Research Paper Business Analytics

Detection of lung nodules in CT scans: Comparison between different machine learning techniques

Francien Meijer

Date: 13-03-17

Supervisor: dr. Mark Hoogendoorn

VU VRIJE UNIVERSITEIT AMSTERDAM

Faculty of Sciences,
Vrije Universiteit Amsterdam

Abstract

The use of CT scans can contribute to accurate and early detection of lung cancer. However, it costs a lot of time to analyse these scans manually. A good alternative of this manually performed analysis of lung cancer in CT scans, is the use of computer-aided detection (CAD) systems. Radiologists make use of these CAD systems in order to detect anomalies, and diagnose earlier and faster. In this research project a CAD system was developed for lung cancer detection. The aim of this study was to compare the differences in performance, in the classification of nodules in CT scans, between four different machine learning techniques: neural network, support vector machine, naïve Bayes and linear discriminant analysis. A dataset published by the Lung Image Database Consortium (LIDC) was used for this research. This dataset consists of diagnostic and lung cancer screening thoracic computed tomography (CT) scans with marked-up annotated lesions. In this research, the regions of interest (ROIs) were segmented with morphological top-hat filtering and binarization. Thereafter, classification of the ROIs was performed by different machine learning techniques. The characteristics of the ROIs were used as features for this classification. The main results show that the four different machine learning techniques: neural network, support vector machine, naïve Bayes and linear discriminant analysis, reach a sensitivity of respectively 68%, 74%, 73% and 67%. Based on the 95% confidence intervals, this research indicates that for this CAD system, the support vector machine and the naïve Bayes model provide a significant better classification performance of the nodules than the other two techniques used.

Contents

1. Introduction

Of all cancer types, lung cancer currently leads to the highest number of deaths (Niki et al., 2001; Kuruvilla & Gunavathi, 2014). Kuruvilla & Gunavathi (2014) noted that 12.3% of the total number of cancer diagnosis over the world are diagnoses of lung cancer. Additionally, 17.8% of all cancer deaths are the result of this type of cancer (Parkin, 2001; Kuruvilla & Gunavathi, 2014). These statistics show that research into the detection and the diagnosis of lung cancer is of great importance. Early detection and especially the specific location of the nodules can be very important for disease staging and for determining an appropriate treatment (Karimaghaloo et al., 2016). In this way, early detection can lead to higher chances of survival and thus an enhanced survival rate (Kuruvilla & Gunavathi, 2014; Niki et al., 2001).

A method that can be applied to enhance the chances of survival of lung cancer patients, is the screening of thoracic CT images. According to Gurcan et al. (2002) this method can improve early detection of lung nodules. Additionally, Niki et al. (2001) indicated that since 1993, the use of thoracic CT images improved the early detection of lung cancer. However, although the use of CT scans can contribute to accurate and early detection of lung cancer, it costs a lot of time to analyse these scans manually. As CT scans exist of many images, it is a demanding task for radiologists to screen the lungs for lung cancer (Gurcan et al., 2002; Armato et al., 2004). As a consequence, a high work load can lead to errors in the detection of lung cancer (Armato et al., 2004).

A good alternative of this manually performed analysis of lung cancer in CT scans, is the use of computer-aided detection (CAD) systems (Gurcan et al., 2002). CAD systems assist radiologists in earlier and faster detection and diagnoses of anomalies (Kuruvilla & Gunavathi, 2014). CAD systems often entail of two separate phases, the feature selection phase and classification phase (Tan et al., 2011; Armato et al., 2004). The first phase consists of feature selection techniques, which select the regions of interest (ROIs) in the CT image. In the second phase, these selected ROIs are classified (Tan et al., 2011; Armato et al., 2004). In some research this is a classification between nodules and non-nodules (Ozekes & Osman, 2010; Gurcan et al., 2002) and in other research between malignant and non-malignant nodules (Niki et al., 2001; Suzuki et al., 2005).

In the classification phase different machine learning techniques can be used, for example support vector machines, neural networks and naïve Bayes classification (Ozekes & Osman, 2010). Studies that have used different classification techniques, presented different results. Unfortunately, a good comparison between these different studies, and thus the different machine learning techniques is not possible as each study used different data. In order to make a good comparison between the different machine learning techniques, this research will use the same data applied to four different machine learning techniques. Based on the results the following research question will be answered: Which of the machine learning techniques is the best practice for the classification of the ROIs in nodules or non-nodules?

This research paper contributes to the research domain of CAD systems for lung cancer detection. The results are important as CAD systems can provide earlier and more accurate detection of lung cancer compared to the manual analysis of radiologists. Additionally, knowing the most sensitive classification technique will benefit the detection. Consequently, more nodules will be found and this will lead to a better treatment plan, which at the end will lead to a higher survival rate.

The outline of this paper is as follows: Firstly, a description of the used data is presented. Secondly, the preparation of the data is explained, which includes segmentation of the ROIs and feature extraction of all the regions of interest. Thirdly, the different machine learning classifiers which are used in this research, are described. Fourthly, the experimental setup will be explained, following by the experimental results for the hyper parameters and the classifiers. Thereafter in the discussion, the results of this research are compared to the results in the literature. At last a conclusion, based on the experimental results, is given.

2. Dataset

2.1. Data Description

A dataset published by the Lung Image Database Consortium (LIDC) was used in this research. The information about this dataset is retrieved from the articles of Armato III et al. (2011) and Armato III et al. (2015). The dataset "consists of diagnostic and lung cancer screening thoracic computed tomography (CT) scans with marked-up annotated lesions" (Armato III et al., 2015). The Lung Image Database Consortium made this dataset available for development, training and evaluation of CAD systems for lung cancer detection and diagnosis. The dataset consists of sets of images from clinical thoracic CT scans of 1010 patients. All these images have corresponding XML files "that records the results of a two-phase image annotation process, performed by four experienced thoracic radiologists" (Armato III et al., 2011, p. 917). The size of the dataset is 124 GB and this corresponds with a total of 244527 images. Due to the size of the dataset, the data of 75 persons were randomly selected to analyse in this research. The data, of these 75 persons combined, consist of in total 1138 CT slices with annotation(s) to analyse. Each CT slice has a size of 512 x 512 pixels. The mean size of the annotations is 213 pixels and the annotation with the greatest size has 1319 pixels. 7.47% of the CT slices has more than one annotation on the CT slice. Figure 1 shows an example of one slice of a CT scan and the annotation of the corresponding slice.

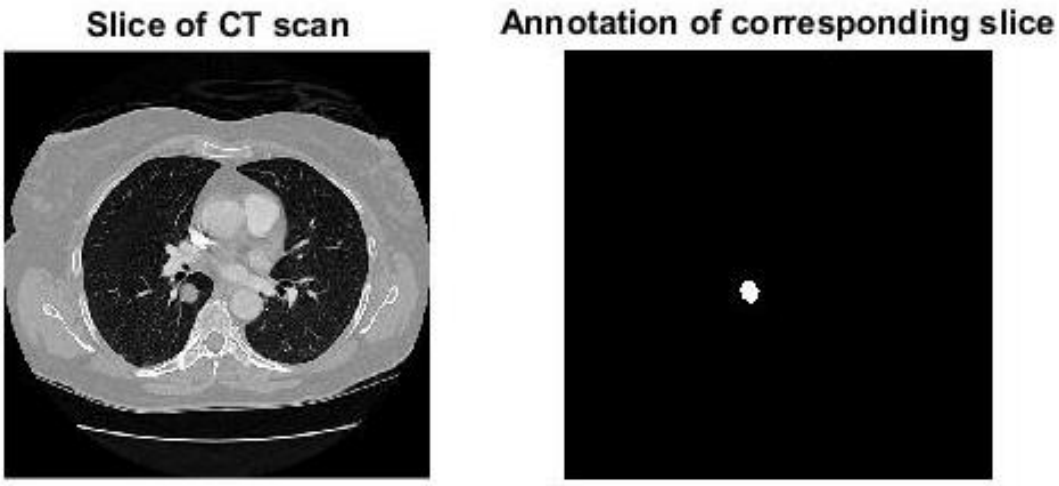

Figure 1: One slice a of CT scan and the corresponding annotation

2.2. Data Preparation

Before the machine learning techniques can be applied to the data, the data needs to be prepared. The data preparation is a process of many steps: the segmentation process, determining the regions of interests and their characteristics and labelling these ROIs. In the following sections these data preparation steps are explained in detail.

2.2.1. Segmentation

The segmentation process itself can also be divided into different steps. The first step is morphological top-hat filtering of the grayscale image. Several other filter techniques have been applied to the dataset, for example the Gaussian and entropy filter. However, eventually the morphological top-hat filter showed the best results on this dataset. This filter technique computes the morphological opening of the image and subsequently subtracts the result from the original image. This operation is performed with a structuring element. The "structuring element is a binary valued neighbourhood, [..] in which the true pixels are included in the morphological computation, and the false pixels are not" (MathWorks, 2016a, p. 545). In this research the structuring element '*periodicline'* is used. This is a flat structuring element that contains 2 X (P + 1) members, where P is the size of the structuring element (MathWorks, 2016a; Kuruvilla & Gunavathi, 2014). The left image in figure 2 shows a CT scan after morphological top-hat filtering.

The second step is converting the grayscale image into a binary image. Therefore the threshold level needs to be calculated. This is done by Otsu's method. This method calculates the threshold level that "minimizes the intraclass variance of the black and white pixels" (Kuruvilla & Gunavathi, 2014, p. 203). Once the threshold level is calculated, all pixels in the input image that have an intensity higher than the threshold level are valued as 1. Consequently, all pixels in the input image that have an intensity lower than the threshold level are valued as 0. In this way the binary image is constructed. The right image in figure 2 shows a CT scan after both steps of the segmentation process are performed.

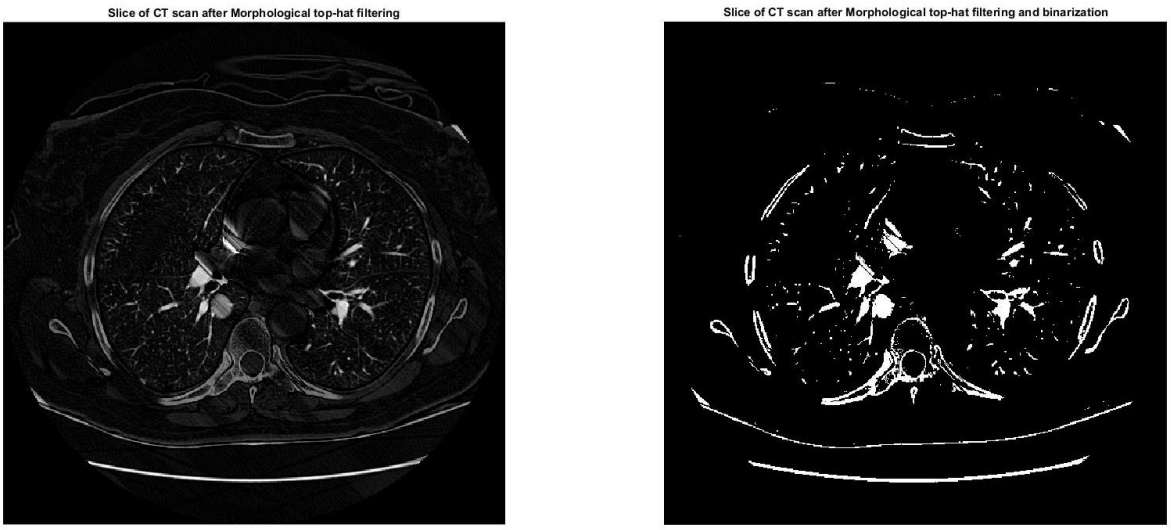

Figure 2: Slices of CT scan after the segmentation process. The left image shows a CT scan after morphological top-hat filtering. The right image is a binary image, retrieved after both segmentation process steps.

2.2.2. Regions of Interest and their characteristics

After morphological top-hat filtering and converting the grayscale image to a binary image, there remain many ROIs. Some of these regions are nodules, but most of these regions are not. These non-nodules can be, for example, blood vessels or bronchia. To make a distinction between regions who are nodules and regions who are non-nodules, the machine learning techniques make use of various characteristics. In this step of the data preparation process these various characteristics are extracted from the ROIs. These characteristics are: *Area, Centroid, BoundingBox, ConvexArea, Eccentricity, EquivDiameter, Extent, MajorAxisLength, MinorAxisLength, Orientation, Perimeter, MaxIntensity, MeanIntensity, MinIntensity, WeightedCentroid* and *StandardDeviation-PixelValues* (see Table 1). The election of these characteristics is based on different scientific articles. For example, Kuruvilla et al. (2013) used the mean of the pixel values and the standard deviation of the pixel values as features. Furthermore, Armato et al. (2001) used the eccentricity of the region as feature and Kanazawa et al. (1998) used the number of pixels in the region as feature. The characteristics, explained in table 1, are determined for every single region of interest.

Table 1: Characteristics of the ROIs (MathWorks, 2016b)

Property Name	Description
Area	A scalar that specifies the actual number of pixels in the region
Centroid	Specifies the center of mass of the region. The first element of Centroid is the x-coordinate of the center of mass and the second element is the y-coordinate of the center of mass.
BoundingBox	The smallest rectangle containing the region, specified as a 1-by-Q*2 vector, where Q is the number of image dimensions.
Eccentricity	A scalar that specifies the eccentricity of the ellipse that has the same second-moments as the region. The eccentricity is the ratio of the distance between the foci of the ellipse and its major axis length. The value is between 0 and 1.
EquivDiameter	A scalar that specifies the diameter of a circle with the same area as the region.
Extent	A scalar that specifies the ratio of pixels in the region to pixels in the total bounding box.
MajorAxisLength	A scalar that specifies the length in pixels of the major axis of the ellipse that has the same normalized second central moments as the region.
MinorAxisLength	A scalar that specifies the length in pixels of the minor axis of the ellipse that has the same normalized second central moments as the region.
Orientation	A scalar that specifies the angle between the x-axis and the major axis of the ellipse that has the same second-moments as the region. The value is in degrees, ranging from -90 to 90 degrees.
Perimeter	A scalar that specifies the distance around the boundary of the region.
MaxIntensity	A scalar that specifies the value of the pixel with the greatest intensity in the region.
MeanIntensity	A scalar that specifies the mean of all the intensity values in the region.
MinIntensity	A scalar that specifies the value of the pixel with the lowest intensity in the region.
WeightedCentroid	Specifies the center of the region based on location and intensity value. The first element of WeightedCentroid is the x-coordinate of the center of region and the second element is the y-coordinate of the center of region.
StandardDeviation-PixelValues	A scalar that specifies the standard deviation of all the pixel intensities in the region.

2.2.3. Labelling

The third step of the data preparation is labelling. In this step the ROIs are labelled as a nodule (value 1) or a non-nodule (value 0). Each slice of a CT scan in the dataset has corresponding XML files that records the results of an image annotation process performed by four experienced thoracic radiologists. In this research the sum of all the annotations of the radiologists is used. Also for these annotations the characteristics are extracted. For each slice of the CT scan, there is a search process aimed to find the region of interest that has the nearest centroid to the centroid of the annotation. Several times the segmentation procedure did not find the annotation at all. To prevent that a non-nodule get the class of being a nodule, the following action was taken: If the region of the annotation and the region of the nearest region of interest do not have overlap, it means that the segmentation

method did not find the annotation and the slice was removed from the CT scan. This performed action is based on the research of Armato et al. (2003), who checked if the geometric center-of-mass of the bounding region of the annotation, or the location of the maximum-gray-level pixel within the bounding region of the annotation, were in the region of the estimated nodule. If so, the nodule was correctly classified. In this research this method did not find the actual nodule(s) in 192 of the total number of 1138 CT slices. So 83.13% of the CT slices was segmented correctly. In this way the classification model will not train on the wrong information. Figure 3 provides an example of a CT slice wherein the segmentation method did not find the annotation. The green region is the real nodule, annotated by the radiologists and the red region is the labelled nodule, determined by the segmentation process.

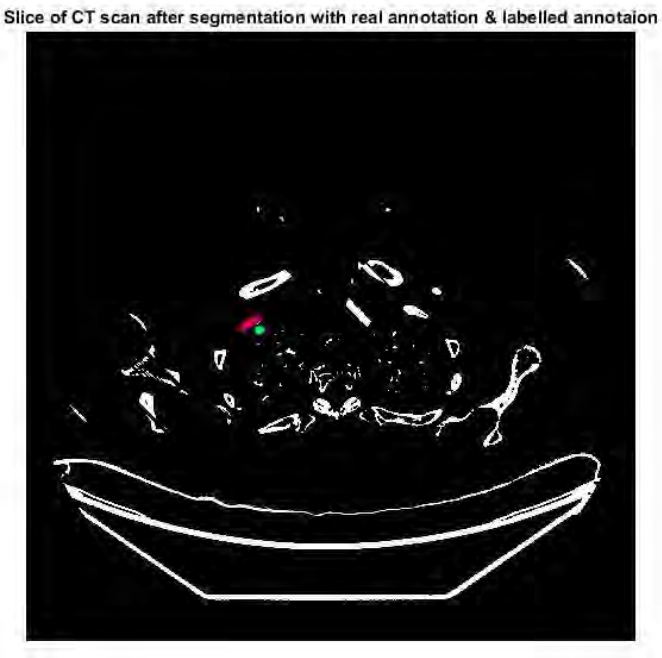

Figure 3: Slices of CT scan after the segmentation process with real annotation (green) and labelled annotation (red).

3. Approach

In this research four different machine learning techniques were used to classify the ROIs as either nodules or non-nodules. These four techniques are the support vector machine, neural network, naïve Bayes and linear discriminant analysis classifier. These techniques were chosen since other research already used these techniques to classify nodules. However none of these studies used the exact same combination of machine learning techniques in their research as this research (Gurcan et al., 2002; Kuruvilla et al., 2013; Ozekes & Osman, 2010). In the paragraphs below a short explanation of these techniques and corresponding algorithms is given based on the book of Bishop (2006) and several articles (Cortes & Vapnik, 1995; Rosenblatt, 1961; Kononenko, 1993; Mika et al., 1999).

3.1. Support Vector Machines

The first machine learning technique that was used, is the support vector machine (SVM). The following information is, apart from the book of Bishop (2006), also based on the article of Cortes & Vapnik (1995). The SVM is a machine learning technique that can solve binary classification problems. The goal of the SVM is finding an optimal hyperplane that makes a distinction between the two classes. To find the optimal hyperplane, the SVM maps the input vectors non-linearly to a high dimensional feature space. Using the SVM there are two possibilities: in the first place, the possibility in which a distinction between the classes is possible without any errors. In this case, the optimal hyperplane is a linear classification function. The optimal hyperplane has maximal margin between the input vectors of the classes. The data points, that are situated on the maximal margin are called the support vectors (Figure 4). In the second place, there is the possibility in which a distinction between the classes is not possible without any errors. In this case the SVM tries to find the hyperplane that separates most of the data and that creates the lowest error. In this research, the goal of the SVM was to find the best classification function that makes a distinction between the data that represents nodules and non-nodules.

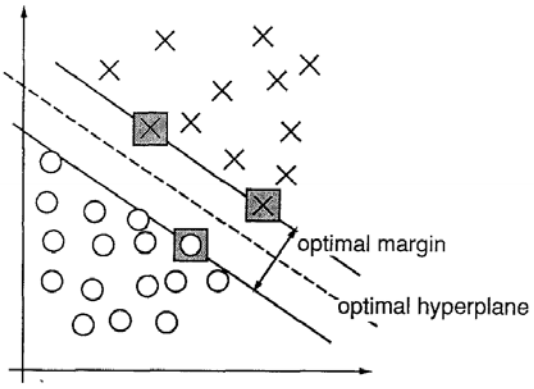

Figure 4: A linear separable dataset with optimal hyperplane. The squares on the margin are the support vectors.

3.2. Neural Network

The second machine learning technique that was used, is a neural network. The following information is based on the article of Rosenblatt (1961) and the book of Bishop (2006). The structure of a neural network is based on mathematical representations of information processing in biological systems. It can be used for different purposes such as pattern recognition, identification, classification, speech vision and control systems. In this research a neural network was used to classify ROIs in nodules or non-nodules.

The basic neural network model can be described as a series of functional transformations of the data. The neural network starts with the input layer, which exists of all the input variables. Subsequently, the neural network can have one or more hidden layers. These hidden layers themselves, consist of one or more hidden nodes. Finally, the neural network has an output layer, which exists of all the output variables.

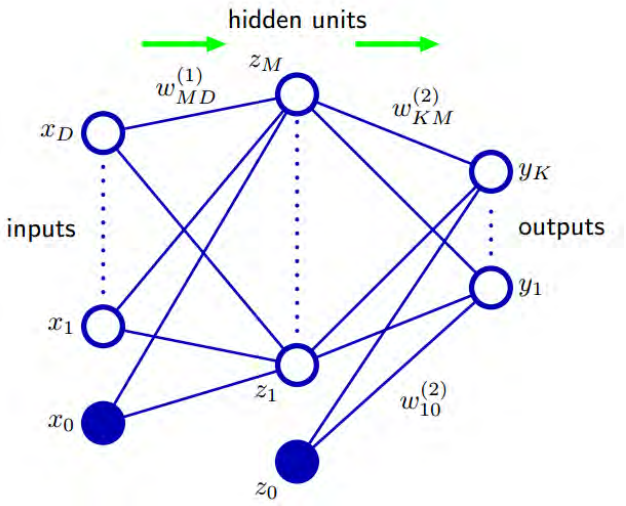

Figure 5: A neural network with one hidden layer

The first step in the neural network is constructing linear combinations of the input variables called activations. In the second step the activations are transformed using an activation function, for example the logistic sigmoid function. The neural network repeats these two steps for each hidden layer. The hidden neurons can have different activation functions. At the end, the transformed activations are combined linearly. This is done in order to calculate the output unit activations. As in all layers, the output unit activations are transformed using an activation function and the final output of the network is calculated. To train the weights of the layer, a training algorithm is used. This training algorithm is an iterative process and has the goal to minimize the error function by adaptation of the weights in the different steps of the process.

3.3. Naïve Bayes

The naïve Bayes model is the third machine learning technique that was used. The following information is, apart from the book of Bishop (2006), also based on the article of Kononenko (1993). The goal of this classifier is using the input variables with their corresponding class to construct the posterior probabilities using Bayes' theorem:

$$Posterior = \frac{Likelihood * Prior}{Evidence}$$

To construct the posterior probabilities, the class-conditional densities, also called the likelihood, together with the prior probabilities for the classes, have to be determined. Firstly, the prior probabilities can be estimated from the proportion of each class in the training set. Secondly, to construct the likelihood, the naïve Bayes method assumes that the predictors are conditionally independent from the class. After the likelihood and the prior probabilities are determined, the posterior probabilities can be calculated using Bayes' theorem. In other words this means that the algorithm estimates the likelihood and the prior using the maximum likelihood. In the next step, the algorithm determines the posterior probability of that input belonging to each class. Subsequently the data is marked as being a nodule or non-nodule based on which class has the highest posterior probability. In the naïve Bayes algorithm, different distributions can be used. For example, the Gaussian predictor conditional distribution and the kernel predictor conditional distribution.

3.4. Linear Discriminant Analysis

The fourth machine learning technique that was used, is linear discriminant analysis (LDA). The following information is based on the article of Mika et al. (1999) as well as on the book of Bishop (2006). The goal of this technique is finding a linear transformation that discriminates best between the nodule and non-nodule classes. The model performs a dimensionality reduction by projecting the input vector with more than one dimension, to one dimension. However, this projection can lead to information loss and some classes are better separated in the original dimension. Therefore, the idea of this classifier is to create a function that will maximize the separation between the means of the two classes and simultaneously provides a small variance within each class. This is called the Fisher's linear discriminant. The classification procedure is achieved in the transformed space.

4. Experimental Setup

In this chapter the decisions made regarding the parameters values, will be explained.

4.1. Segmentation

As mentioned in the segmentation process, the structuring element '*periodicline*' was used for the morphological top-hat filtering. This flat structuring element contains 2 X (P + 1) members, where P is the size of the structuring element (MathWorks, 2016a; Kuruvilla & Gunavathi, 2014). The value of P is set to 10. These settings are based on the research of Kuruvilla et al. (2013). Kuruvilla et al. (2013) used a structuring element of the size P = 2. Also in this research, the segmentation process initially started with a structuring element of the size P = 2. However, after some trial and error the process showed that P = 10 results in a better segmentation.

4.2. Under Sampling

After all the data preparation steps were executed, a new dataset was created to work with. This dataset had only a small number of ROIs, which were actual nodules. This is not strange, since most of the slices of the CT scan had one or two nodules and the number of regions of interest selected in one slice was very big. Therefore, a lot of regions of interest were non-nodules. In case that all the data is used to train the model, the model will train itself mostly on data that classifies regions of interest as non-nodules. In order to prevent this, all the regions of interest that are actual nodules were selected. In addition to these actual nodules, seven times as many regions of interest were added to the data set which are non-nodules. In this way a new data set is created. The final under sampled data set, consisted of 1007 number of regions that were actual nodules and the ratio of nodules and non-nodules in the data set is set to 1:7. Other ratio's have been explored, but increasing or decreasing the ratio has pushed the classification too much to one of the classes.

4.3. Training set vs. Test set

Once under sampling was finalized, the dataset set was divided into two parts. The first part of the dataset was used as training set and the other part was used as test set. This was done in order to avoid over fitting. In accordance to the rule of thumb, for splitting a dataset in a training and test set, the dataset is randomly split in $\frac{2}{3}$ training set versus $\frac{1}{3}$ test set (Tan et al., 2011).

4.4. Parameter Settings

In three of the machine learning techniques, parameter setting was performed relatively straightforward. Two different optimization algorithms were used, in order to find the optimal hyper parameters of the support vector machine, naïve Bayes model and the linear discriminant analysis model. In the first place, Bayesian optimization was used for the support vector machine. In the second place, grid search was used for the naïve Bayes model as well as the linear discriminant analysis model. Both optimization algorithms were applied to the training set, along with five-fold cross-validation.

Parameter setting in the fourth machine learning technique, the neural network, was performed differently. Additional decisions had to be made for the parameters, for example the number of hidden layers and the number of hidden neurons in the layer. In this research the input layer of the network consisted of 21 different input variables: the characteristics of the ROIs. There was one hidden layer and for this hidden layer different number of neurons were tried. The optimal number of hidden neurons in the hidden layer was determined in the range 1-100 along with five-fold cross-validation on the training set. The output layer had two neurons, classifying the ROIs in nodules or non-nodules. In short, a two-layer feed-forward network was trained. In this network the hidden neurons had the hyperbolic tangent activation function and the output neurons had the logistic sigmoid activation function. The training algorithm of the network was scaled conjugate gradient back propagation.

5. Experimental Results

Four different machine learning techniques were trained to predict which ROIs are actual nodules and which are non-nodules. The subsequent step was to apply these techniques to the prepared test set. In order to make a good comparison between the different techniques the statistical measures precision, sensitivity, specificity and accuracy were calculated to measure the performance of each model. In this chapter, the results of the hyper parameter optimisation are presented. Thereafter, the statistical measures of each machine learning technique are presented. For correctly interpreting the statistical measures, the following definitions are important:

- True Positive (TP): The number of ROIs that is correctly classified as nodules.
- False Positive (FP): The number of ROIs that is not correctly classified as nodules.
- True Negative (TN): The number of ROIs that is correctly classified as non-nodules.
- False Negative (FN): The number of ROIs that is not correctly classified as non-nodules.

5.1. Results Hyper Parameter Optimisation

In table 2, the optimal hyper parameters for each machine learning technique are presented. With regard to the neural network, the previous chapter explained that different numbers of hidden neurons were used in the hidden layer. In figure 6, the number of hidden neurons used in the hidden layer is plotted against the performance of the neural network. As the figure shows, 91 hidden neurons lead to the highest accuracy.

Table 2: Optimal hyper parameters for each machine learning technique

Support Vector Machines		**Naïve Bayes**		**Linear Discriminant Analysis**		**Neural Network**	
Hyper Parameters	Value	**Hyper Parameters**	Value	**Hyper Parameters**	Value	**Hyper Parameters**	Value
Box Constraint	978.7281	**DistributionNames**	Normal	**Delta**	0.00001	**Number of hidden nodes**	91
KernelScale	12.2638	**Width**	-	**Gamma**	0		
Kernel Function	Gaussian	**Kernel**	-				
PolynomialOrder	-						
Standardize	True						

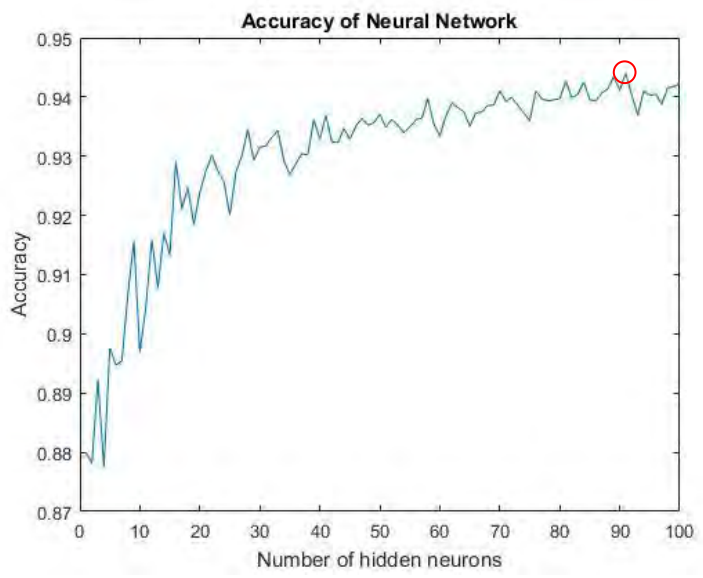

Figure 6: Optimal number of hidden neurons

5.2. Results Machine Learning Techniques

In table 3, an overview of the results of the different techniques is presented. The table shows the results of the different statistical measures for all models.

Table 3: Statistical measures for all models

Model	TP	FN	TN	FP	Precision	Sensitivity	Specificity	Accuracy
Support Vector Machine	262	90	2288	45	85.34%	74.43%	98.07%	94.97%
Neural network	241	111	2276	57	80.87%	68.47%	97.56%	93.74%
Naïve Bayes	257	95	2160	173	59.77%	73.01%	92.58%	90.02%
Linear Discriminant Analysis	237	115	2263	70	77.20%	67.33%	97.00%	93.11%

Due to the fact that the dataset is unbalanced (there are more non-nodules than nodules), the 'accuracy' is not a good statistical measure to determine the best machine learning technique. Most literature in this domain use the statistical measure 'sensitivity' to determine the performance of the model and subsequently to compare it with other models (Ozekes & Osman, 2010; Gurcan et al., 2002). Therefore, also in this research the best model is determined based on the sensitivity. To check whether there is a significant difference between the sensitivities of the models, a 95% confidence interval for each sensitivity of each model was calculated. Table 4 shows the sensitivities and the 95% confidence intervals.

Table 4: Confidence intervals of the sensitivity of all models

Model	Sensitivity	95% Confidence Interval of the sensitivity
Support Vector Machine	74.43%	[72.75%, 76.12%]
Neural network	68.47%	[66.67%, 70.26%]
Naïve Bayes	73.01%	[71.30%, 74.72%]
Linear Discriminant Analysis	67.33%	[65.52%, 69.14%]

6. Discussion

In this research a computer-aided detection system for lung cancer detection was developed. Based on prior research, the CAD system shows promising results for the detection of lung nodules in CT scans (Boroczky et al., 2006; Ko & Betke, 2001; Fiebich et al., 1999). This research made a comparison between four different machine learning techniques, while using the same data. It included segmentation of the ROIs, feature extraction and classification of the ROIs in nodules or non-nodules with different machine learning techniques (support vector machine, neural network, naïve Bayes and linear discriminant analysis).

6.1. Performance of the segmentation method

As explained in the data preparation chapter, the segmentation of the ROIs was carried out in 2D. This segmentation method did not find the actual nodule(s) in 192 of the total number of 1138 CT slices. This gives an accuracy of 83.13%. In comparison with the existing literature, other research have found corresponding or higher accuracies. For example, Gurcan et al. (2002) achieved an accuracy of 90% in the segmentation stage. Another research, done by Armato et al. (2001), achieved a percentage of 21.1% of actual nodules, who were not found in the segmentation procedure. This corresponds to an accuracy of 78.9%. Furthermore, Tan et al. (2011) found in the detection stage accuracies of 88.8%, 96.5%, 96.8% and 98.8% respectively at agreement levels 1, 2, 3 and 4. These agreement levels indicated the minimal number of radiologists that marked the nodule. Agreement level one, in which the nodules are marked by at least one of the radiologists, match the method used in this research. These accuracies are relatively similar.

The differences in the above accuracies of the segmentation method, can be explained by different reasons. The first explanation could be that different filter techniques were applied before classification (Tan et al., 2011). Dependent on the used filter technique, different parts of the CT image remain available for further use and this could lead to discrepancies in accuracies. The second explanation could be the difference in number of radiologists that marked a region as a nodule. For example, in some research regions were only labelled as nodules in case they were marked by all radiologists participated in the research, instead of only one. As a result, only the obvious nodules were labelled as nodules. Since these obvious nodules are easier to detect during segmentation, this could lead to differences in accuracies. A third explanation can be the fact that the different studies did not use the same datasets. For example, studies used datasets with different number of cases, scanning

protocols and slice thickness (Tan et al., 2011). A final explanation can be the fact that not every research used the same labelling method (Tan et al., 2011).

Additionally there are reasons that explain the fact that not all nodules were found in the segmentation method. Based on earlier research of El-Baz et al. (2013), Firmino et al. (2014) stated that mistakes in detection of nodules can be caused by the presence of small nodules, ground-glass opacity nodules, nodules attached to vessels, and nodules attached to parenchymal wall and diaphragm. It is a difficult task to segment small nodules, because of the spatial discretization used for the CT imaging. In this case, a voxel represents several tissue types, which result in averaging of their intensity values. Furthermore, Cascio et al. (2007) noted difficulties in the detection of ground glass nodules due to the fact that they have poorly defined borders and they are of low attenuation. To reach higher accuracies more research is needed in filter techniques.

6.2. Best machine learning technique

The results of this research show that the support vector machine and the naïve Bayes model are most accurate for classifying the ROIs in nodules and non-nodules after the mentioned segmentation method was applied. It is difficult to make a comparison with the existing literature, because little research has been done in comparing different machine learning techniques for classification of ROIs in nodules or non-nodules. Additionally, Tan et al. (2011) noted that a comparison with other CAD systems is difficult, because not every research used the same data, the same labelling methods, the same scoring methods or the same validation standards.

Nevertheless, some comparisons are possible as Ozekes & Osman (2010) compared the classification performance of four different machine learning techniques: neural network, support vector machine, naïve Bayes and logistic regression. Their research showed that the neural network model provides the best performance in comparison with the other techniques in nodule classification. Furthermore, Tan et al. (2011) compared the classification performance of three different machine learning techniques: feature-deselective neuro evolving augmenting of topologies method, support vector machines and fixed-topology neural networks. In this research of Tan et al. (2011), the fixed-topology neural network model gives the best performance in comparison with the other techniques in nodule classification. The results of these studies, however, do not corresponds with the current research, that found best performance using the support vector machine and the naïve Bayes model.

6.3. Critical points

In this research 2D segmentation is used. However, 3D segmentation would result in more accurate results. 3D feature extraction can provide better information to make a distinction between nodules and non-nodules. Due to time constraints 2D segmentation was applied. Another critical point concerns the filter technique that is used in this research. In figure 4 of the data preparation chapter, various structures of the body are visible that were unimportant in this research, for example bones. By using another filter technique or a combination of techniques, these unimportant regions could have been removed. This would augment the accuracy of the classification phase. These two critical points may have contributed to the low detection scores of this research. Therefore this research did not beat the manual detection by radiologists.

6.4. Recommendations for further research

Based on the abovementioned critical points, there are some recommendations for further research. These recommendation improve the comparisons between the different machine learning techniques. The first recommendation is applying 3D segmentation, which may lead to an increased accuracy of the segmentation procedure. The second recommendation is to conduct more research into the most accurate filter techniques that can be used. The third recommendation is to conduct more research in the field of different agreement levels of radiologists and comparing these results. Finally, as this research project did, more research is needed using the same data with different techniques. In this way, the performance of different machine learning techniques in nodule detection can be compared.

7. Conclusion

Based on this research, in which specific filter techniques were applied and specific characteristics were used as features for classification, the following can be concluded: The support vector machine and the naïve Bayes model give the best performance for classification of the ROIs in nodules or non-nodules.

8. References

Armato, S. G., Altman, M. B., & La Riviere, P. J. (2003). Automated detection of lung nodules in CT scans: effect of image reconstruction algorithm. *Medical Physics*, *30*(3), 461-472.

Armato, S. G., Giger, M. L., & MacMahon, H. (2001). Automated detection of lung nodules in CT scans: preliminary results. *Medical physics*, *28*(8), 1552-1561.

Armato, S. G. III., McLennan, G., Bidaut, L., McNitt-Gray, MF., Meyer, C. R., Reeves, A. P., Zhao, B., Aberle, D. R., Henschke, C. I., Hoffman, E. A., Kazerooni, E. A., MacMahon, H., van Beek, E. J. R., Yankelevitz, D. (2011). The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A completed reference database of lung nodules on CT scans. *Medical Physics*, 38, 915-931.

Armato III, S. G., McLennan, G., McNitt-Gray, M. F., Meyer, C. R., Yankelevitz, D., Aberle, D. R., Henschke, C. I., Hoffman, E. A., Kazerooni, E. A., MacMahon, H., Reeves, A. P., Croft, B. Y., Clarke, L. P. (2004). Lung image database consortium: Developing a resource for the medical imaging research community. *Radiology*, *232*(3), 739-748.

Armato III, Samuel G., McLennan, Geoffrey, Bidaut, Luc, McNitt-Gray, Michael F., Meyer, Charles R., Reeves, Anthony P., … Clarke, Laurence P. (2015). Data From LIDC-IDRI. The Cancer Imaging Archive. Retrieved May 31, 2017, from http://doi.org/10.7937/K9/TCIA.2015.LO9QL9SX.

Bishop, C. M. (2006). Pattern recognition and Machine Learning, *Springer*.

Boroczky, L., Zhao, L., & Lee, K. P. (2006). Feature subset selection for improving the performance of false positive reduction in lung nodule CAD. *IEEE Transactions on Information Technology in Biomedicine*, *10*(3), 504-511.

Cascio, D., Cheran, S. C., Chincarini, A., De Nunzio, G., Delogu, P., Fantacci, M. E., ... & Retico, A. (2007). Automated detection of lung nodules in low-dose computed tomography. Comput. Assist. Radiol. Surg. 2, 351–372.

Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, Moore S, Phillips S, Maffitt D, Pringle M, Tarbox L, Prior F. (2013). The Cancer Imaging Archive (TCIA): Maintaining and Operating a

Public Information Repository, *Journal of Digital Imaging*, Volume 26, Number 6, pp 1045-1057.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3), 273-297.

El-Baz, A., Beache, G. M., Gimel'farb, G., Suzuki, K., Okada, K., Elnakib, A., ... & Abdollahi, B. (2013). Computer-aided diagnosis systems for lung cancer: challenges and methodologies. *International journal of biomedical imaging*.

Fiebich, M., Wietholt, C., Renger, B. C., Armato III, S. G., Hoffmann, K. R., Wormanns, D., & Diederich, S. (1999). Automatic detection of pulmonary nodules in low-dose screening thoracic CT examinations. In *Medical Imaging'99* (pp. 1434-1439). International Society for Optics and Photonics.

Gurcan, M. N., Sahiner, B., Petrick, N., Chan, H. P., Kazerooni, E. A., Cascade, P. N., & Hadjiiski, L. (2002). Lung nodule detection on thoracic computed tomography images: Preliminary evaluation of a computer-aided diagnosis system. *Medical Physics*, *29*(11), 2552-2558.

Firmino, M., Morais, A. H., Mendoça, R. M., Dantas, M. R., Hekis, H. R., & Valentim, R. (2014). Computer-aided detection system for lung cancer in computed tomography scans: review and future prospects. *Biomedical engineering online*, *13*(1), 41.

Freer, T. W., & Ulissey, M. J. (2001). Screening mammography with computer-aided detection: Prospective study of 12,860 patients in a community breast center. *Radiology*, *220*(3), 781-786.

Kanazawa, K., Kawata, Y., Niki, N., Satoh, H., Ohmatsu, H., Kakinuma, R., ... & Eguchi, K. (1998). Computer-aided diagnosis for pulmonary nodules based on helical CT images. *Computerized medical imaging and graphics*, *22*(2), 157-167.

Karimaghaloo, Z., Arnold, D. L., & Arbel, T. (2016). Adaptive multi-level conditional random fields for detection and segmentation of small enhanced pathology in medical images. *Medical image analysis*, *27*, 17-30.

Ko, J. P., & Betke, M. (2001). Chest CT: Automated Nodule Detection and Assessment of Change over Time—Preliminary Experience 1. *Radiology*, *218*(1), 267-273.

Kononenko, I. (1993). Inductive and Bayesian learning in medical diagnosis. *Applied Artificial Intelligence an International Journal*, *7*(4), 317-337.

Kuruvilla, J., & Gunavathi, K. (2014). Lung cancer classification using neural networks for CT images. *Computer methods and programs in biomedicine*, *113*(1), 202-209.

Lampert, T., Stumpf, A., & Gancarski, P. (2016). An Empirical Study of Expert Agreement and Ground Truth Estimation. *IEEE Transactions on Image Processing,* 25 (6): 2557–2572.

MathWorks (2016a). Image Processing Toolbox: User's Guide (R2016b). Retrieved February 2, 2017, from https://nl.mathworks.com/help/pdf_doc/images/images_tb.pdf

MathWorks (2017b). Regionprops (R2017a). Retrieved February 2, 2017, from https://nl.mathworks.com/help/images/ref/regionprops.html.

Mika, S., Ratsch, G., Weston, J., Scholkopf, B., & Mullers, K. R. (1999, August). Fisher discriminant analysis with kernels. In *Neural Networks for Signal Processing IX, 1999. Proceedings of the 1999 IEEE Signal Processing Society Workshop.* (pp. 41-48). IEEE.

Niki, N., Kawata, Y., & Kubo, M. (2001, June). A CAD system for lung cancer based on CT image. In *International Congress Series* (Vol. 1230, pp. 631-638). Elsevier.

Ozekes, S., & Osman, O. (2010). Computerized lung nodule detection using 3D feature extraction and learning based algorithms. *Journal of medical systems*, *34*(2), 185-194.

Rosenblatt, F. (1961). *Principles of neurodynamics. perceptrons and the theory of brain mechanisms* (No. VG-1196-G-8). CORNELL AERONAUTICAL LAB INC BUFFALO NY.

Suzuki, K., Li, F., Sone, S., & Doi, K. (2005). Computer-aided diagnostic scheme for distinction between benign and malignant nodules in thoracic low-dose CT by use of massive training artificial neural network. *IEEE Transactions on Medical Imaging*, *24*(9), 1138-1150.

Tan, M., Deklerck, R., Jansen, B., Bister, M., & Cornelis, J. (2011). A novel computer-aided lung nodule detection system for CT images. *Medical physics*, *38*(10), 5630-5645.