nodule candidate is assigned as a correct match and the nodule

candidate of a slice will be matched with the nodule candidate  $X$ . Otherwise, it is assigned as a false match and the nodule candidates will not be matched. In the end, all the nodule candidates in each slice of the single CT scan are matched and the 3D-shape based features of these matching nodule candidates are derived. The following 3D-shape based features are extracted:

1. *Area mean*: The average area of the matched nodule candidates. In mathematical form:

$$\mu_{Area} = \frac{1}{N} \sum_{i=1}^N Area_i \quad (4)$$

where,  $Area_i$  is the area of the nodule candidate in slice  $i$  with  $N$  the total number of slices in the single CT scan.

2. *Perimeter mean*: The average perimeter of the matched nodule candidates. In mathematical form:

$$\mu_{Perimeter} = \frac{1}{N} \sum_{i=1}^N Perimeter_i \quad (5)$$

where,  $Perimeter_i$  is the perimeter of the nodule candidate in slice  $i$ .

3. *Area variance*: The variance of the area of the matched nodule candidates. In mathematical form:

$$Var_{Area} = \frac{1}{N-1} \sum_{i=1}^N |Area_i - \mu_{Area}|^2 \quad (6)$$

4. *Perimeter variance*: The variance of the perimeter of the matched nodule candidates. In mathematical form:

$$Var_{Perimeter} = \frac{1}{N-1} \sum_{i=1}^N |Perimeter_i - \mu_{Perimeter}|^2 \quad (7)$$

These features give information about the change of the nodule candidate's shape over the slices. The shape of lung nodules tends to be more changing over the slices whereas the shape of blood vessels stays roughly the same over the slices. This is because the original position of a cancer nodule is located at a specific section in the lung. Thus, the cancer nodule is the most visible

in the slice that represents this specific section whereas it becomes less visible in slices that are further from this section. An overview of a blood vessel in a CT scan is illustrated in Figure 5 and an overview of a nodule in a CT scan is illustrated in Figure 6.

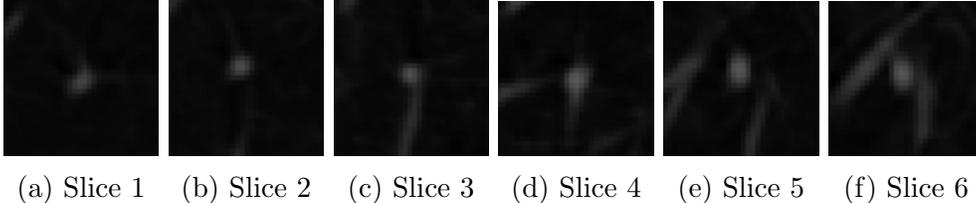


Figure 5: A blood vessel in a CT scan. Note that the structure of the blood vessel is roughly the same on each slice. This indicates that the object is a blood vessel running to the slices.

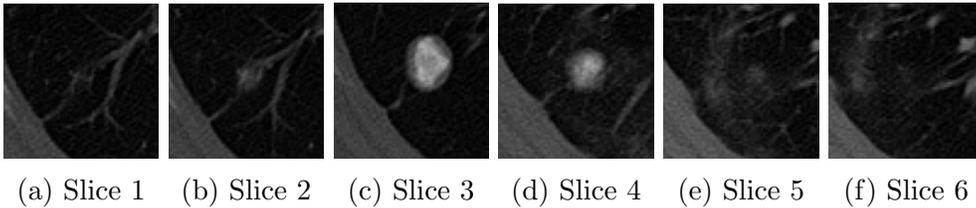


Figure 6: A nodule in a CT scan. Note that the structure of the nodule starts to appear in slice 2 and gets larger, then start to diminish after slice 5.

### Texture based features

Next, the texture based features of the nodule candidates are extracted using Gray level co-occurrence matrix (GLCM). GLCM is a second order statistical measure that is introduced by Haralick [33]. It investigates texture that considers the spatial relationship with pixels by extracting statistical measures (texture based features) [34]. The extracted texture based features for a GLCM with  $m$  rows and  $n$  columns are:

1. *Contrast*: The local variations in the GLCM. It computes the intensity contrast between a pixel and its neighbor pixel for the whole image. In

mathematical form:

$$Contrast = \sum_{i=1}^m \sum_{j=1}^n (i - j)^2 * p(i, j) \quad (8)$$

where,  $p(i, j)$  is the pixel of the GLCM at location (i,j).

2. *Correlation*: The joint probability occurrence of the specified pixel pairs. In mathematical form:

$$Correlation = \sum_{i=1}^m \sum_{j=1}^n \frac{(i - \mu_i)(j - \mu_j)p(i, j)}{\sigma_i \sigma_j} \quad (9)$$

where,  $\mu_i, \mu_j$  and  $\sigma_i, \sigma_j$  are the mean and standard deviations of GLCM along row wise i and column wise j.

3. *Energy*: The textural uniformity. In mathematical form:

$$Energy = \sum_{i=1}^m \sum_{j=1}^n (p(i, j))^2 \quad (10)$$

4. *Homogeneity*: The closeness of gray levels in the spatial distribution over image. In mathematical form:

$$Homogeneity = \sum_{i=1}^m \sum_{j=1}^n \frac{p(i, j)}{1 + |i - j|} \quad (11)$$

These features have shown to be useful in cancer detection [35, 36].

### 3.5 Classification

In this stage, the extracted nodule candidates in each slice are classified as being a nodule or a non-nodule. The models that are used for this classification are Support Vector Machine (SVM) and Multi-Layer Perceptron (MLP). These models are explained in detail in section 3.5.2.

### 3.5.1 Target Creation

Before the models are able to classify the nodule candidates, they have to be trained and the training requires a target output. Therefore, a target variable was created to identify whether a nodule candidate is in fact a nodule or not. In order to identify this, the united annotation was used as identifier. The nodule candidate that has the closest centroid to the unit annotation was indicated as a nodule. However, to be confident that this nodule candidate was the annotated nodule, we took the intersection of the set of pixels location of this nodule candidate and the set of pixels location of the annotation. These pixels location are presented with linear indices. Next, the fraction of intersection similarity was computed and this is given by:

$$Intersection_{sim} = \frac{\text{Number of intersected pixels}}{\text{Total number of annotation pixels}} \quad (12)$$

We assumed that the fraction of intersection similarity must be at least 0.8 in order to be confident that the nodule candidate is in fact the annotated nodule. The choice of a minimum fraction of 0.8 fraction instead of 1.0 is because the annotation is a union of the annotations of the four radiologist which could have some deviation with the nodule candidate. However, this deviation is small and does not have intersection similarity below 0.8. The advantage of using the united annotation is that the representation of the nodule in the annotation is guaranteed. The disadvantage is that the united annotation does not always represent the precise boundary of the nodule. However, this is not a problem as radiologists still have to evaluate the detected nodule by the CAD system. Next, the target variable is created where nodule candidates that are in fact nodules are assigned with value 1 and nodule candidates that are non-nodules are assigned with value -1.

### 3.5.2 Models

#### Support Vector Machine

Support Vector Machine (SVM) is a popular data classification method. It is a supervised machine learning algorithm and takes a set of input data

(features of nodule candidates) with corresponding output data (target variable) and predicts the two possible classes (nodule or non-nodule) for each given nodule candidate. These classes are separated by a hyperplane. SVM seeks the hyperplane that maximizes the distance between each class and the hyperplane. This hyperplane is known as a maximum-margin hyperplane, where margin is defined as the sum of the minimum distances from each set to the hyperplane. Figure 7 illustrates a maximum-margin hyperplane. Besides the regular linear classification, SVM can perform non-linear classification using a kernel function. This function maps the input into high-dimensional features spaces which makes it linear separable. In this paper, three different kernel functions are used: the linear kernel function, the radial basis function (RBF) and the multilayer perceptron kernel function (MLP).

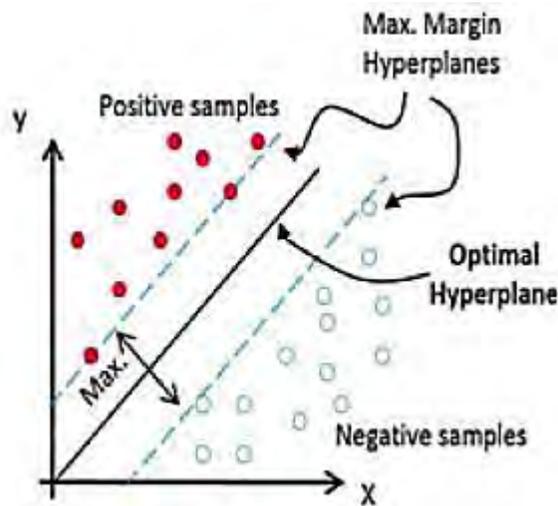


Figure 7: Maximum-margin hyperplane

## Multi-Layer Perceptron

The Multi-Layer Perceptron (MLP) is an Artificial Neural Network (ANN). ANN is a computational model which structure and performance characteristics are similar to those of the biological neural networks and is able to learn complex transformation of inputs to certain outputs. The architecture of a MLP is illustrated in Figure 8.

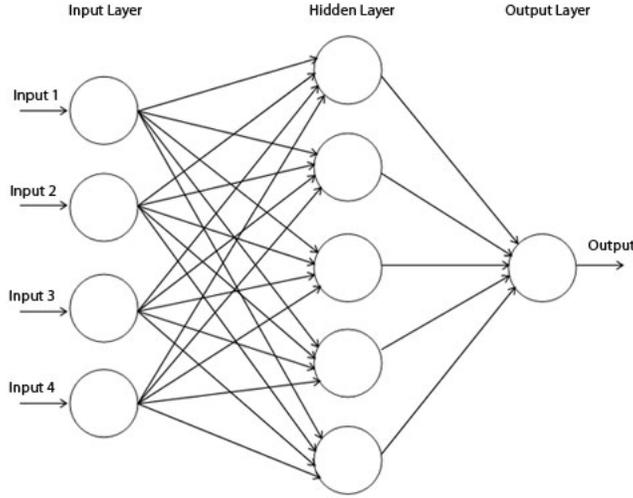


Figure 8: Architecture of a MLP with a minimum of 3 layers (input, hidden and output). The input later has 4 neurons, the hidden layers has 5 neurons and the output layer has 1 neuron. Note that the MLP can contain multiple hidden layers.

MLP consists of processing elements called neurons. These neurons are used for the transformation from input to output and they are organized in layers. Between these layers neurons are connected to each other, each with an associated weight. The connections are always directed from lower layers to upper layers. The output of each neuron  $c$  in the hidden layer(s) and output layer is then a (non)linear function over the dot product of the weights of the connections with the outputs of the neurons in the previous layer. In mathematical form:

$$c = \phi \left( \sum_i w_i a_i + b \right), \quad (13)$$

where  $a_i$  are the inputs of the neuron,  $w_i$  are the weights of the neuron and  $b$  is the bias.  $\phi$  is the (non)linear function, also called the activation function as it determines the activation level of the neuron.

Next, the network is trained using a learning algorithm that updates the weights such that the error between the value of the output neuron and the

target value is minimized. Scaled Conjugate Gradient (SCG) learning algorithm is used which is an iterative algorithm that searches along conjugate directions. This algorithm is better and faster than the normal gradient descent [37].

In the network for the classification of nodule candidates, the input neurons in the input layer representing the features of nodule candidates are first transformed using the logistic activation function. This transformation maps the input data into a space where it becomes linearly separable represented in the hidden layer(s). Next, the hidden layer to output layer transform the linear separable inputs into a probability output value using again the logistic activation function. This probability output value represents the probability of being a nodule.

### 3.5.3 Unbalanced Dataset Handling

The unbalanced dataset problem occurs in lung nodule classification as the number of non-nodule samples is much larger than nodule samples. In our dataset, the proportion of nodule classes is extremely low (0.58%). Training on such an unbalanced dataset will result in a model that is unable to predict the nodule class as the model is more biased towards the majority class. In order to combat this problem, the dataset is re-sampled by under-sampling the majority class. The majority class (non-nodule) samples are randomly removed from the dataset until the proportion between the majority class and the minor class is equal. The choice for a fully balanced dataset is made to ensure that the models will not be affected by any factor of unbalancing. However, randomly removing the majority class samples can cause unbalancing in the nodule type samples (juxtaleural nodules and non-juxtaleural nodules). Therefore, the dataset is first divided into two datasets. The first dataset contains only juxtaleural candidates and the second dataset contains only non-juxtaleural candidates. In these datasets 37.77% of the samples are juxtaleural nodules and 62.23% of the samples are non-juxtaleural nodules. Next, the non-nodules samples are undersampled in both datasets until the proportion between majority class and minor class are equal. Thus, an equal proportion in juxtaleural non-nodule samples and juxtaleural nodule samples in the first dataset and an equal proportion in non-juxtaleural non-nodule samples and non-juxtaleural nodule samples

in the second dataset. Next, these two balanced datasets are aggregated into one dataset that contains 20704 samples from 874 patients. This balanced dataset is used for training and testing the models.

## 4 Experiments

In this section the evaluation methods and the experimental setup for the models are described.

### 4.1 Evaluation methods

To obtain the performance of the models, a validation method and an evaluation metric are needed. The 5-fold cross-validation validation method is used. This validation method randomly partitions the patients in the dataset into 5 equally folds. Each time, one of these 5 folds is used for testing and the other 4 folds are used for training until all the folds have been tested once. With k-fold cross-validation it shows how well the models perform on data from new patients.

The confusion matrix is used as evaluation metric. The terminology of the confusion matrix is as follows:

1. *True Positive (TP)*: Correctly predicting a nodule as a nodule.
2. *True Negative (TN)*: Correctly predicting a non-nodule as a non-nodule.
3. *False Positive (FP)*: Predicting a non-nodule as nodule.
4. *False Negative (FN)*: Predicting a nodule as non-nodule.

From this confusion matrix, three performance measures are computed:

1. *Precision*: The proportion of correct predicting nodule cases. The equation is given by:

$$Precision = \frac{TP}{TP + FP} \tag{14}$$

2. *Sensitivity*: The proportion of nodule cases that were correctly classified. The equation is given by:

$$Sensitivity = \frac{TP}{TP + FN} \quad (15)$$

3. *Specificity*: The proportion of non-nodule cases that were correctly classified. The equation is given by:

$$Specificity = \frac{TN}{FP + TN} \quad (16)$$

These three performance measures are used to measure the performance of the models in the experimental setup. Sensitivity is used as primary outcome. The reason is that sensitivity is more important than specificity in CAD systems as predicting a false negative is much worse than predicting a false positive. Moreover, precision is included as it gives information about the false positive findings and the relevance of the detected nodules. This is relevant because a high false positive rate is one of a major problem in CAD systems [38].

## 4.2 Experimental setup

The dataset is split in a train set and a test set. The train set consists of 80% of the total number of patients (700) and the test set consists of the remaining 20% of patients (174). Next, three experimental setups are applied.

The first experimental setup is the parameter settings. We have experimented with different parameter combinations for both models using 5-fold cross-validation on the train set. In the Multi-Layer Perceptron model, the parameters are the number of hidden layers and the number of neurons in each layer. In the Support Vector Machine model, the parameter is the different kernel functions.

The second experimental setup is the features importance. First, the feature categories are varied in the MLP model to detect the importance of each feature category for classifying nodule candidates. Next, the importance of each individual feature is explored. This is done by removing each individual

feature as input feature for the MLP model and observe how this influences the performance of the model. Again 5-fold cross-validation on the train set is applied.

The last experimental setup is the comparison between the two models. Both models are evaluated on the test set using the optimal parameters found in the parameter settings.

## 5 Results

In this section the obtained results for the models are presented. Section 5.1 describes the result for the different parameter settings in both models. Section 5.2 describes the result for the different selected features in the MLP model and in section 5.3 the two models are compared.

### 5.1 Parameter settings

The first parameter setting is the threshold  $\alpha$  for matching the same nodule candidate in each slice of a single CT scan. This threshold  $\alpha$  is set to 5 and it is chosen based on manually observing the position of the nodule candidates in different slices.

Next, the performance for different number of hidden layers and their size (number of neurons) in each layer is investigated. 5-fold cross-validation is used on the train set and the average performance measures over these 5-folds are taken. Multiple runs have been considered. Moreover, the input features and the output threshold of the model are kept fixed during the tests to avoid influence from external factors. The result is shown in Figure 9.

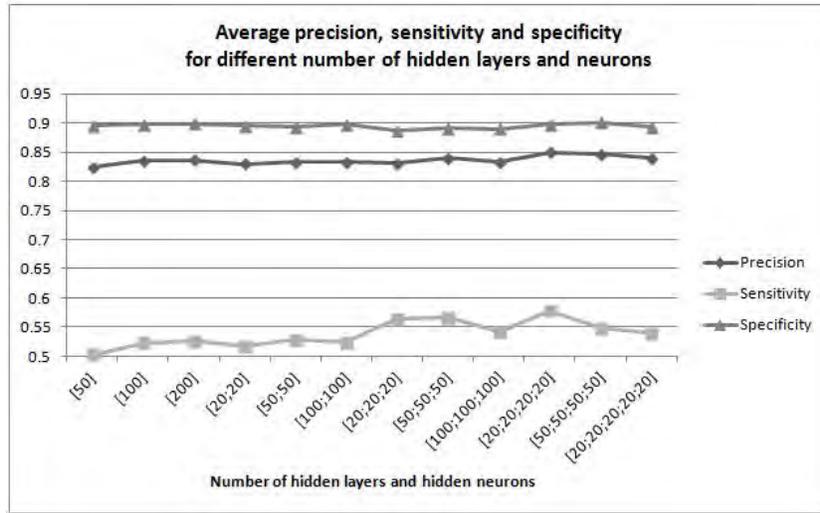


Figure 9: Average precision, sensitivity and specificity against different number of hidden layers and neurons using 5-fold cross-validation, all input features and 0.1 output threshold over 10 runs.

We observe that the model requires more than two hidden layer as the sensitivity is improving when the number of hidden layers is increased. However, increasing the number of hidden layers above 4 layers is irrelevant as the performance starts to decay and overfit. The best performance is acquired with 4 hidden layers with 20 neurons at each layer. Moreover, increasing the number of neurons in 4 hidden layers is not necessary because the performance starts to decrease. Furthermore, the precision and specificity are relatively constant over the different number of hidden layers and neurons.

Next, the performance for the different Kernel functions in SVM is explored. The average performance measures of the 5-folds in the train set over 10 runs are taken and the result is shown in Figure 10.

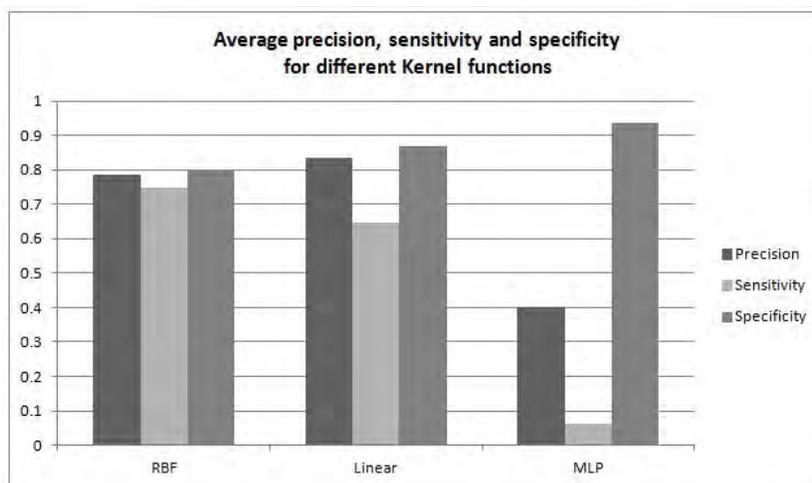


Figure 10: Average precision sensitivity and specificity for different Kernel functions using 5-fold cross-validation and all input features over 10 runs.

MLP kernel has the lowest sensitivity (0.06) of all the three kernel functions. This indicates that MLP kernel is not suitable for classifying lung nodules. RBF and linear kernel have a specificity of 0.80 and 0.87, respectively. This shows that linear kernel performs better at predicting non-nodule cases. However, the sensitivity of RBF (0.75) is larger than the linear kernel (0.64) showing that RBF is better at predicting nodule cases. Thus, we observe a trade-off between sensitivity and specificity. However, the difference between specificity is smaller than the difference in sensitivity and the difference between precision is small (0.78 for RBF and 0.83 for linear). Moreover, sensitivity is the primary outcome. Therefore, RBF is preferred.

## 5.2 Features Importance

The sensitivity for different feature categories is observed in the MLP model. The result is shown in Figure 11.

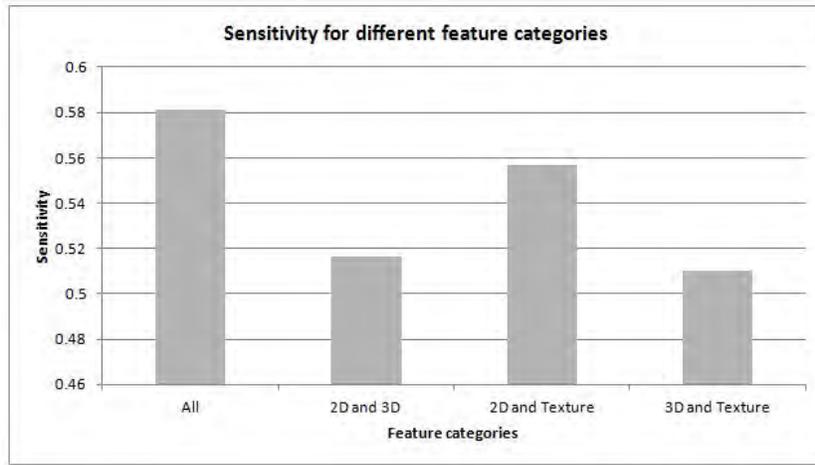


Figure 11: Sensitivity for different feature categories using MLP with 4 hidden layers and 20 neurons at each layer with 0.1 output threshold and 5-fold cross-validation over 10 runs.

We observe that all three feature categories contribute in classifying nodule candidates. However, the most important category is 2D-shape based features as the sensitivity decreases from 0.58 to 0.51 when 2D-shape based features are excluded from the model. Moreover, excluding 3D-shape based features and texture based features result in a decrease in sensitivity.

Next, each individual feature is evaluated by observing the decrease in sensitivity when this individual feature is eliminated from the model. The result is shown in Table 1.

Eliminated feature	Sensitivity decrease
Correlation	-10.76%
Perimeter	-9.13%
Equivalent Diameter	-8.32%
Area	-6.67%
Minor Axis Length	-5.75%
Eccentricity	-5.61%
Area mean	-3.98%
Homogeneity	-3.68%
Energy	-3.38%
Major Axis Length	-3.10%
Area variance	-3.05%
Perimeter variance	-2.57%
Perimeter mean	-1.96%
Contrast	-1.47%

Table 1: Sensitivity decrease for each feature using MLP with 4 hidden layers and 20 neurons at each layer with 0.1 output threshold and 5-fold cross-validation over 10 runs.

We observe that correlation is the most important feature as eliminating correlation result in a sensitivity decrease of 10.76%. Moreover, the 3D-shape based features perimeter mean and variance have a less important degree than area mean and variance. The reason could be that area gives a better indication about the change of nodule shape over the slices as shown in Figure 6. Furthermore, perimeter, equivalent diameter and area are all important 2D-shape based features.

### 5.3 Model Comparison

The SVM model and the MLP model are evaluated on the test set over 100 runs using the optimal parameters found in section 5.1. Moreover, the performance of the models in classifying juxtapleural nodule candidates (nodules attached to the thoracic wall) and non-juxtapleural nodule candidates is investigated. First, the nodule candidates are categorized into either juxtapleural nodule candidates or non-juxtapleural nodule candidates. Next, the

performance measures in the two categories are computed for both models. The result is shown in Table 2.

All nodules			
Model	Precision	Sensitivity	Specificity
SVM	0.79	0.75	0.80
MLP	0.84	0.55	0.89
Juxtapleural nodules			
Model	Precision	Sensitivity	Specificity
SVM	0.60	0.45	0.70
MLP	0.41	0.08	0.95
Non-juxtapleural nodules			
Model	Precision	Sensitivity	Specificity
SVM	0.87	0.93	0.86
MLP	0.86	0.86	0.86

Table 2: The average model performance over 100 runs on the test set: SVM with RBF kernel and MLP with 4 hidden layers, 20 neurons at each layer and 0.1 output threshold.

The choice of 0.1 output threshold in MLP is because this threshold gave the best result in terms of sensitivity without having high negative effect on specificity and precision.

We observe that SVM performs better than MLP for all the nodule categories. The reason is the bad performance in detecting juxtapleural nodules for MLP (0.08 sensitivity). This shows that MLP is unable to classify juxtapleural nodule candidates while SVM is still able to classify 45% of the juxtapleural nodule candidates correct. This result shows that both models have difficulties in detecting juxtapleural nodules. However, both models have good performance in detecting non-juxtapleural nodules with 0.93 sensitivity for SVM and 0.86 sensitivity for MLP. Moreover, the precision in both models are large, 0.87 and 0.86 for SVM and MLP, respectively. This shows that the number of false positives is low in both models. Furthermore, the specificity in both model is large. These results have shown that both models are able to make a clear distinction between nodules and non-nodules for non-juxtapleural nodule candidates.

Next, the 95% confidence interval of these test results is derived. The result is shown in Table 3.

All nodules			
Model	Precision	Sensitivity	Specificity
SVM	[0.78 0.79]	[0.74 0.75]	[0.79 0.80]
MLP	[0.80 0.89]	[0.51 0.58]	[0.88 0.90]
Juxtapleural nodules			
Model	Precision	Sensitivity	Specificity
SVM	[0.59 0.61]	[0.43 0.45]	[0.69 0.71]
MLP	[0.36 0.46]	[0.06 0.10]	[0.93 0.96]
Non-juxtapleural nodules			
Model	Precision	Sensitivity	Specificity
SVM	[0.86 0.87]	[0.93 0.94]	[0.85 0.86]
MLP	[0.81 0.90]	[0.81 0.90]	[0.85 0.86]

Table 3: 95% confidence interval of the model performance over 100 runs on the test set: SVM with RBF kernel and MLP with 4 hidden layers, 20 neurons at each layer and 0.1 output threshold.

We observe that the 95% confidence interval for SVM is smaller than MLP in all the nodule categories and performance measures. This indicates that SVM is more accurate in classifying nodule candidates than MLP. However, MLP is still accurate as the confidence intervals are relatively small.

## 6 Discussion

The aim of this paper was to develop a computer aided detection (CAD) system for lung nodules detection using 2D methods. In these systems two machine learning techniques were implemented: Support Vector Machine (SVM) and Multi-Layer Perceptron (MLP). SVM performed better than MLP, it achieved a sensitivity of 75% while MLP achieved a sensitivity of 55%. Moreover, SVM is more accurate than MLP as the 95% confidence intervals are smaller in SVM. However, the performance of only detecting non-juxtaleural nodules are good in both SVM and MLP systems. They achieved 93% and 86% sensitivity for SVM and MLP, respectively. Moreover, both systems have a low false positive rate: precision of 87% and 86% for SVM and MLP, respectively. The problem of both systems is the detection of juxtaleural nodules. SVM achieved 45% sensitivity and MLP only achieved 8% sensitivity. The reason is the Otsu segmentation method that was used. This method was unable to segment juxtaleural nodule candidates from the thoracic wall. This resulted in providing the thoracic wall as juxtaleural nodule candidates for the models which made it difficult for the models to detect juxtaleural nodules. This shows that the segmentation of nodule candidates plays a crucial role in building a good CAD system.

Besides nodule candidates segmentation, features extraction is also important as it gives the characteristics of the nodule candidates. The result has shown that 2D-shape based features, 3D-shape based features and texture based features are all important features. Texture based features are important in identifying objects in medical images [39] and shape based features are important as nodules are usually small round or oval-shaped while lung tissues are usually flat-shaped.

Since the developed CAD systems have difficulties in detecting juxtaleural nodules, we recommend future studies to focus more on this part of nodule detection. Other segmentation methods might be able to separate the nodule from the thoracic wall that could lead to better performance. Another recommendation is to apply 3D lung nodule candidate detection instead of 2D lung nodule candidate detection for juxtaleural nodules. The reason is that the lung nodule candidates become more visible in 3D which makes it easier to segment.

## 7 Conclusion

Both developed computer aided detection (CAD) systems (SVM and MLP) have shown that they were able to detect lung nodules that are not attached to the thoracic wall (non-juxtaleural nodules). These systems have achieved a low number of false positives (high precision) while having a high sensitivity for non-juxtaleural nodules detection. However, this does not hold for detecting lung nodules that are attached to the thoracic wall (juxtaleural nodules). The developed CAD systems have difficulties in detecting juxtaleural nodules. Therefore, these systems should be further developed in juxtaleural nodules detection.

Despite that CAD systems could make huge improvements in lung cancer detection in the future, these systems should not replace radiologists or be used for final interpretation. The reason is that experience and expertise are always needed. Therefore, CAD systems should always remain as second opinion for the radiologists. CAD systems are supposed to assist radiologists and not supposed to replace them.

## References

- [1] Cancer Research UK. *Let's beat cancer sooner*. URL: <http://www.cancerresearchuk.org/health-professional/cancer-statistics/worldwide-cancer#heading-Zero>.
- [2] integraal kankercentrum Nederland. *Mortality lung*. URL: [http://www.cijfersoverkanker.nl/selecties/sterfte\\_luchtwegen/img577521ba94d95?language=en](http://www.cijfersoverkanker.nl/selecties/sterfte_luchtwegen/img577521ba94d95?language=en).
- [3] integraal kankercentrum Nederland. *Survival lung*. URL: [http://www.cijfersoverkanker.nl/selecties/dataset\\_3/img587cd18186163](http://www.cijfersoverkanker.nl/selecties/dataset_3/img587cd18186163).
- [4] J. K. Lyo G. D. Rubin et al. "Pulmonary Nodules on Multi-Detector Row CT Scans: Performance Comparison of Radiologists and Computer-aided Detection". In: *Radiology* 234 (2005), pp. 274–283.
- [5] Hanai K Abe Y1 et al. "A computer-aided diagnosis (CAD) system in lung cancer screening with computed tomography". In: *ANTICANCER RESEARCH* 25 (2005), pp. 483–488.
- [6] McCulloch C. C., Kaucic R. A., Mendonca P. R. S., Walter D. J., and Avila R. S. "Model-based detection of lung nodules in computed tomography exams. Thoracic computer-aided diagnosis". In: *Acad. Radiol.* (2004), pp. 258–266.
- [7] Murphy K., van Ginneken B., Schilham A. M. R., de Hoop B. J., Gietema H. A., and Prokop M. "A large-scale evaluation of automatic pulmonary nodule detection in chest CT using local image features and k-nearest-neighbour classification". In: *Med. Image Anal.* (2009), pp. 757–770.
- [8] Ge Z., Sahiner B., Chan H. P., Cascade P. N., Bogot N., Kazerooni E. A., Wei J., and Zhou C. "Computer-aided detection of lung nodules: False positive reduction using a 3D gradient field method and 3D ellipsoid fitting". In: *Med. Phys.* (2005), pp. 2443–2454.
- [9] Filho AO., Silva AC., de Paiva AC., Nunes RA., Gattass M. "3D shape analysis to reduce false positives for lung nodule detection systems". In: *Med. Biol. Eng. Comput.* (2016).
- [10] Qi Dou, Hao Chen, Lequan Yu, Jing Qin and Pheng Ann Heng. "Multi-level Contextual 3D CNNs for False Positive Reduction in Pulmonary Nodule Detection". In: *IEEE Transactions on Biomedical Engineering* (2016).

- [11] M. Firmino, G. Angelo, H. Morais, MR. Dantas and R. Valentim. “Computer-aided detection (CADe) and diagnosis (CADx) system for lung cancer with likelihood of malignancy”. In: *Biomed Eng Online* (2016).
- [12] Kim, H., Nakashima, T., Itai, Y., et al.. “Automatic detection of ground glass opacity from the thoracic MDCT images by using density features”. In: *International Conference on Control, Automation and Systems* (2007), pp. 1274–1277.
- [13] S.Sivakumar et al. “Lung Nodule Detection Using Fuzzy Clustering and Support Vector Machines”. In: *International Journal of Engineering and Technology (IJET)* (2013), pp. 179–185.
- [14] Jaesung et al. “Segmentation of Individual Ribs from Low-dose Chest CT”. In: *Medical Imaging 2010: Computer-Aided Diagnosis* (2010),
- [15] Pu, J., Roos, J., Yi, C.A., et al. “Adaptive border marching algorithm: automatic lung segmentation on chest CT images”. In: *Comput. Med. Imaging Graph* 32 (2008), pp. 452–462.
- [16] Gori, I., Bellotti, R., Cerello, P., et al. “Lung nodule detection in screening computed tomography”. In: *IEEE Nuclear Science Symposium Conference Record* (2006).
- [17] Wei, G.Q., Fan, L., Qian, J. “Automatic detection of nodules attached to vessels in lung CT by volume projection analysis”. In: *Comput. Med. Imaging Graph*. 32 (2008), pp. 452–462.
- [18] Y.-H. Na K.-T. Bae J.-S. Kim. “Pulmonary nodules: automated detection on CT images with morphologic matching algorithm, preliminary results”. In: *Radiology* 236 (2005), pp. 286–294.
- [19] N. Birkbeck, M. Sofka, T. Kohlberger, J. Zhang, J. Wetzl, J. Kaftan and S. Kevin Zhou. “Robust Segmentation of Challenging Lungs in CT Using Multi-stage Learning and Level Set Optimization”. In: *Computational Intelligence in Biomedical Imaging* (2014), pp. 185–208.
- [20] R. Vig P. Aggarwal and H.-K. Sardana. “Semantic and Content-Based Medical Image Retrieval for Lung Cancer Diagnosis with the Inclusion of Expert Knowledge and Proven Pathology”. In: *In proc. of the IEEE second international conference on Image Information Processing ICIIP* (2013), pp. 346–351.
- [21] F. Taher and R. Sammouda. “Identification of Lung Cancer Based on Shape and Color”. In: *In proc. of the 4th International Conference* (2010).

- [22] Kim, D.Y., Kim, J.H., Noh, S.M., et al. “Pulmonary nodule detection using chest CT images”. In: *Acta. Radiol.* 44 (2003), pp. 252–257.
- [23] Ozekes, S., Osman, O. “Computerized Lung Nodule Detection Using 3D Feature Extraction and Learning Based Algorithms”. In: *Medical Systems* 34 (2010), pp. 185–193.
- [24] Li, Q., Li, F., Doi, K. “Computerized detection of lung nodules in thin-section CT images by use of selective enhancement filters and an automated rule-based classifier”. In: *Acad. Radiol.* 15 (2008), pp. 165–175.
- [25] Arimura, H., Katsuragawa, S., Suzuki, K., et al. “Computerized scheme for automated detection of lung nodules in low-dose computed tomography images for lung cancer screening”. In: *Acad. Radiol.* 11 (2004), pp. 617–629.
- [26] McCulloch, C.C., Kaucic, R.A., Mendonca, P.R., et al. “Model-based detection of lung nodules in computed tomography exams. Thoracic computer-aided diagnosis”. In: *Acad. Radiol.* 11 (2004), pp. 258–266.
- [27] Boroczky, L., Zhao, L., Lee, K.P. “Feature subset selection for improving the performance of false positive reduction in lung nodule CAD”. In: *IEEE Trans. Inf. Technol. Biomed.* 10 (2006), pp. 504–511.
- [28] Dazhe Zhao Wei Li Peng Cao and Junbo Wang. “Pulmonary Nodule Classification with Deep Convolutional Neural Networks on Computed Tomography Images”. In: *Computational and Mathematical Methods in Medicine* (2016).
- [29] The Cancer Imaging Archive. *LIDC IDRI*. URL: <https://wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI>.
- [30] J. Zhang and J. Hu. “Image Segmentation Based on 2D Otsu Method with Histogram Analysis”. In: *International Conference on Computer Science and Software Engineering* (2008).
- [31] H. Al-shamlan and A. El Zaart. “Feature Extraction Values for Breast cancer Images”. In: *International Conference on Bioinformatics and Biomedical Technology* (2005).
- [32] Matlab Documentation. *Shape Measurements*. URL: <https://nl.mathworks.com/help/images/ref/regionprops.html>.
- [33] K. Shanmugam R. M. Haralick and I. Dinstein. “Textural Features of Image Classification”. In: *IEEE Transactions on Systems SMC-3* no.6 (1973).
- [34] R. M. Haralick and L. G. Shapiro. *Computer and Robot Vision*. Vol. 1. 1992, p. 459.

- [35] Ashish K. Maurya Ritesh Maurya Surya Kant Singh and Ajeet Kumar. “GLCM and Multi Class Support vector machine based automated skin cancer classification”. In: *International Conference on Computing for Sustainable Global Development (INDIACom)* (2014).
- [36] Md. Hasanul Kabir Mir Rayat Imtiaz Hossain Imran Ahmed. “Automatic Lung Tumor Detection Based on GLCM Features”. In: *Asian Conference on Computer Vision* (2015).
- [37] M. F. Moller. “A scaled conjugate gradient algorithm for fast supervised learning”. In: *Neural Networks* 6 (1993), pp. 525–533.
- [38] Stuart A. Taylor et al. “Influence of computer-aided detection false-positives on reader performance and diagnostic confidence for CT colonography”. In: *AJR Am J Roentgenol* (2009), pp. 1682–9.
- [39] William Henry Nailon. “Texture Analysis Methods for Medical Image Characterisation”. In: *Biomedical Imaging* (2010).