Czech Technical University in Prague

Faculty of Electrical Engineering

Department of Cybernetics

Diploma thesis

# Reduction of False Positives in Lung Nodule Detection Algorithm

MUDr. Bc. Latnerová Iva

Supervisor: Doc. Ing. KYBIC JAN Dr.

**České vysoké učení technické v Praze**
**Fakulta elektrotechnická**

**Katedra kybernetiky**

# ZADÁNÍ DIPLOMOVÉ PRÁCE

**Student:**　　　　　　Bc. Iva  L a t n e r o v á

**Studijní program:**　Biomedicínské inženýrství a informatika (magisterský)

**Obor:**　　　　　　　Biomedicínské inženýrství

**Název tématu:**　　　Snižování počtu falešně pozitivních detekcí plicních nodulů

### Pokyny pro vypracování:

1. Úkolem práce je vytvořit algoritmus detekce plicních nodulů z 3D CT dat na základě existujícího, který bude mít podobnou citlivost, ale významně nižší počet falešně pozitivních detekcí.
2. Seznámit se algoritmem detekce nodulů popsaným v doporučené literatuře a jeho implementací, která bude poskytnuta vedoucím.
3. Zkontrolovat a opravit trénovací a testovací data.
4. Najít, vizualizovat a popsat chybně detekované struktury a analyzovat příčiny chybné klasifikace. Statisticky zhodnotit užitečnost jednotlivých deskriptorů.
5. Navrhnout nové deskriptory, vhodné pro rozlišení nyní chybně detekovaných případů, implementovat je, a ověřit jejich efektivitu.
6. Vylepšit způsob trénování klasifikátoru, zejména navrhnout a implementovat semi-interaktivní („bootstrap") trénování za pomoci experta.
7. Vyzkoušet použití různých klasifikátorů, například „Random Forest". Experimentálně změřit úspěšnost celého systému detekce v závislosti na použitých deskriptorech, klasifikátoru, databázi a dalších parametrech. Vyhodnotit nezávisle parenchymální a juxtapleurální noduly.

Práce bude probíhat ve spolupráci s autorem existujícího software Ing. Martinem Dolejším.

### Seznam odborné literatury:

[1] M. Dolejší,J. Kybic: The Lung TIME-Annotated Lung Nodule Dataset and Nodule Detection Framework. In Proceedings of SPIE, SPIE February 2009.
[2] M. Dolejší, J. Kybic: Reducing False Positive Responses in Lung Nodule Detector System by Asymmetric AdaBoost. In Proceedings of 2008 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2008
[3] M. Dolejší, J. Kybic: Automatic Two-Step Detection of Pulmonary Nodules. In Proceedings of SPIE, Medical Imaging 2007: Computer - Aided Diagnosis, vol. 6514, 2007.
[4] M. Dolejší: Detection of Pulmonary Nodules from CT Scans. Master thesis CTU-CMP-2007-05, 2007.
[5] R. O. Duda, P. E. Hart, D. G. Stork: Pattern Classification. Second edition. John Wiley & Sons, Inc., 2001.

**Vedoucí diplomové práce:**　doc. Dr. Ing. Jan Kybic

**Platnost zadání:**　do konce zimního semestru 2014/2015

L.S.

prof. Ing. Vladimír Mařík, DrSc.　　　　　　　　　　　　　prof. Ing. Pavel Ripka, CSc.
**vedoucí katedry**　　　　　　　　　　　　　　　　　　　　　**děkan**

V Praze dne 25. 6. 2013

**Czech Technical University in Prague**
**Faculty of Electrical Engineering**

**Department of Cybernetics**

# DIPLOMA THESIS ASSIGNMENT

**Student:**               Bc. Iva  L a t n e r o v á

**Study programme:**        Biomedical Engineering and Informatics

**Specialisation:**         Biomedical Engineering

**Title of Diploma Thesis:**   Reduction of False Positives in Lung Nodule Detection Algorithm

### Guidelines:

1. The aim of this task is to create an algorithm for lung nodule detection from 3D CT data based on an existing one, with a similar sensitivity but a significantly lower number of false positive detections.
2. Learn about the original algorithm described in the literature below. Study and learn to use its implementation, supplied by the supervisor.
3. Check and correct testing and training data.
4. Find, visualize, and describe incorrectly detected structures and analyze the reasons of the incorrect classification. Statistically evaluate the usefulness of the individual features.
5. Propose new descriptors useful for discriminating currently incorrectly detected objects from true negatives, implement these descriptors and experimentally test their performance.
6. Improve the classifier training, especially propose and implement semi-interactive ('bootstrap') training using feedback from a human expert.
7. Test different classifiers, for example 'Random Forest'. Experimentally evaluate the whole-system performance as a function of the detectors, classifier, databases and other parameters. Evaluate indipendently parenchymal and juxtapleural nodules.

The work will be done in collaboration with the author of the original software, Ing. Martin Dolejší.

**Bibliography/Sources:**

[1] M. Dolejší,J. Kybic: The Lung TIME-Annotated Lung Nodule Dataset and Nodule Detection Framework. In Proceedings of SPIE, SPIE February 2009.
[2] M. Dolejší, J. Kybic: Reducing False Positive Responses in Lung Nodule Detector System by Asymmetric AdaBoost. In Proceedings of 2008 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2008
[3] M. Dolejší, J. Kybic: Automatic Two-Step Detection of Pulmonary Nodules. In Proceedings of SPIE, Medical Imaging 2007: Computer - Aided Diagnosis, vol. 6514, 2007.
[4] M. Dolejší: Detection of Pulmonary Nodules from CT Scans. Master thesis CTU-CMP-2007-05, 2007.
[5] R. O. Duda, P. E. Hart, D. G. Stork: Pattern Classification. Second edition. John Wiley & Sons, Inc., 2001.

**Diploma Thesis Supervisor:**  doc. Dr. Ing. Jan Kybic

**Valid until:**  the end of the winter semester of academic year 2014/2015

L.S.

prof. Ing. Vladimír Mařík, DrSc.                              prof. Ing. Pavel Ripka, CSc.
        **Head of Department**                                          **Dean**

Prague, June 25, 2013

## Prohlášení autora práce

Prohlašuji, že jsem tuto práci vypracovala samostatně a že jsem uvedla všechny použité informační zdroje v souladu s Metodickým pokynem o dodržování principů při přípravě vysokoškolských závěrečných prací

V Praze, dne _____

Podpis:_____

## Poděkování

Na tomto místě bych ráda využila možnosti poděkovat všem, kteří se zasloužili o to, že tato práce vznikla, tedy v první řadě vedoucímu práce Doc. Ing. Janu Kybicovi Dr., a to za hodnotné připomínky a především za jeho čas a trpělivost. Dále Ing. Martinu Dolejšímu za cenné rady. A v neposlední řadě všem blízkým, kteří se mnou sdíleli světlé i tmavé okamžiky vzniku této práce.

# List of abbrevations

HRCT (High Resolution Computed Tomography)

HU (Hounsfield unit) – units describing CT density

MPR (multiplanar reconstruction) – postprocessing procedure creating coronal, saggital or oblique planes from the original axial data

MIP (Maximum Intensity Projection) – postprocessing procedure displaying the pixel of maximal intensity value along the line

Ground-glass nodule – nodule having the appearance of opaque glass

Axial (transversal) plane – any horizontal plane dividing the body into superior and inferior part

Coronal (frontal) plane – any vertical plane dividing the body into ventral and dorsal part

Sagittal plane – any vertical plane dividing the body into left and right part

GT (ground truth) – information about real nodules and real non-nodule structures created in advance by an expert

TPs (true positives) – all truly detected nodules

TNs (true negatives) – all truly detected non-nodule structures

FPs (false positives) – all non-nodule structures detected as nodules

FNs (false negatives) – all missed nodules

Partial volume artefact – CT artefact which occures when only a partial volume of the dense structure lying off-centre is in the way of the X ray beam

PCA (Parcial Component Analysis) – dimensionality reduction method

ROC (Receiver Operating Characteristic) – graphical curve plot ilustrating the performance of classifier system

AUC (Area Under Curve) – area under ROC curve

# Abstract

Lung nodules are the lung parenchyma structures found by radiodiagnostic imaging methods, especially Computed Tomography. Lung nodules are of various etiologies and can be found in various lung diseases . At worst they represent a primary or secondary tumorous proces of the lung. That is why it is necessary to find all suspicious lung nodules.

The aim of this study is to create the automatic lung nodule detection algorithm, based on the existing one. In my work I first analyse the baseline algorithm results to find all the shortcomings that can be improved to receive better output results. These findings are applied to create new classification method. This metod is based on reducing the number of existing nodule characteristics, modifying the training data and applying the suitable classifier to receive as good sensitivity and as low number of false positive detections as possible. For that purpose, combinations of several dimensionality reduction methods and several classifiers are studied.

New method have the same sensitivity, but significantly lower number of false positives than the existing one.

# Abstrakt

Plicní noduly jsou struktury nacházející se v plicním parenchymu, které lze zobrazit pomocí radiodiagnostických metod, nejlépe pomocí výpočetní tomografie (CT). Etiologie plicních nodulů může být různá. V nejhorším případě jsou součástí rakovinného procesu, ať už primárního, či sekundárního. Z tohoto důvodu je nutné všechny podezřelé plicní noduly spolehlivě najít a diagnostikovat.

Cílem této práce je na základě již existujícího algoritmu sestavit nový automatický systém na detekci plicních nodulů. Součástí práce je analýza dosavadních výsledků a postupů, která má za cíl najít všechny nedostatky, jejichž odstraněním by se výsledky algoritmu zlepšily. Na základě této pečlivé analýzy jsou všechny poznatky využity k tvorbě nového algoritmu. Ten je založený na snižování dimenze příznakového prostoru a použití vhodného klasifikátoru k redukci množství detekovaných dat. Testovány jsou kombinace několika metod snižujících dimenzi příznakového prostoru a několika různých klasifikátorů.

Nový algoritmus má srovnatelnou senzitivitu, ale o poznání nižší hodnoty falešně pozitivních detekcí, než algoritmus současný.

# Contents

# 1    INTRODUCTION

## 1.1    Radiodiagnostic imaging tests

Radiodiagnostic imaging tests are mostly noninvasive (not disturbing body tissue) procedures done when we suggest some patology. Imaging methods could help us find a suspicious area (leasion) that might be pathologic. They help us detect the lesion, visualize the spread of the lesion, determine other affected structures and based on all provided information they help us predict the curability and prognosis of the patient.

### 1.1.1    Radiodiagnostic imaging of the lung

The basic imaging methods and the gold standard of chest examination is plain radiograph (chest X-ray) and the Computed Tomography (CT scan). Both metods are based on absorbtion of X-ray beams passing through the human body.

Plain PA (posteroanterior) radiograph is the most common imaging method with a small radiation dose (0.02-0.1 mSv). X-ray provides us a 2-dimensional image of the body, which means, that all the structures the X-ray beam passes through are summed. This limitation of plain chest radiograph is eliminated by CT. CT provides a 3D cross-sectional view of the body. Compared to the plain radiograph, the resolution of CT is much higher. On the other hand the radiation dose is higher too (about 5 mSv for one chest CT).

These days High Resoluted Computer Tomography (HRCT) is commonly used for lung imaging. HRCT is a computed tomography method maximalizating the spatial resolution and ease the patology detection.

Another imaging method for lung investigation is Pozitron Emission Tomography (PET). PET is a functional imaging method that provides the information about metabolic activity of the tissue. It is based on consumption of a radioactive tracer (mostly used is fluorodeoxyglucose - FDG). PET investigation is indicated in tumorous processes, because increased glucose metabolism indicates the malignant potential of the leasion. It is usually combined with CT (PET/CT method), where CT provides the anatomic information and PET provides the metabolic information.

## 1.2    Lung nodule

Lung nodule is one of many pathologies that could be found in CT scan. It is a small, mostly spherical area of solid tissue localized in the lung parenchyma. In X-ray or CT scan every nodule looks like a light „spot" surrounded by the normal (dark) parenchyma of the lung. In both it is defined as a leasion of reduced transparency (or increased density).

### 1.2.1 Lung nodule imaging

Most nodules are discovered as an incidental finding in X-ray or CT. X-ray could detect only nodules larger than circa 1 cm in diameter if they are not summated with mediastinal structures. When a nodule is first discovered on the plain X-ray, it needs further investigation (usually CT scan). CT helps us to find much smaller leasions, even the micronodules, smaller than 3 mm in diameter. Lung nodule is among the most commonly seen structures in CT scans. PET has it´s limitations too. Ideal lesions for PET are larger than circa 0.5 cm in diameter, smaller lesions cannot be reliably evaluated, because they could be false negative due to limited spatial resolution of the PET camera.

### 1.2.2 Does the lung nodule always mean a cancer?

There are two main types of pulmonary nodules – benign and malignant. In the population lung nodules are usually considered as a lung cancer. Everyone knows about the cancer and is afraid of it, the lung cancer is not an exception. Cancer in general is the leading cause of death worldwide with about 7.6 million deaths every year. Even though lung cancer is second most common fatal malignancy for both men and women, it accounts about 13% of all new diagnosed cancers and it is the leading cause of cancer deaths. Nowdays, the most important risk factor for the lung cancer is tobacco (1).

### 1.2.3 Benign and malignant nodules

When speaking about pulmonary nodule, lung cancer is only the „tip of an iceberg". In fact, most of the nodules are benign. Approximately 50% of the nodules, that were surgically removed, turn out to be benign (2). It was reported, that only 1% of small nodules (<5 mm in diameter) in patients with no history of cancer are malignant and the risk of malignancy rises with the size of the nodule (3). On the other hand approximately 50% of incidentally detected nodules (>8 mm in diameter) are malignant (4). However, even the small nodule could be cancerous and might represent the primary lung cancer in it´s early state, or the secondary lung cancer, a metastase. Metastasis is a cancer with an origin somewhere else in the body which spread to the lungs. CT is the most sensitive modality for detecting lung metastases. Pulmonary metastases has its origin most commonly in the carcinoma of the breast, kidneys, colon, stomach, pancreas or thyroid gland. The mortality of both (primary and secondary lung cancer) can be reduced if they are detected and threated early.

When we first find some suspiciuous nodule, we have to recommend the further testing (HRCT, another CT scan in 3 months, PET scan, bronchoscopy, or even a biopsy).

### 1.2.4 Various appearance of one structure

We observe many signs, that could be very helpful to determine whether the nodule is probably malignant or not (Picture 1). This general knowledge of the lung nodule issue helps us understand how difficult is to find a reliable automatic algorithm for lung nodules detection.

### Size

One of the most important characteristics of the nodule is size. Mostly the size of first founded nodules are at intervals 5-10 mm. The bigger nodule, the worse prognosis. Nodules smaller than 2 mm are called the miliary nodules, nodules exceeding 3 cm in diameter are called mass, or the tumour.

### Density

Nodules could differ in density (from the solid high bone density nodules with densities about hundreds of Hounsfield units (HU) to the non-solid nodules, such called „ground glass" nodules, or even negative density nodules with fat deposits). The nodule density could be homogenoeous (which means that it does not change in the whole mass of the nodule), or heterogeneous (we could find various densities because of cavities, or calcifications).

### Shape

The shape of typical pulmonary nodule is usually spherical or round, but we can see irregular shapes, or even the spiculated nodes, which are very suspect to be malignant. The margins could be lobulated, or smooth.

### Localization

Localization of the leasion have to be taken into consideration too. We see nodules completely surrounded by lung parenchyma with no touch to the hilum, mediastinum or pleura (intraparenchymal nodules), or the leasions touched to the pleura (subpleural nodules). The distribution of multiple nodules could be perilymfatic, centrilobular or random. Single intraparenchymal nodule not associated with adenopathy, pleural effusion or atelectasis is called the solitary pulmonary nodule.

### Dynamic of growth

Important information provides us of course the dynamic of growth. Comparing the finding with prior images gives us an important information about etiology, or treatment success. Cancerous nodules grow fast, the doubling time of the malignant lesions is mostly 1-6 months (5). On the other hand, nodules which stay same in shape and size more than two years are considered to be benign (6).

### Enhancement

Another helpful information is the value of enhancement after intravenous bolus injection of the contrast agent (however most lung scans are „native", without any intravenous contrast).

### Individual patient anamnesis

Last but not least when finding a lung nodule is the patient´s anamnesis, job anamnesis, the abusus (smoking), history of malignancy, age etc.

## 1.3 Lung nodule diferential diagnosis

When finding any patological process, diferential diagnosis helps us to generate a list of diseases which have to be taken into consideration. In diferential diagnosis all processes (benign and malignant) have to be included. Benign nodules can be found in various lung diseases, they could be solitary benign neoplasms (hamartomas, chondromas), or multiple inflammatory granulomas (in tuberculosis, histoplasmosis, sarcoidosis). Septic emboli, or intrapulmonary lung nodes are nodule-like structures that could mimick nodules. Another nodule mimicking structures are rib fractures, costochondral junctions, or AV malformations.

Malignant leasions include mainly peripherial bronchial carcinoma, alveolar cell carcinoma, pulmonary carcinoid tumor, lymphoma and metastases (2).
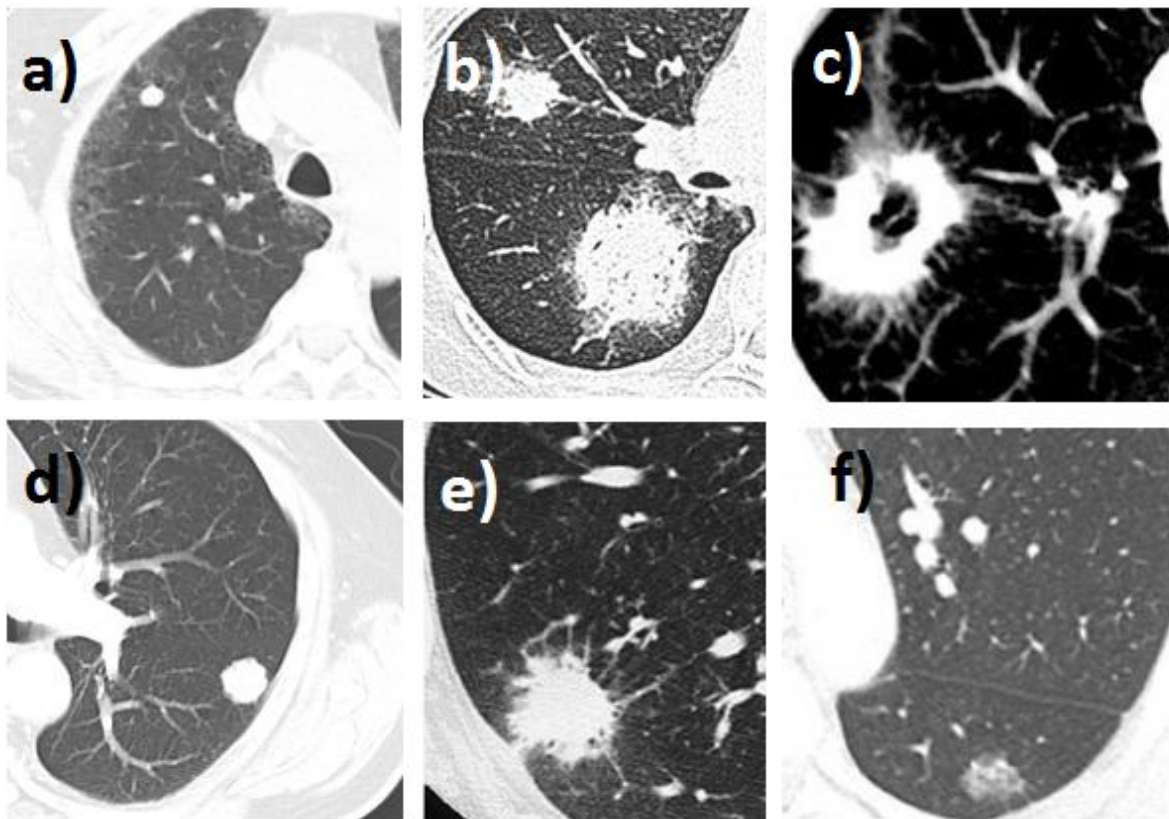


**Picture 1 – various appearance of the lung leasions a) stable benign solid nodule (7) b) two partly solid and partly ground-glass cavitating leasions representing a mycotic infection called angioinvasive aspergilosis (8) c) spiculated cavitating mass representing adenocarcinoma (9) d) solid nodule with irregular borders representing the primary lung cancer (7) e) spiculated pulmonary mass representing the primary lung cancer (10) f) ground glass nodule could represent the bronchoalveolar cell carcinoma (BAL) (11)**

## 1.4 MIP projection

Nodules come in many forms, there are ones that are easy to detect (large, round, dense and sharply marginated nodules) and problematic ones hardly classified even for an experienced human eye. Readers sensitivity for detecting pulmonary nodules could be increased by using the special postprocessing volume rendering techniques such as Maximum Intensity Projection (MIP) (12). MIP displays the voxel of maximum density $Z_{xy}$ along the Z-axis of a given volume $n$. Reader looses the sense of depth of the original data, but receives the sense of 3D. Using this technique reader could easily differeciate between rounded (eg. nodules) and tubular structures (eg. vessels).

$$Z = \{z_1, z_2, \dots z_n\}$$
$$Z_{xy} \in n$$
$$Z_{xy} = max(Z)$$

(1.1)

In the Picture 2 it is possible to compare standard multiplanar reconstruction image (MPR) and maximal intensity projection image (MIP), both in axial and coronal plane.
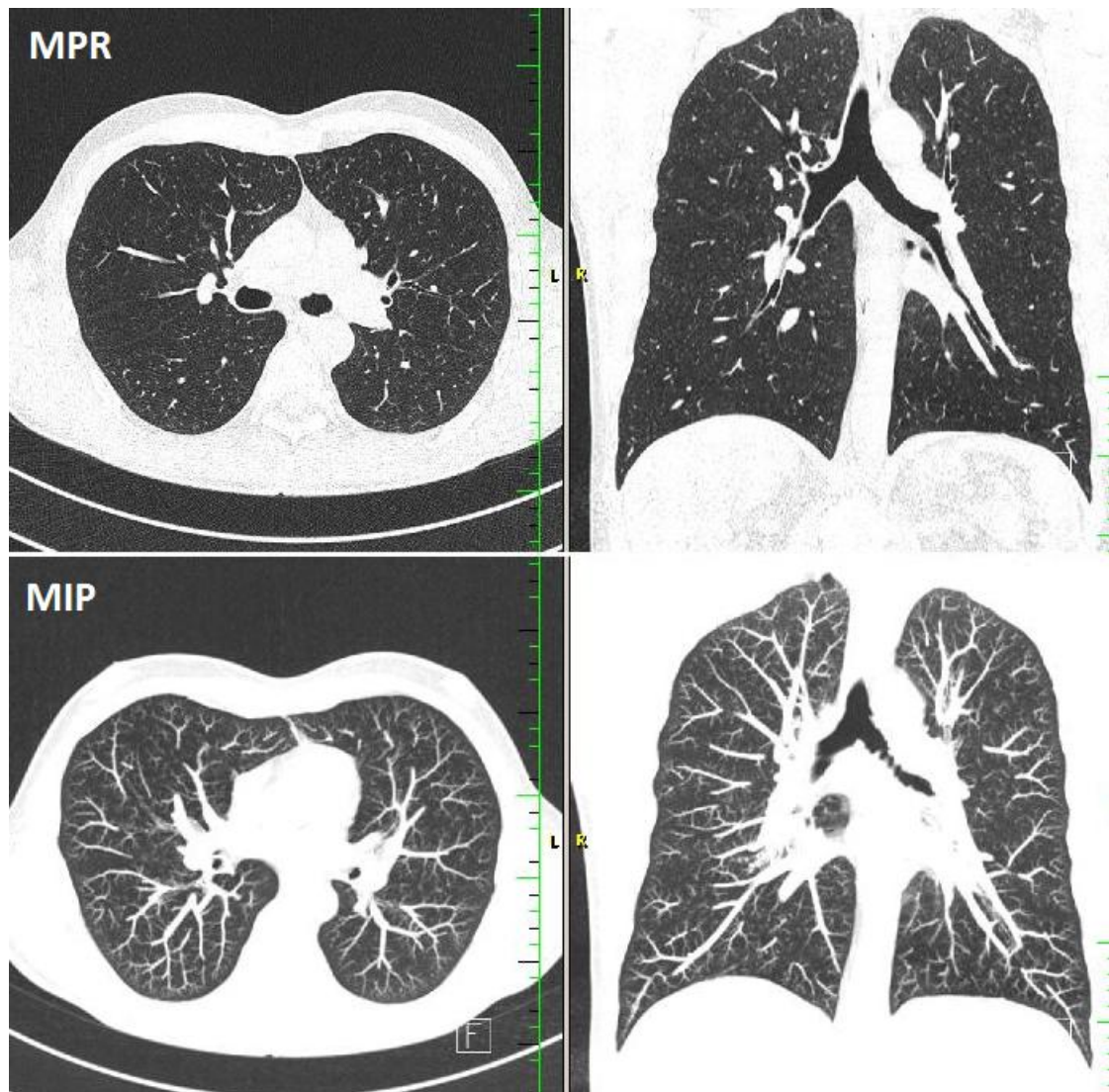


Picture 2 – MPR/MIP difference

## 1.5   Computer aided diagnosis

Computer-aided diagnosis (CAD) is a term for all the technologies and procedures in medicine that help human professionals to interpretate the medical image. The CAD research begun in early 1980s and it is still one of a major research subjects in medical imaging and diagnostics. It is believed that CAD will serve as a useful tool for diagnostic examinations in everyday praxis, not to replace the human experts, but to help them focus on suspicious structures.

### 1.5.1   No perfect results so far

Huge number of CADs for automatic lung nodule detection appeared over the years, the sensitivity reported by the authors differs in various algorithms and data sets from 71 to 95% (Table 1). Performance rates depend on the dataset type, on number of true nodules, nodule characteristics and other variables. Bigger, round and solid solitary nodules are easy to detect, while small, irregular nodules often fall into the false negative group. It was reported, that CAD system still have problems with detecting ground-glass nodules (13).
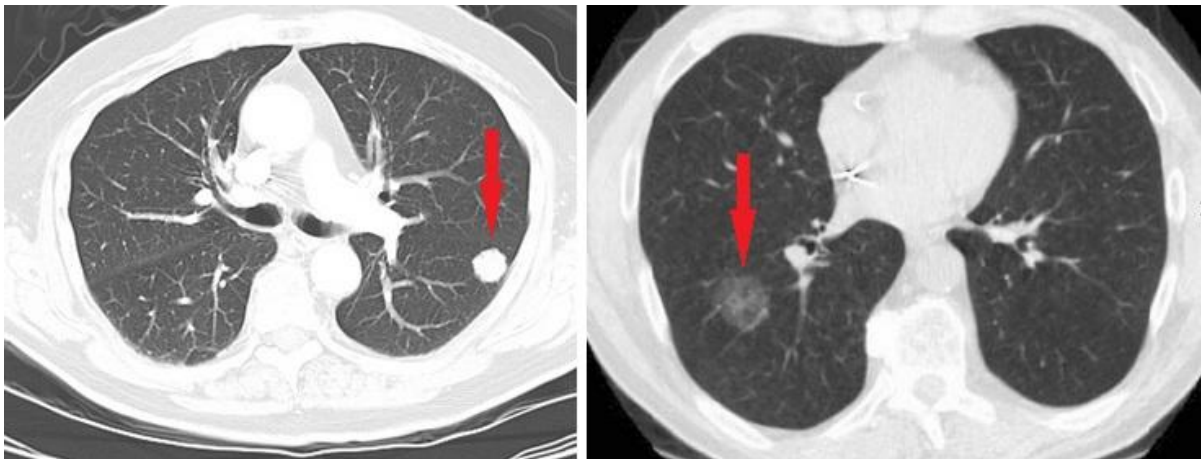


Picture 3 – examples of solid (left) and ground-glass (right) parenchyma nodules

For finding all suspicious leasions we need as high sensitivity as possible. Over the years CAD systems reached really good sensitivity rate levels, but unfortunately, high sensitivities are associated with high numbers of false positive detections. All at once, the authors are trying to reduce FPs to minimum.

We really need as good sensitivity as possible and we also need to reduce the FP rate, but not at the cost of sensitivity decreasing. But we don´t have to be perfectly sensitive in every case. When there are only few leasions, we need to find all of them, but when there are tens? It is not necessary to detect all of them, because nobody will care about the exact number. We also don´t need to get rid of all the FPs - not only nodules, but also other suspect findings have to be noticed.

## 1.5.2 Lung nodule detection algorithms compared

In following table some authors and their presented results are mentioned.

| Author | Sensitivity [%] | FPs/scan | Notes |
|---|---|---|---|
| Bisheng Zhao et al. (14) | 84.2 | 5 | Only 60-80 slices per scan, 266 simulated nodules |
| Golosio et al. (15) | 79 | 4 | Combined scan database, nodules >3mm, performed on LIDC database |
| Kyongtae et al. (16) | 95.1 | 6.9 | 164 true nodules>3mm, 20 scans used |
| Tan et al. (17) | 87.5 | 4 | 259 true nodules>3mm |
| Hirose et al. (18) | 71.4 | 0.95 | 21 scans used, nodules in 15 scans |
| El-Baz et al. (19) | 82.3 | 12 | 130 true nodules |
| Gori et al. (20) | 85.2 | 6 | 102 true nodules>5mm |
| Yuan et al. (21) | 72.6 | 3.19 | 628 true nodules |
| Choi (22) | 95.28 | 2.27 | 151 nodules>3mm, LIDC database |
| Cascio et al. (23) | 88 | 2.5 | 148 nodules |
| Dolejší et al. (24) | 94.03 | 5.46 FPs/slice | |
| Dolejší et al. (25) | 74.3 89.6 | 2.6 FPs/slice 9 FPs/slice | The presented baseline algorithm |

Table 1 – various algorithm results

Zhao (14) published 84.2% sensitivity with 5 FPs/scan, the maximum number of slices per scan was 80, there were 266 nodules (2-7 mm in diameter) and all the nodules were simulated. The maximal detection sensitivity of 94.4% was reached, however the total number of FPs per scan was 906.

Golosio (15) combined 83 scans from LIDC and 23 other scans, the mean number of slices was 310. He only included nodules >3mm in diameter. The overall sensitivity of 79% was reached with 4 FPs/scan.

Tan (17) published 87.5% sensitivity with 4 FPs/scan in LIDC database. He used 235 scans for training and 125 scans containing 259 nodules (3-30 mm in diameter) for testing. He reported maximal detection sensitivity 88.8-98.8% with mean 457 FPs per scan.

El-Baz (19) published 92.3% sensitivity with 12 FPs/scan, 200 scans were used, there were abnormalities in 21 of them. The total number of nodules was 130.

Gori (20) reported 85.3% sensitivity with 6 FPs/scan for intraparenchymal nodules and 85.2% sensitivity with 13.6 FPs/scan for subpleural nodules. He used 39 CT scans with 300 slices per scan in average, 34 scans were containing the inraparenchymal nodules and 20 scan were conaining the subpleural nodules. The dataset consisted of 102 nodules.

Yuan (21) published 72.6% sensitivity with 3 FPs/scan in 150 CT scans containing 628 true nodules. He was comparing CAD and reader sensitivity differences and found out, that CAD is much better in detecting hilar and central nodules, while human reader has higher

sesitivities in detecting peripheral and subpleural nodules. The average number of slices per CT scan was 404.

Choi (22) used 58 LIDC nodule containing scans (with nodules >3mm in diameter) with average number of slices per scan about 200 were used for testing. The total number of nudules was 151. The maximal detection sensitivity of 97% was reached. The number of FPs/scan was 60. After the classification step the number of sensitivity was 95.28% with 2.27FPs/scan. An SVM classifier was performed.

Cascio (23) used 84 LIDC scans containing 148 nodules, the mean number of slices per one scan was 310. The detection sensitivity was 97% with 6.1 FPs/scan. After FPs reduction and classification process the values dropped to 88% sensitivity and 2.5FPs/scan. He performed a classifier based on neuronal network.

Dolejší (24) published 94.27% sensitivity with 7.57 FPs/slice for TIME and LIDC datasets combined, then 94.03% with 5.46 FPs/slice for TIME only (containing two independent datasets), 89.62% with 12.03 FPs/slice for LIDC (containing 38 scans with the mean number of 223 slices per scan) and 78.68% sensitivity with 4.61 FPs/slice for ANODE09 dataset (containing 50 scans with 451 slices per scan in average).

### 1.5.3 Future diagnostic tools for helping the experts

There were many tries to automate the process of nodule-finding, but the daily routine of the human radiologists has not changed yet. The naked-eye-detection still remains the golden standard od the diagnostics, the MIP projection helps a lot.

Automatic systems definitelly have it´s future in cooperation with human experts - in detecting suspicious structures, finding new leasions, rating the growth or evaluation the threatment effectivity.  It will ease the work of humans and decrease a time needed to examine the CT scans. It also can provide the „double reading" of the diagnostic images.

## 2 REDUCTION OF FALSE POSITIVES IN LUNG NODULE DETECTION ALGORITHM

The aim of my work is to analyse the existing lung nodule detection algorithm (12), to find its potential shortcomings and to analyse its results. At the same time to set up the parameters to receive required results and based on this findings to design new method for further nodule classification.

### 2.1 Work process

The whole work process is like finding the diagnosis. First of all we have to familiarize and get to know all the details, then it is turn for analysing the problem, find and consider all the pathologies and finally based on all the information to find the main diagnosis. When we know exactly, where the problem is, we can target it and initiate threatment.

My work process will be the same. First I get to know how the original algorithm works and get familiar with the input data and functions. Then I analyse the primary results including the detected and classificated data (4.1), segmentation process (4.3) and the computed characteristics (4.4). Afterwards I create new classification algorithm using all the results of the previous analysis.

# 3 BASELINE METHOD

## 3.1 Nodule detection algorithm

I was given an access to the new CAD system designed and created by Ing. Martin Dolejší from the Czech Technical University in Prague (12). This fully automatic nodule detection algorithm is used to detect nodules in CT scans. The algorithm was implemented in MATLAB and it is divided into scripts and functions controlled by a single GUI.

### 3.1.1 Algorithm overview

The whole algorithm process consists of several separate steps. The major steps are two: finding the nodule candidates and eliminating the false positive detections by nodule classification.

The algorithm first separates the lung parenchyma from other anatomic structures (lung segmentation process using simple thresholding method), then it finds the nodule candidates using the thresholding and blob detection (based on multi-scale filtering and Gaussian filter) techniques. The local maxima are treated as nodule candidates (24). Each detected nodule candidate is considered to be a probable center of a nodule. For each nodules candidate geometrical and image characteristics are counted. Nodule candidates detector was designed to have a good sensitivity and as small number of false positives as possible. To reduce the number of false positive detections the additional step of classification is used. The author applied two classifiers - the classifier based on Fisher Linear Discriminant (FLD classifier) and the classifier based on Multiple Thresholding (AdaBoost).

Fisher Linear Discriminant uses only one linear discriminant function. The value of this function can be is positive or negative. If positive, the detection is classified as „nodule".

The number of thresholds in Multiple Thresholding method is *2n*, when *n* features is used. If the detection point is between the two borders, this detection is classified as „nodule".

Results of the existing algorithm is 95.9% sensitivity with 12 FPs/slice for the detector only, 74.3% sensitivity with 2.6 FPs/slice for FLD classifier and 89.6% sensitivity with 9 FPs/slice for the multi-threshold classifier.

Knowing the coordinates of real nodules and the non-nodule structures (the GT data) we can compare this data with the algorithm results and find out how successful the algorithm is. The GT information was created by an expert in Scan View[1].

---

[1] created by RNDr. Jan Krásenský, jkras@lf1.cuni.cz, http://www.scanview.cz/

### 3.1.2   Implementation

First I used the algoritm as it was provided and tried to understand the steps. I did not use all the scripts, only those that offer required information.

„*PrepareNodDetection.m*" is a first script responsible for the data input, it reads the data. In „*nodDetParalelScript.m*" the own detection and then the classification of the nodules takes place. Classificator is used there to estimate if the detected structure is a nodule or not. „*PrepareREGanalysis.m*" is the statistical analysis of all detected structures. This script compares all the detections with the ground truth (GT). It determines true positives (TPs), false positives (FPs) and false negatives (FNs), if there are any. It counts the sensitivity rate and number of FPs for each CT scan and for the whole set. Then I used the grafic interface „*nodVizGUI.m*" to visualize the detected structures.

## 3.2   CT imaging data

The data used for algorithm testing is provided by the Faculty Hospital, Motol, Prague. There are 4 data sets available. I only choose one, the biggest one including the human adult CT scans (PND0). This set could be considered as a representative sample.

All input 3D data are composed of $N$ slices of standart size (512 x 512 pixels). Where $N$ depends on the length of scanning data (N=386±172), the section thickness is 1 mm, voxel size (0.6 x 0.6 x 1). There are 98 scans in the database, with a total number of images=34036. The data contains 229 nodules of different size (1-11mm in diameter), shape and in different localities.

All the data is available on the cmp.server */mnt/datagrid/Motol*
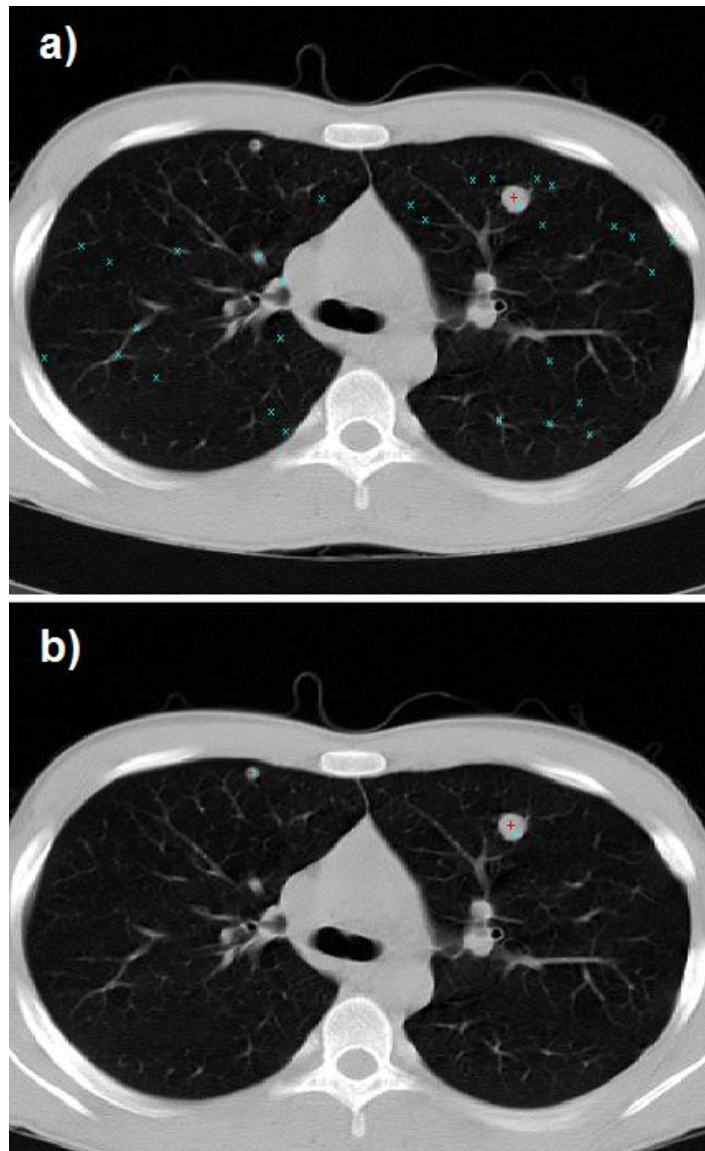
# 4 ANALYSIS OF THE BASELINE RESULTS

The first aim of this work is to analyse the results of the existing algorithm, which are still not perfect. Finding the possible deficiencies of detection and classification process can improve the results. The analysis is devided into 3 steps – the analysis of detected structures, the analysis of data segmentation and the analysis of computed features.

## 4.1 Analysis of detected structures

The basic problem of all CADs is to distinguish true nodules from nodule-like structures (such as vessels, shadows, rib margins, etc.). Checking the detected structures visually we are able to find out, what structures are mostly marked as nodule candidates, how they look like and what could be done to classify them truly as nodules or non-nodules in future. For visual scoring the detected structures I was using the included grafic interface *nodVizGUI.m*.

### 4.1.1 Necessity of the classification step

Using only the detection method without classification, there are hundreds of nodule candidates detected. The number of candidates (and also the number of FPs) decreased a lot after the classification step. For example in the chosen scan the number of FP detections decreased from 3575 to 33. After the classification step only TP markers remain in the displayed slide (Picture 4). We still can remove some of the remaining incorrectly detected structures to get better results.

### 4.1.2 True positives

TPs are the real nodules truly detected by the algorithm. When looking at all the TPs, we can devide them into two groups:

„*Examplary*" nodules look like a nodule, they have the typical nodular shape and high density, they are bigger in size (usually 3-10mm), intraparenchymal and sharply separated from the surroundings. These nodules could be benign or malignant. All the exemplary nodules I have found in the data were detected, none was missed (Picture 5).

„*Atypical*" nodules don´t have the properties mentions above (Picture 6). Most of them are small, subpleural localized, elongated, low density structures and could be easily missed. The small rounded subpleural ones look benign. (Picture 7).

21
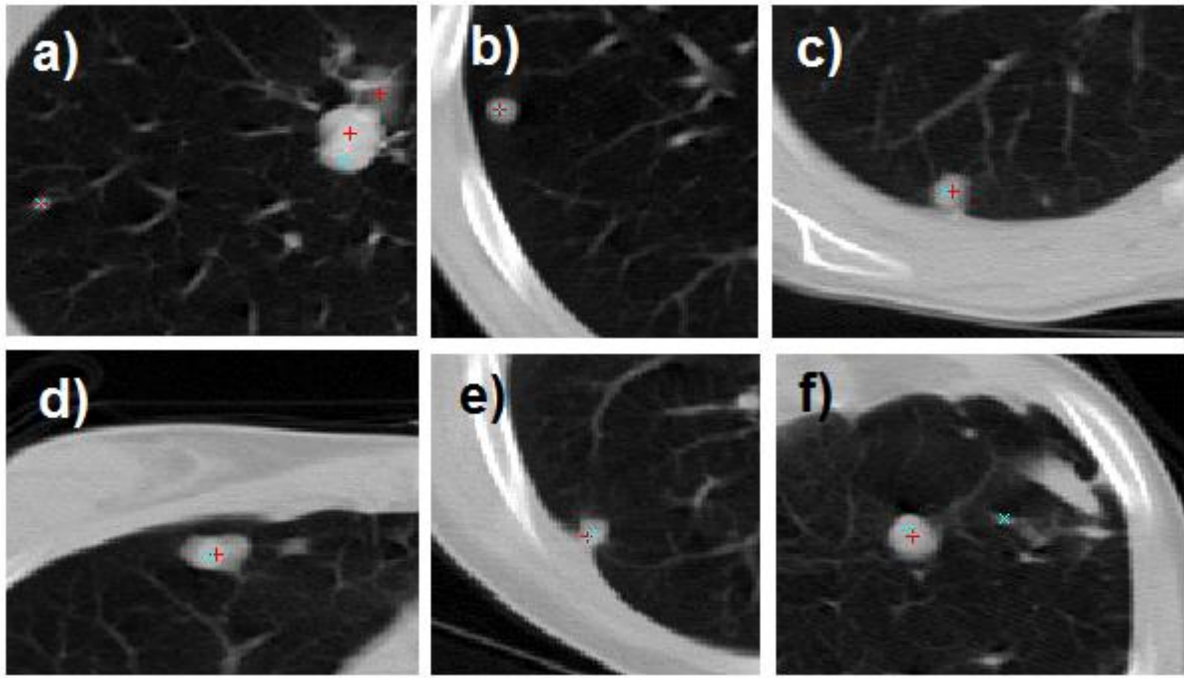
Picture 5 – „exemplary" nodules and their coordinates a) big parahilar node, PND007 (182,263,57) b) subpleural node not connected to pleura, PND007 (45,282,191) c) subpleural node connected to pleura, PND024 (134,383,61) d) big dense oval node, PND024 (164,168,124) e) subpleural node, PND024 (76,357,162) f) parenchymal node, PND024 (391,239,193)
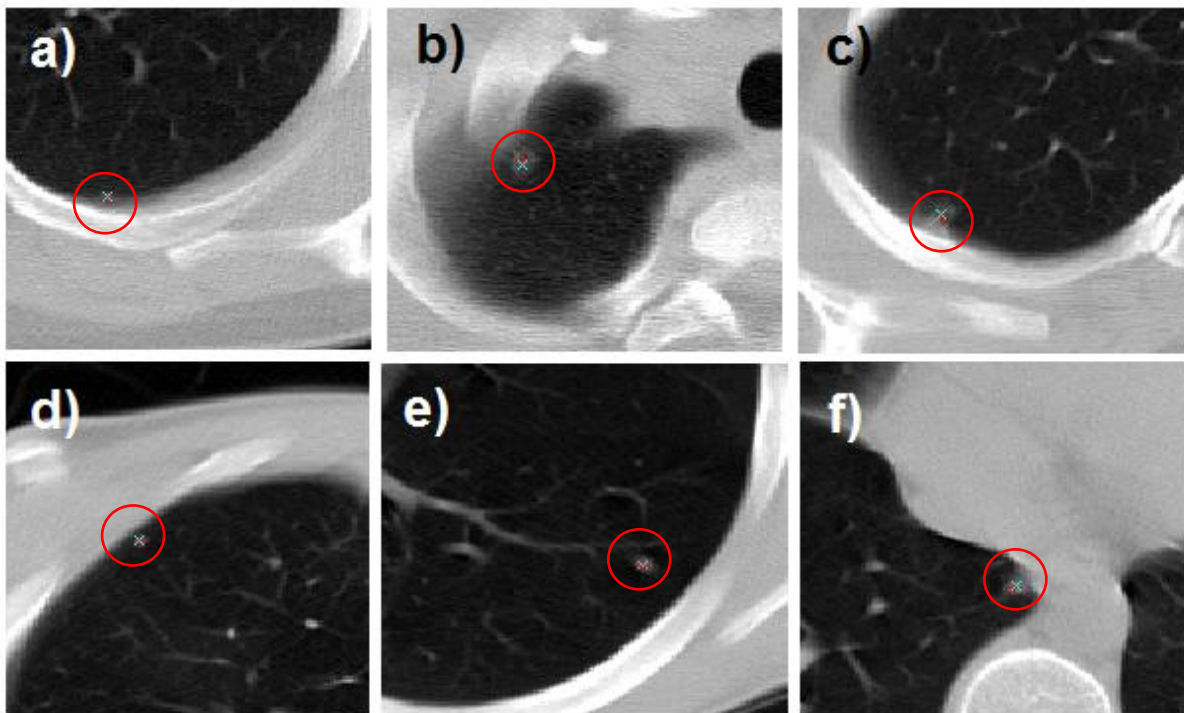


Picture 6 –„atypical" nodules and their coordinates a) small pleural benign looking nodule, PND041 (341,392,42) b) apical low dense nodule, PND051 (170,262,12) c) subpleural nodule connented to pleura, PND075 (137,345,30) d) small benign looking micronodule, PND007 (130,192,32) e) small oval nodule, PND007 (427,338,104) f) nodule connected to the mediastinum, PND024 (253,275,173)

### 4.1.3  False negatives

FNs are the real nodules missed by the algorithm. These nodules were always small, hypodense, or they did not have the typical nodular shape (mostly the „atypical" ones).
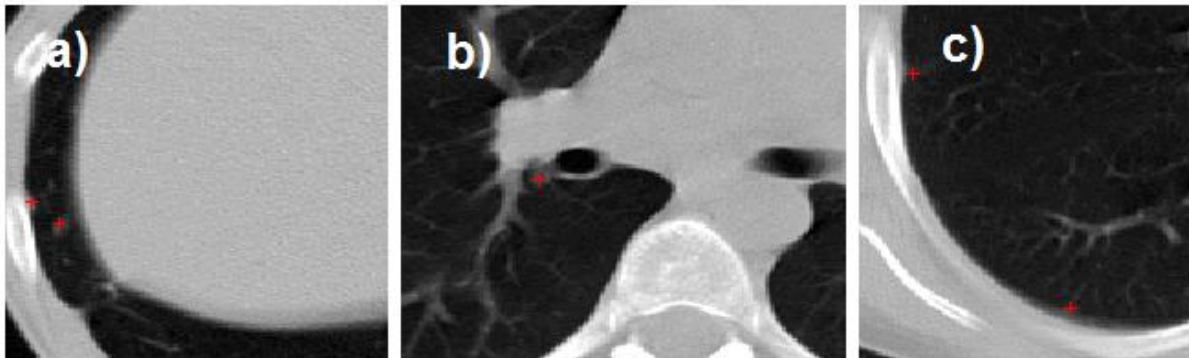
**Picture 7 – missed nodules in various localities a) basal nodules, PND007 (41,281,252) b) node in the right hilus of the lung, PND024 (200,301,92) c) subpleural structures, PND075 (70,282,9), (133,387,90)**

There also were some structures I can´t sign as nodules although they were in GT group. The algorithm didn´t find them and I agree with it. Unfortunatelly they raise the number of FNs (Picture 8). All of the undetected structures in this picture look benign.
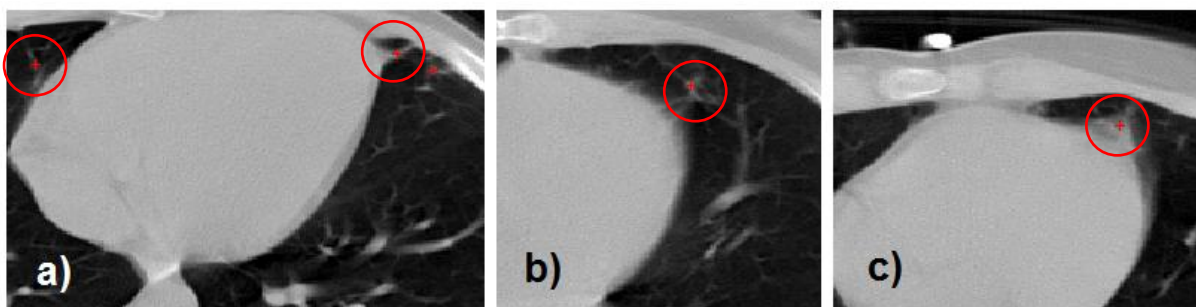


**Picture 8 – non-nodules signed as nodules in GT data, all are the areas of increased density, but not nodular, PND075**

### 4.1.4  False positives

FPs are the non-nodule structures detected as nodules. Browsing the scans one by one and seeking the shortcomings I found out that some of the structures signed as nodules looked similar. There were the same FP structures in all patients - vascular structures (pulmonary veins and arteries), pleural adhesions, small dense areas of hypoventilation or postspecific changes. Even structures with non-nodular appearance were detected (trachea, fibrous bands, pleural structures, ribs or mediastinum margins). The known CT artefacts (above all the involuntary motion artefacts caused by breathing or the heart beat, or partial volume effect artefacts) also increase the FPs.

*Mediastinal structures*

Mediastinal structures are parts of the mediastinum, the big area in the middle of the chest between the two lungs. It contains the heart, big vessels (aorta, pulmonary vessels, vena

cava), trachea and principal bronchi, esophagus and lymhatic nodes. Mostly the margins of mediastinum were detected, but there were also some detections localized into the center of the dense huge mass of mediastinum. We could reduce the mediastinal structures detections by enhancing the segmentation process, mediastinal structures such as heart, hilar structures, or other mediastinal soft tissues will be removed (Picture 9).
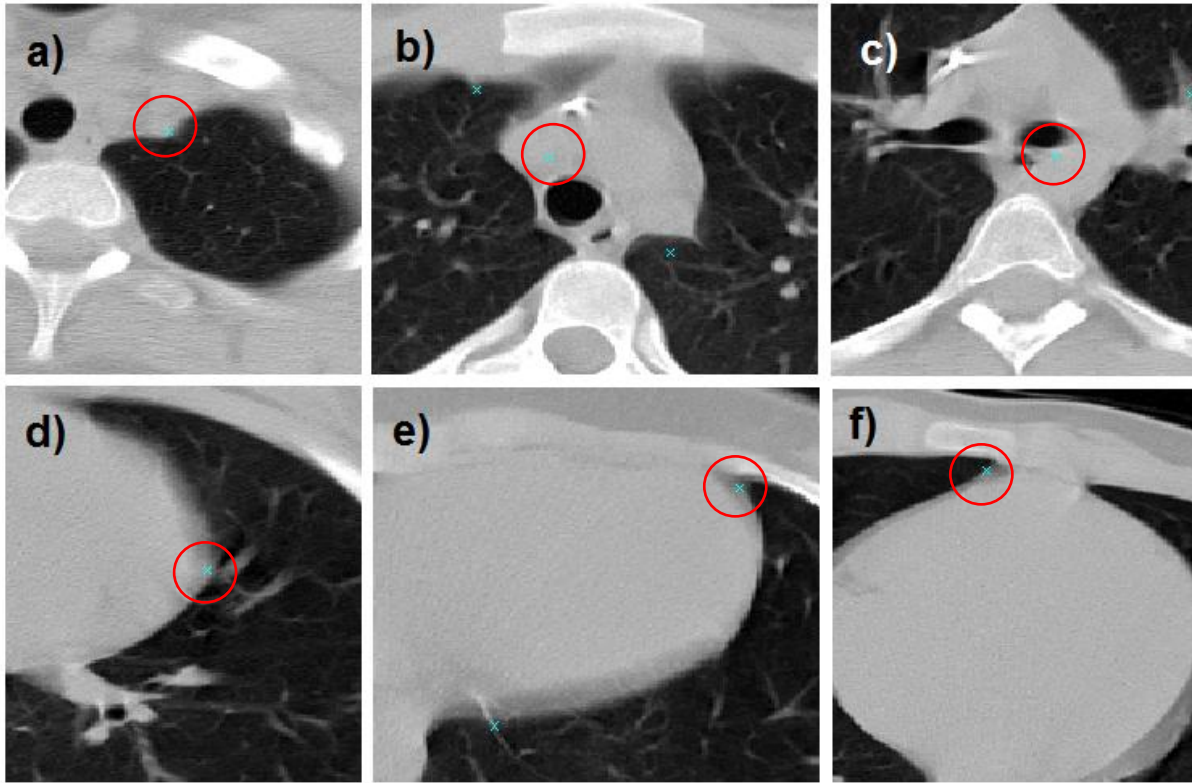


**Picture 9 – FPs mediastinal structures a) left marginally part od the upper mediastinum, PND002 (312,272,19) b) vena cava superior, PND004 (219,285,56) c) mediastinal structure detection between aorta and the left principal bronchus, PND098 (269,283,71) d) margin of the heart, PND002 (379,244,120) e) margin of the heart, PND002 (396,162,161) f) margin of the heart, PND049**

## *Lung margins and ribs*

Another often seen FP detections belong to pleura and the inner margins of ribs, especially in terrain of CT artefacts. These FPs are widely attached to the border of the lung and there are not nodular (Picture 10). Multiple detections of the same structure were also seen. For example in c) the same rib was detected 4 times.
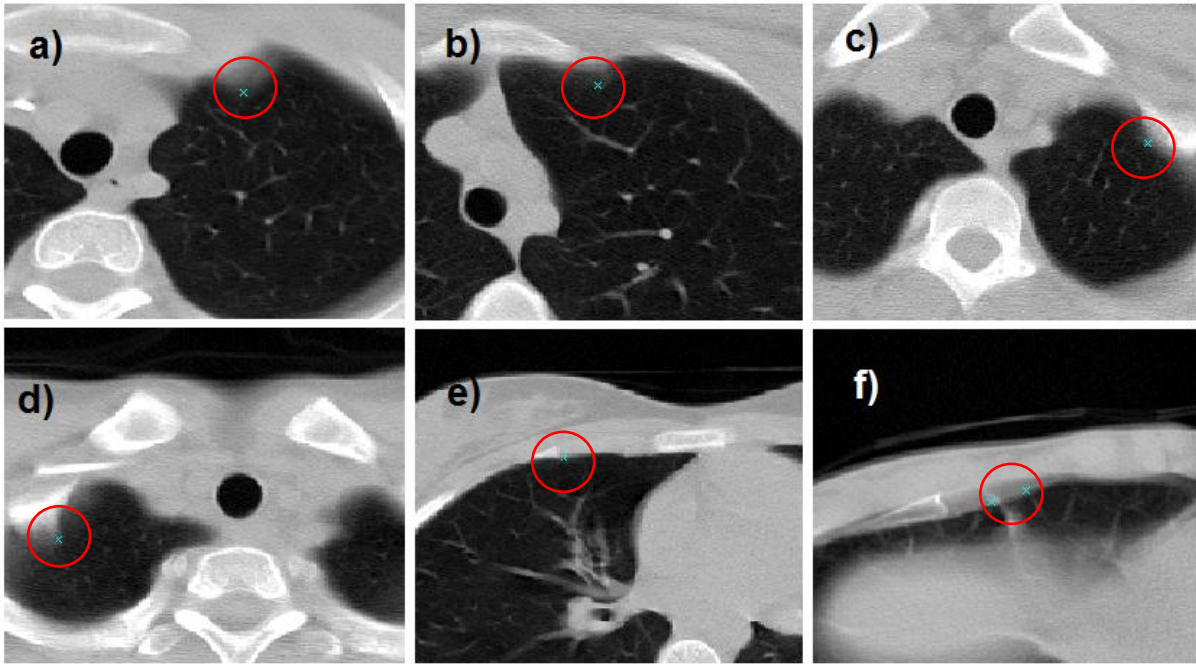
## *Vessels, vessel branching points and bronchovascular structures*

Vessel could be on 2D mistaken to the nodule because of its round shape and smooth borders, but in 3D the shape is different – if we trace the vessel we can get from perifery to hilus and back, it passes through the whole lung from the centre to the periphery giving the branches and changing the diameter gradually. In 3D space, nodules have a near spherical shape, whereas vessels have a tubular shape. The shape descriptors in all three planes could help to recognise these two (Picture 11).
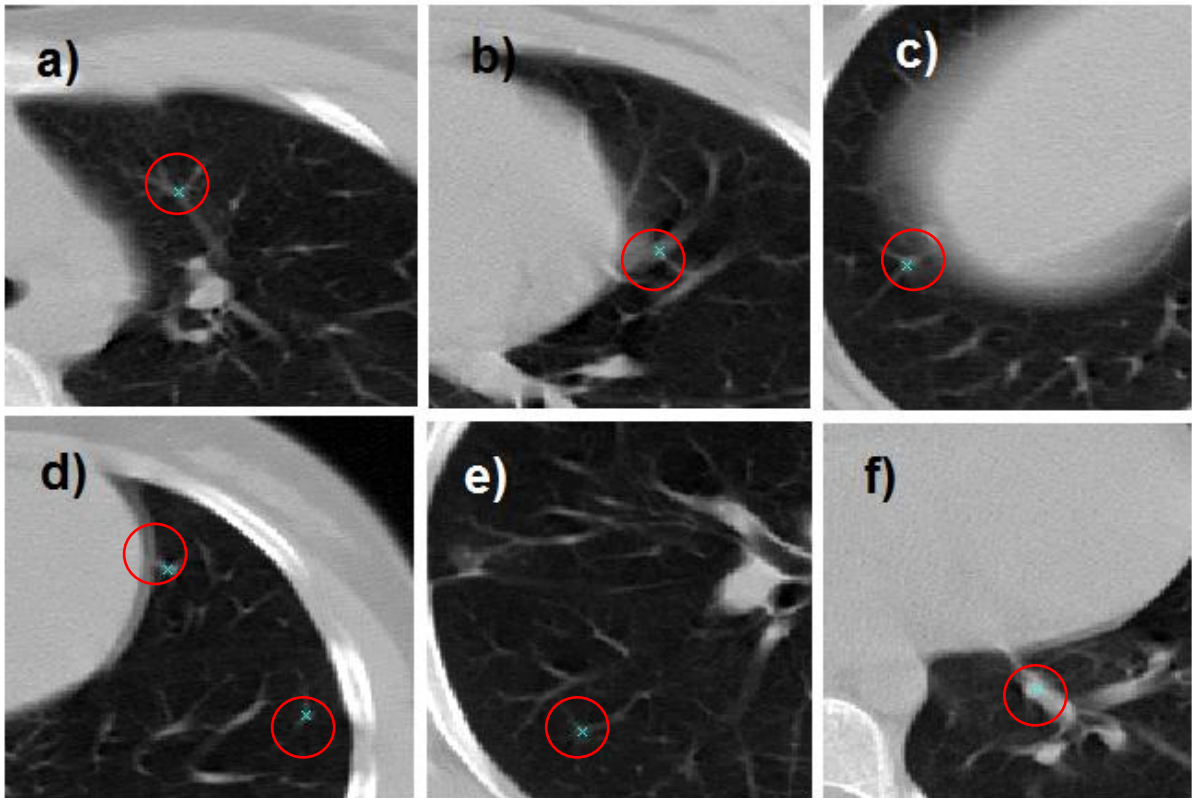
**Picture 11 – vessels a) vessel branching, PND002 (330,240,68) b) PND002 (386,246,118) c) PND002 (79,306,178) d) perifery vessels PND002 (409,197,159), (459,262,159) d) PND075 (117,350,111) f) central vessel PND004 (176,286,45)**

## Thin structures

The rest of detected structures have the thin shape in common, they could be long or short and they are mostly fibrous bands, adhesions, scars or interlobia. They are usually attached to the lung parenchyma border. The help to differentiate these thin structures from real nodules could be the same as for avoiding the vessels (shape characteristics in all planes). Examples of thin structures are shown in Picture 12.
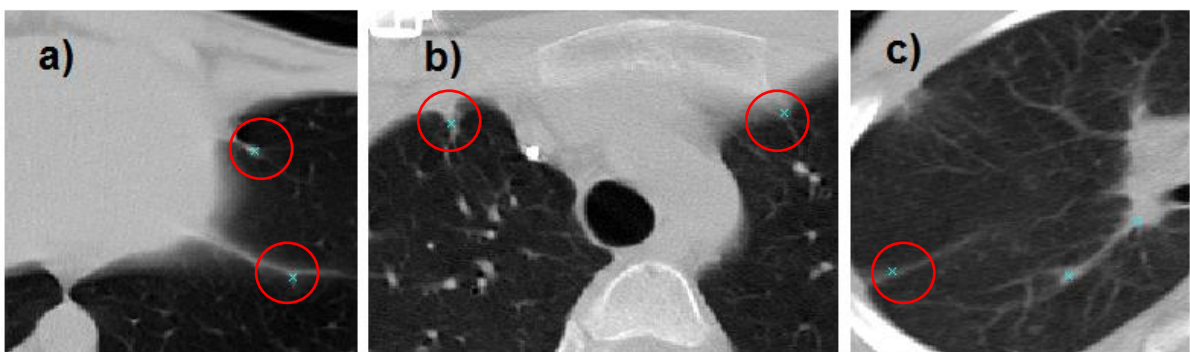


**Picture 12 – thin structures examples a) pericardial, PND007 b) pleural, PND090 c) PND004**

Usually there are two lung lobes in the left lung and three in the right lung, between the lung lobes we can find interlobia. There is one interlobium in the left lung and two or three (when a variant accessory lobe is present) in the right one.



Picture 13 – interlobia a) two detections of the left interlobium, PND007 (384,187,227), (400,177,227) b) PND049 (285,266,107) c) diaphragm, PND002 (411,276,169)

Bundles connecting pericardium to the posterior surface of the sternal bone (sternoperidcardial ligaments) were repeatedly detected as nodules (Picture 14).
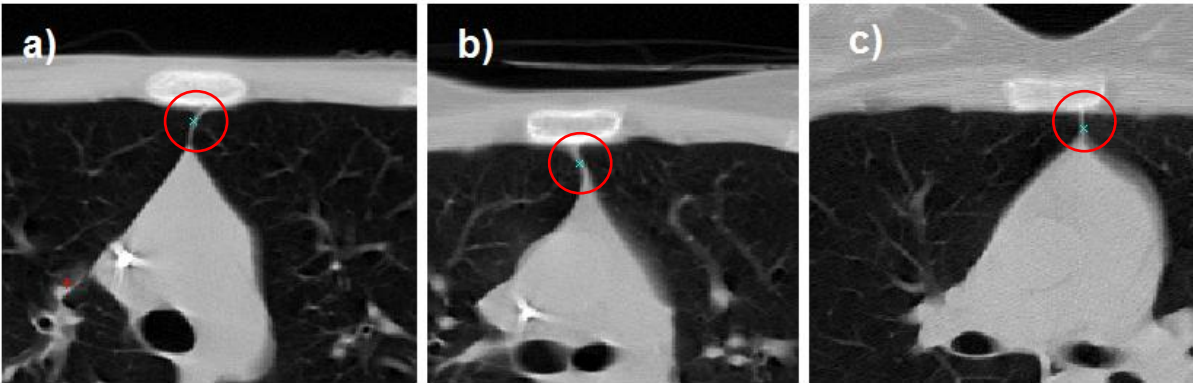


Picture 14 – sternopericardial ligaments a) PND007 (254,141,75) b) PND024 (260,180,71) C) PND041 (275,165,95)

Another mistaken structure was trachea (Picture 15).
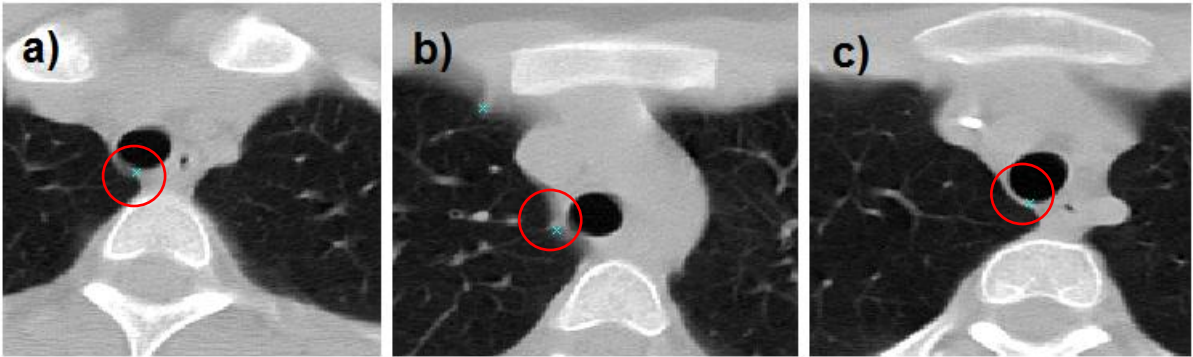


Picture 15 – tracheal ring a) PND002 (258,285,34) b) PND002 (241,226,54) c) PND024 (256,291,26)

### 4.1.5 Repeated detections

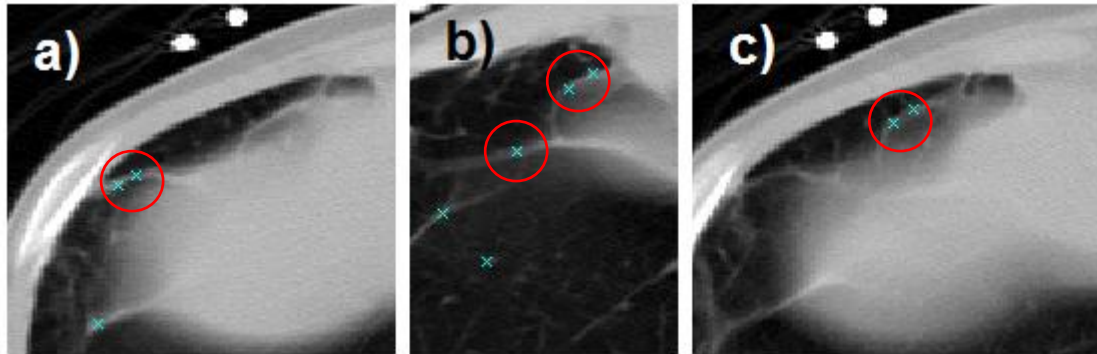Some structures were detected repetitively. These were usually fibrous bands, interlobia or other long thin structures. These multiple detections also raise the number of FPs.



Picture 16 - Multiple detections of the same structure catched on one slideInteresting findings

### 4.1.6 Shortcomings found in GT data

Observing the scans I have found some suspicious structures not marked as GT nodules. These structures, even they are maybe not the typical nodules, deserve an attention. They have to be detected and at least they will deserve further observation.

*Big parahilar nodule-shaped masses*

In the patient PND051, there were 72 GT nodules, the algorithm found 61 of them, it missed 11. There were many nice „exemplary" nodules, but there also were the big nodule-shape masses in both hila, which were not signed as a GT nodules. The algorithm detected each of this „big ball" even more than 10 times. In any case it is necessary to find these big structures. They could be large lymphatic nodes, that often indicate a serious pathology (Picture 17).
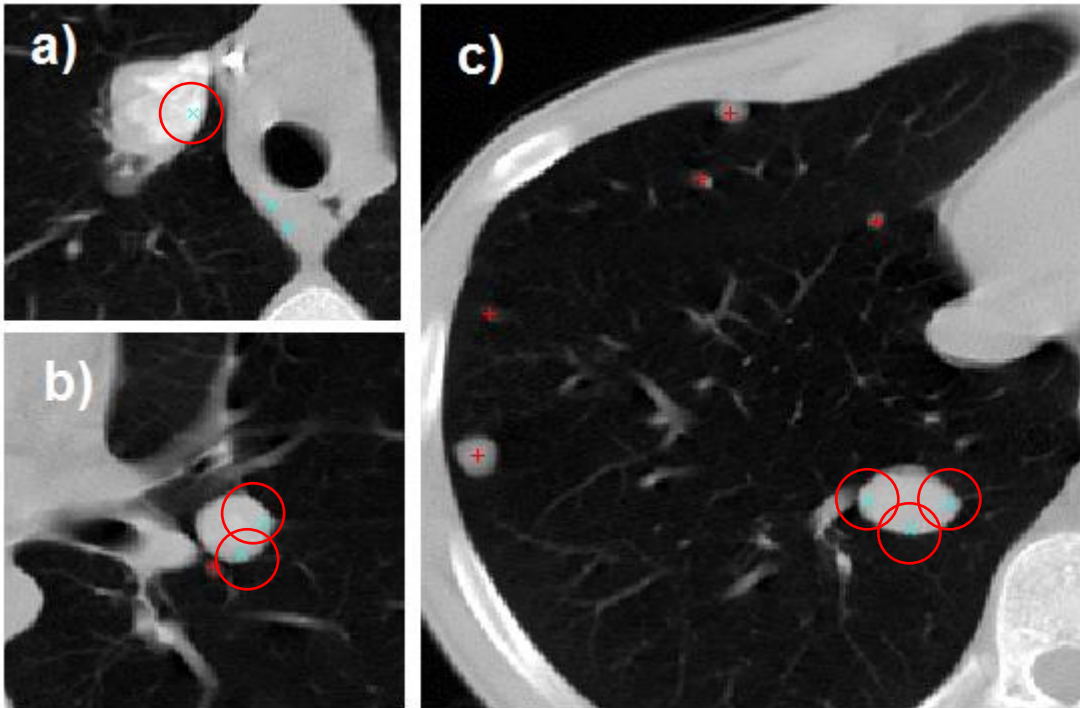
**Picture 17 - big parahilar nodules in PND051 a) also the dorsal mediastinal structures are sign as nodules b) two detections of the big nodule with no GT sign, however there is a smaller GT nodule close c) three detections of the big nodule in one scan, other smaller GT nodules in perifery (they were also detected in other slices)**

## Other suspect pathologies

The other mentioned pathologies can represent the nodular hypoventilations, pleural nodules, postspecific changes and others. In any case they all are suspicious and should be noticed (Picture 18), (Picture 19). It also depends on the number of these structures. The judgement will differ whether there are more pathologies or if the whole rest lung parenchyma is normal. This decision is sometimes hard even for the human expert and the judgement differ based on the experience of the radiologist.
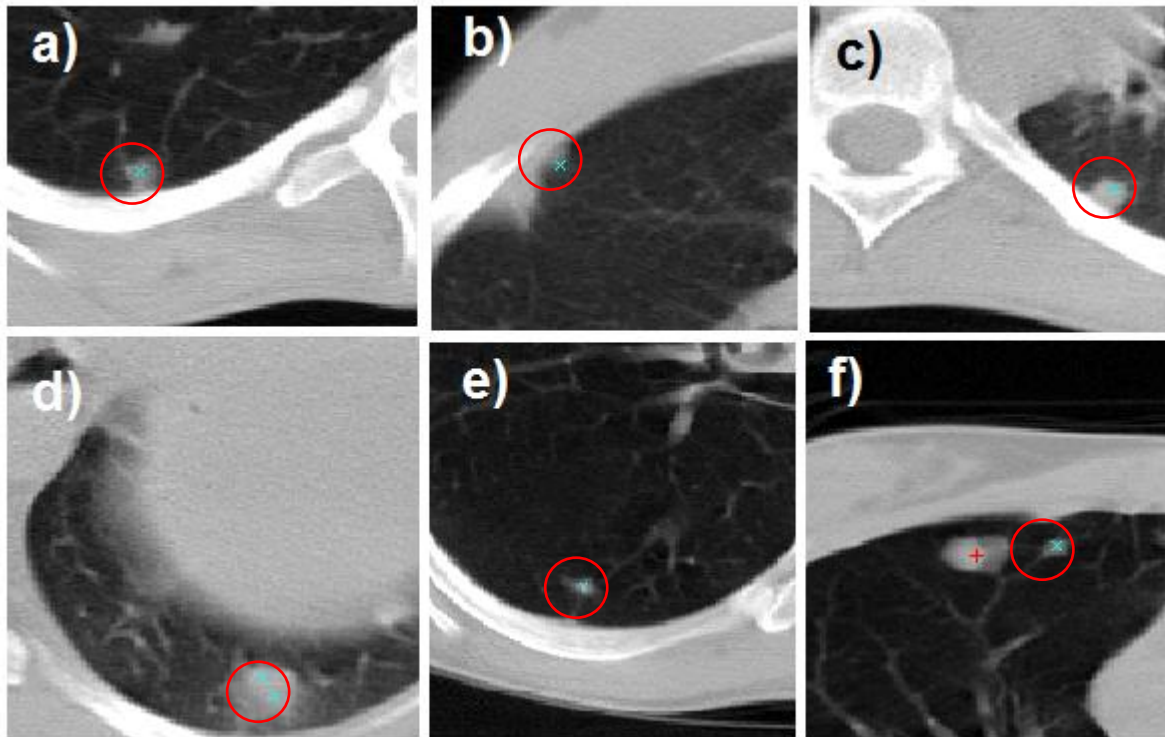
**Picture 18 – FPs detected structures not signed as GT nodules, but suspect to be problematic a) subpleural nodule, PND004 (133,396,52) b) pleural changes, PND004 (81,260,111) c) subpleural nodule, PND004 (302,401,113) d) nodular structure situated on the basis of the left lung, PND004 (365,392,186), (369,399,186) e) small oval nodule, PND024 (137,380,79) f) small nodule close to the bigger one, which was also detected, PND024 (193,163,126)**
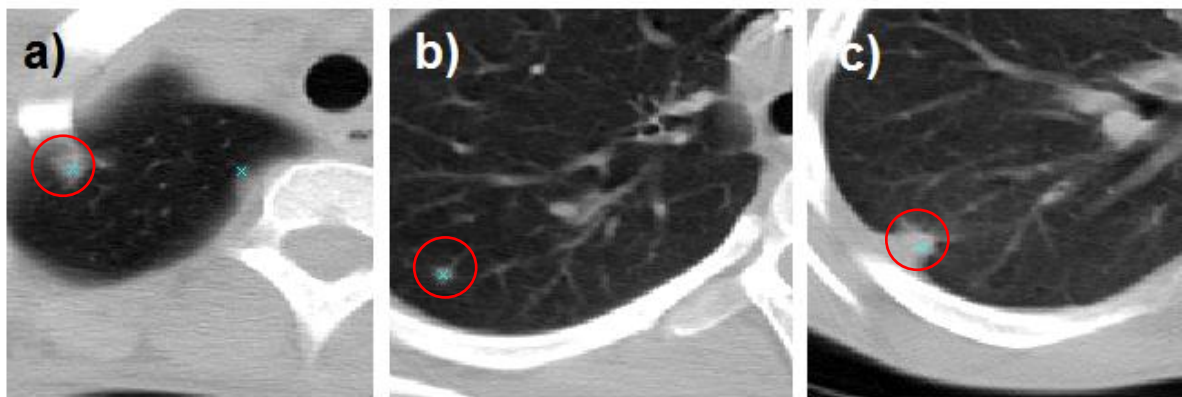


**Picture 19 – FPs detected structures not signed as a GT a) apical plastic pleural changes or a nodule? PND004 (134,335,23) b) small nodule? PND004 (93,388,72) c) postinfiltrative changes, pleural mass, hypoventilation or a nodule? PND004 (64,386,126))**

## *Lung segmentation*

Some of the scans I have checked contain nodules, but I have found no pleural effusion, no infiltration, no large tumourous processes or post-lobectomies, no lung fibrosis or other common pathologies that could lead to missclassification. I only found the paramediastinal condensation (atelectasis or infiltration) in the right lung and a big (tumorous or better atelectatic) basal pleural mass (Picture 20). I was surprised these pathologies did not increase the number of FPs at all and FP rate didn´t differ from the others scans. This may

reflect the good segmentation process when lungs are separated from other anatomic structures. On the other hand if the important pathological process is segmented away, it could be missed.
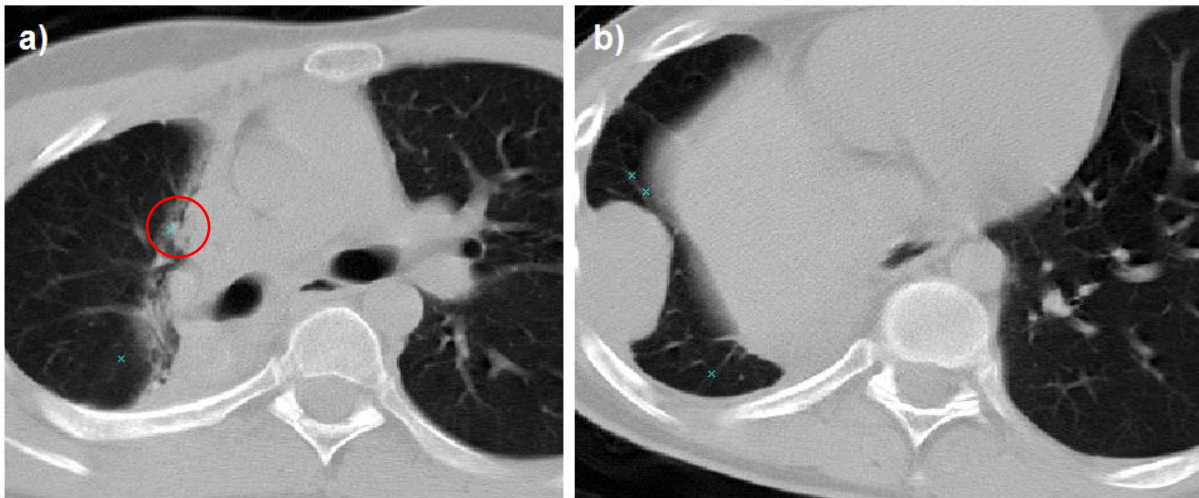


**Picture 20 – PND051 a) large paramediastinal condensation in the right lung and only one detection of this dense mass, probably it was segmented away b) big dense pleural mass in the basis of the right lung with no false positive detection**

### 4.1.7 Detection results

Next to the typical shortcomings of all CAD systems such as vessels, fibrous bands, scars and motion artifacts (26) our FPs were also related to the tracheal ring, mediastinal structures or rib margins. A recent study demonstrated, that radiologist will accept 11% of FP marks in nodule candidates they are not sure being a nodule (27). Radiologists will not miss any big nodule (>5 mm) as well as they will not accept a FP mark of an evident non-nodule candidate. This follows, that only FPs of small nodule-like structures can influence the radiologist´s judgement – maybe we should keep these FPs.

In every case, we need to get rid of all typical and unnecessary FPs. Detecting mediastinal and hilar structures could be fixed by improving the segmentation method. Counting the shape characteristics could be used to differentiate between nodules and vessels or scars. Typical nodular structures were found among the FPs too, this finding resulted to the GT data correction (7.2).

## 4.3 Analysis of the structure segmentation

The algorithm classifies detected candidates according to their characteristics (features). Features are computed for every candidate based on the segmentation process, that provides us the information which pixel belongs to the area of interest and which one is „outside". Segmentation process uses thresholding method and produces a binary segmentation mask. Correct segmentation is the first step to get correct feature values that fit the reality.

### 4.3.1 Data segmentation

For each detected nodule candidate a set of features is computed. Some features (such as size, density, volume) are computed from the whole detected structure using the „individual" segmentation mask for each candidate. Some are computed from selected areas using the „universal" segmentation masks (ballmasks, masks for the detection point, annulus, octants). The aim of segmentation analysis is to find out if the segmentation masks reflect the reality and features are computed from the well segmented area. All masks were received from the original algorithm in „*getNodFeatures.m*" script. All the masks are 3D, but in following pictures only 2D axial planes are displayed.

### *Universal masks*

Examples of binary „universal" masks are drawn in the cutouts. There are universal masks of balls, annular rings and octants used (Figure 1, Figure 2). All the various ball and annular ring masks have the centre at the detecion point and differ in diameter ($d_1$=2mm, $d_2$=4mm, $d_3$=8mm for balls, $d_1$=2-4mm, $d_2$=4-8mm for rings). The octant masks are counted for 8 regions that are created dividing the three-dimensional space by planes *x,y,z* with the detection point in the centre *D(0,0,0)* (Picture 21).
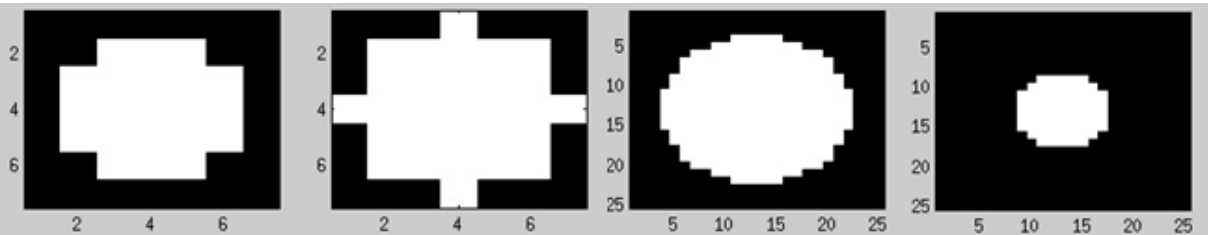


**Figure 1 - examples of „universal" binary ballmasks of various diameter values used for segmentation**
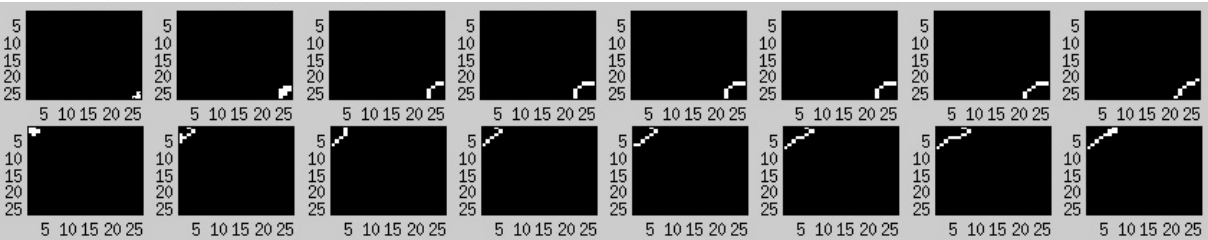


**Figure 2 – exmaples of „universary" binary octant masks**
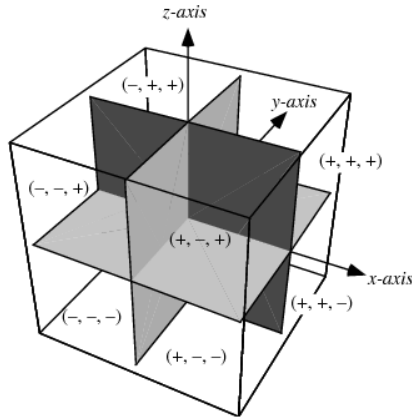
## *Individual masks*

The „individual" masks (Figure 3, Figure 4) are visualized in slices together with the corresponding anatomic structures. For tracing boundaries (boundaries are drawn as red curves) a MATLAB function *bwtraceboundary* is used.
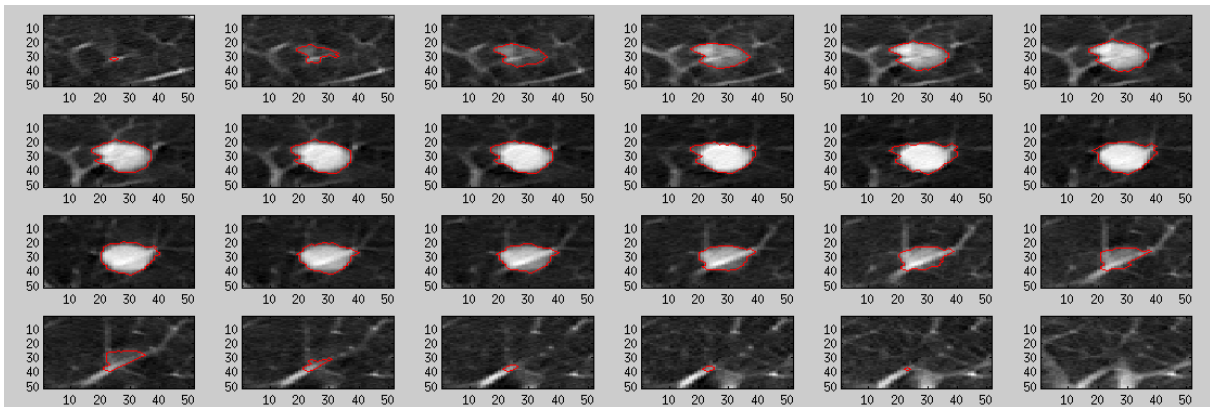


**Figure 3 - shows 24 axial slice example of the detected real nodule and its segmentation mask in each slice, the red curve defines the binary mask boundaries**
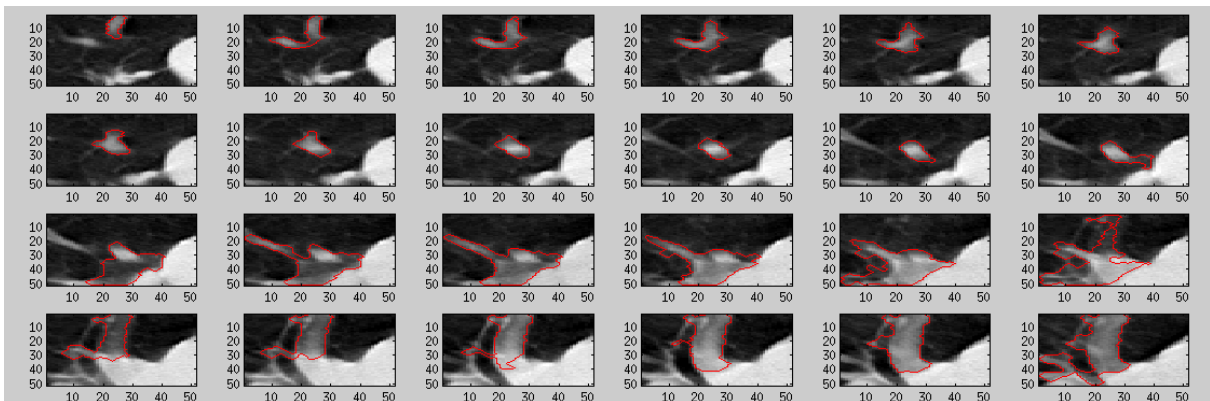


**Figure 4 - picture shows 24 axial slice example of the detected non-nodule structure and its segmentation mask in each slice, the red curve defines the binary mask boundaries**

---

[2] Provided by: http://mathworld.wolfram.com/Octant.html

Another examples show well and bad segmented candidates. Nearly all the real nodules are segmented well (Figure 5). Nodules in a) and d) and their masks have nearly rounded shape with low variance in radius in all axes while non-nodules in b) and c) have irregular shape with big radius differences. Features using individual masks are computed from areas defined by the segmentation mask.
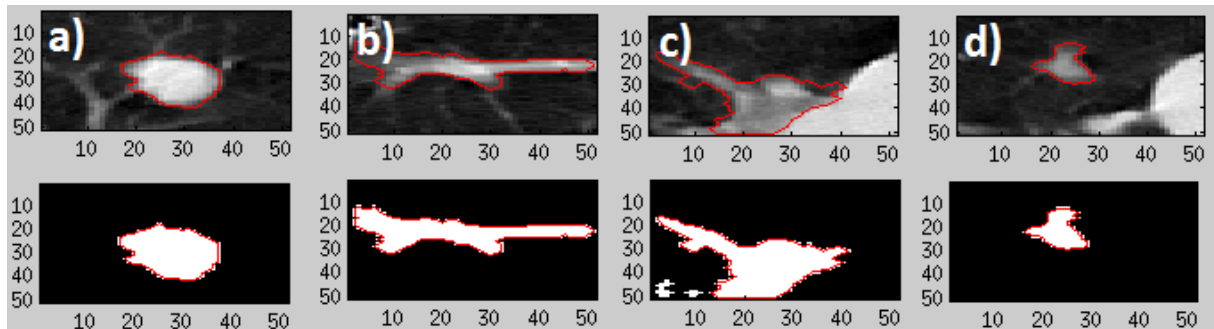


Figure 5 – 4 randomly chosen good segmented exaples of the detected nodule and non-nodule structures (above) and their segmentation binary masks (under), both with the segmentation curve

Nearly all bad segmented candidates are non-nodules (Figure 6). When looking on the pictures showing the real image of the candidate (above), there is no doubt they all are non-nodules. But when looking at the segmentation binary masks of each candidate (under), nearly all, but at least b) and c) have a typical „nodular-shaped" mask and are easily mistaken. These are also the typical FPs in our data – vessels and lung margins.
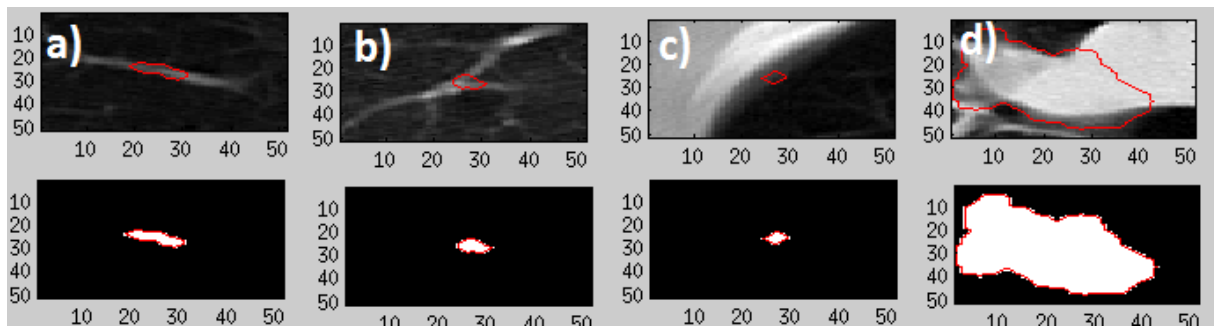


Figure 6 – 4 randomly chosen bad segmented examples of the detected structures, non-nodules in all cases (above) and it´s segmentation binary mask (under), both with the segmentation curve

### 4.3.2 Segmentation analysis results

Data segmentation for the real nodules and also for most examined non-nodule structures correspond to reality. But there were also bad segmented structures, mainly non-nodules, whose segmentation mask doesn´t fit the real shape. Nodules joining other dense structures (mostly in subpleural, or paravascular locality) were mostly problematic. These problems relate to the segmentation process which is based on thresholding. Because of this segmentation imperfections neither the computed features reflect the reality. These candidates can be easily mistaken for nodules and generates the majority of FPs.

Described segmentation problem is gone when universal masks are used. The universal masks with the centre placed to the detection point (ballmask, annulus ring mask) fit all the candidates. They could be helpful especially when the nodule candidate is attached to the border of the lung, or if it is not perfectly segmented. We only have to be careful about the mask diameter (d=2mm and d=4mm fit the smaller nodules and d=8mm fit the bigger ones).

The octant masks are of disputable significance. When used, features are counted only from the part of the structure and from the out-of-nodule surroundings. This I find useless or even harm.

In further computation, I prefer universal masks and individual masks.

## 4.4 Feature analysis

### *List of computed features*

For each nodule candidate a feature vector including 482 features is computed in the baseline algorithm (28). Features describe the characteristics of detected candidates and help the algorithm classify them. They describe size, intensity, homogenity, shape, curvature and dimension on several scales, localization and distance to the boundary of the lung. In most cases mean, median, variation, maximal and minimal values are computed. Some characteristics (size and volume parameters) are computed from the outer ellipsoid obtained from the covariance matrix eigenvalues (24), the rest from the segmented areas. All types of segmentation masks are used.

| Feature number | Feature describing |
|---|---|
| 1-16 | Size, diameter, volume parameters |
| 17-58 | Intensity parameters |
| 59-184 | Curvature in 3 scales (σ=2,4,8 mm) |
| 185-310 | Shape index in 3 scales (σ=2,4,8 mm) |
| 311-436 | Dimension in 3 scales (σ=1,2,3mm) |
| 437-478 | Local scale (LSC) |
| 479-482 | Location |

Table 2 – short list of all the features[3]

Some of the feature computation is shown below:

An elipsoid *E* is fitted to each candidate and the **size** is expressed as the length of 3 half-axes of an ellipse *(a,b,c)* rotatet by angles *α, β* in 3-coordinate system.

Effective radius:
$$r = \sqrt[3]{abc} \tag{1.2}$$

**Intensity values** are representing the HU grayscale densities in real CT image. When *n* is the number of all voxels *N* within the nodule area.

Mean intensity:
$$I = \frac{\sum N \in E}{n} \tag{1.3}$$

When parameters *k1* and *k2* are the principal curvatures computed for each voxel in the segmented lung parenchyma, then **shape index** and **curvature** are computed as:

Shape index:
$$SI = \frac{2}{\pi} \arctan\left(\frac{k_1 + k_2}{k_1 - k_2}\right) \tag{1.4}$$

---

[3] **designed and computed by Dolejší M.**

Curvature:
$$CV = \sqrt{k_1^2 + k_1^2}$$
(1.5)

**Dimension** is estimated based on eigenvalues of Hessian matrix $|\omega_1| > |\omega_2| > |\omega_3|$ and tells us if the object is darker or lighter than the background.

Dimension:
$$DIM = \frac{\sum_{\forall i} \omega_i}{|\omega_1|}$$
(1.6)

**Local scale** is an invariant descriptor of shape and translation obtained from the eigenvalues of Hessian matrix.

### 4.4.1 Visual feature analysis

False positives are classified as nodules based on characteristics that are similar to their „nearest" real nodule neighbours from the training data. On the other hand false negatives are classified as non-nodules due to similarities to the non-nodule training samples. In some computed characteristics we can easily use the naked eye to qualify if the values reflect the reality or not and to estimate differences.

To verify, if the descriptors are consistent with visual assessment, I chose some examples of random FP detections and found first four nearest neighbours (in 482-dimensional feature space) from the TP data set for each FP. The observed features describing size, volume, density, shape and curvature were compared with the structure appearance.

Features that can differentiate presented FP and TP structures are shown in tables below the pictures. The original feature values are of different scales and units. To make the data more informative normalization process was performed. „Example 1" shows both original (orig) and normalized (norm) feature values. In examples 2, 3 and 4 only normalized values can be found.

### *Example 1*

The first structures (Picture 22) are the same density and almost all of them are localized subpleuraly, but they can be visually differentiated by shape. These structures did not significantly differ in observed features describing neither curvature or shape index, nor dimension. The biggest difference between FP and TPs was found in *the feature nb.16* describing the variance of radius, in the table below it is evident that this feature in FP is much higher then in TPs. This corresponds to the FP´s nonregular shape.
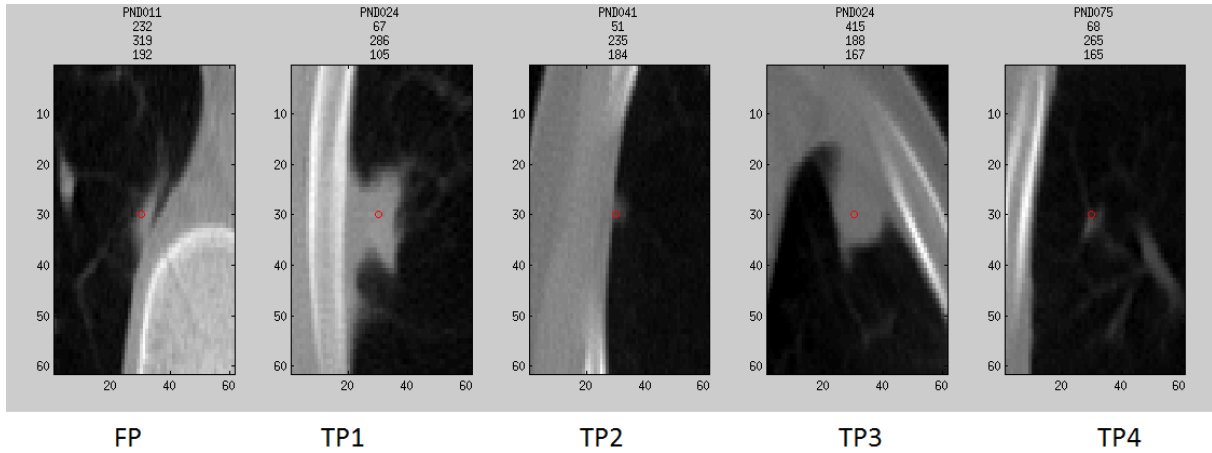
**Picture 22 - randomly selected FP and the four nearest TPs (example 1)**

| Feature nb. | FP | TP1 | TP2 | TP3 | TP4 |
|---|---|---|---|---|---|
| 16 - orig | **11.0437** | 0.2727 | 0.0417 | 4.3895 | 0.5664 |
| 16 - norm | **0.2145** | 0.0050 | 0.0006 | 0.0851 | 0.0107 |

**Table 3 – chosen feature values (example 1), original data (orig) and normalized data (norm) for feature nb. 16 (variance of radius)**

## *Example 2*

Another FP sample (Picture 23) is visually different in all observed features, mainly in size, intensity and curvature. Numerically it differs in mean intensity of the segmented structure (*feature nb.17*) and in the variance of intensities (*feature nb.18*), in mean curvature in all 3 scales (*features nb.59, 101, 143*) and also in variance of dimension in all 3 scales (*features nb.312, 354, 396*). TPs have higher curvature values and lower dimension variance. Surprisingly there was no significant difference in shape index.
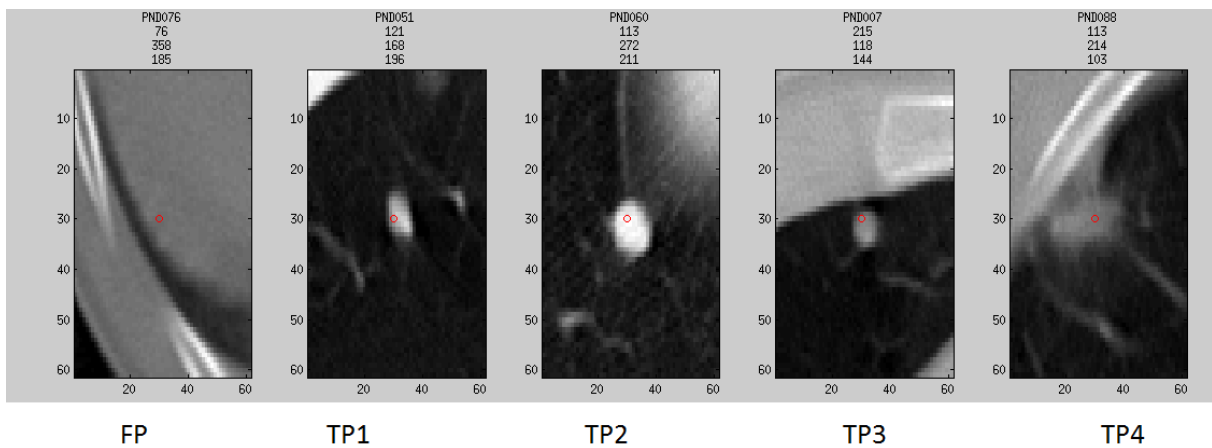


**Picture 23 – randomly selected FP and the four nearest TPs (example 2)**

| Feature nb. | FP | TP1 | TP2 | TP3 | TP4 |
|---|---|---|---|---|---|
| 17 | **0.3469** | 0.1515 | 0.1870 | 0.1412 | 0.2490 |
| 18 | **0.1192** | 0.0634 | 0.0913 | 0.0516 | 0.0584 |
| 59 | **0.3118** | 0.7402 | 0.4084 | 0.7770 | 0.4712 |
| 101 | **0.2960** | 0.8085 | 0.4787 | 0.5797 | 0.5747 |
| 143 | **0.4899** | 0.9017 | 0.6677 | 0.7710 | 0.7367 |

| | | | | |
|---|---|---|---|---|
| 312 | **0.6950** | 0.3159 | 0.4727 | 0.3656 | 0.5259 |
| 354 | **0.6568** | 0.0959 | 0.4102 | 0.5558 | 0.4362 |
| 396 | **0.4275** | 0.0467 | 0.3422 | 0.0222 | 0.2999 |

**Table 4 - chosen feature values (example 2)**

## *Example 3*

Structures in another picture (Picture 24) visually differ in size and shape. Numerically they can be differentiated mainly by radius variety and shape index values. The mean shape index (*feature nb. 185, 227*) and shape index variance (*feature nb.186*) is lower in FP, while the variance of radius (*feature nb.16*), curvature (*feature nb.60*) and dimension (*feature nb.396*) is much higher. I find interesting that even if the nodules TP1, TP2 and TP3 look similar, there is quite a big difference in shape index value and curvature variance, the other features don´t differ at all.
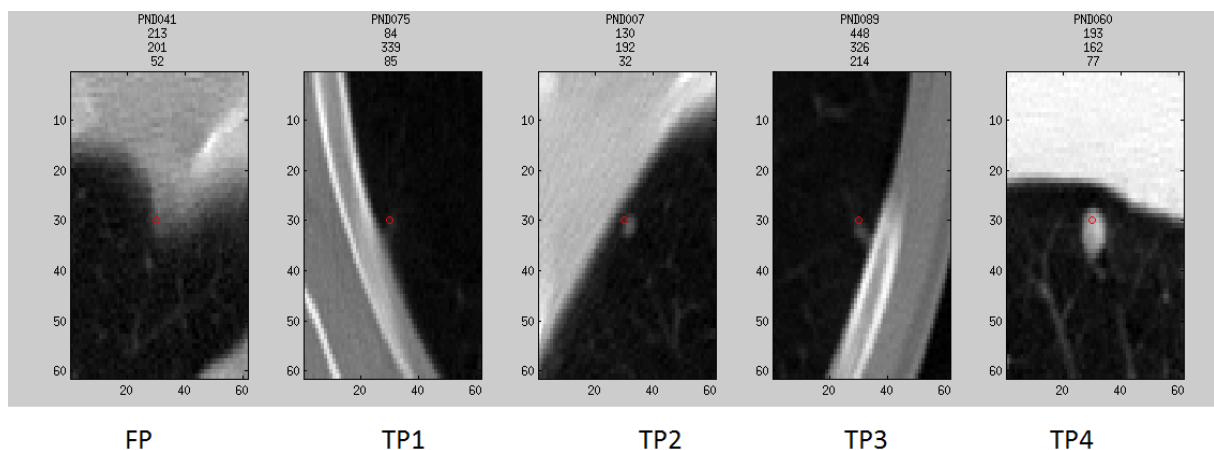


FP TP1 TP2 TP3 TP4

**Picture 24 - randomly selected FP and the four nearest TPs (example 3)**

| Feature nb. | FP | TP1 | TP2 | TP3 | TP4 |
|---|---|---|---|---|---|
| 16 | **0.1387** | 0.0079 | 0.0001 | 0.0001 | 0.0395 |
| 60 | **0.5071** | 0.0940 | 0.1791 | 0.2877 | 0.0839 |
| 185 | **0.0441** | 0.1055 | 0.1960 | 0.0025 | 0.0003 |
| 186 | **0.0001** | 0.0006 | 0.0023 | 0.0025 | 0.0003 |
| 227 | **0.0803** | 0.1197 | 0.1313 | 0.1702 | 0.3422 |
| 396 | **0.1495** | 0.0275 | 0.0124 | 0.0077 | 0.1167 |

**Table 5 - chosen feature values (example 3)**

## *Example 4*

Last example (Picture 25) shows FP (vessel in this case) that visually differs in curvature and shape, it is nearly the same size and density as all the TP nodules. Contrary to the visual parameters, the computed values differ mainly in size parameters – in volume (*feature nb.10*), surface (*feature nb.13*) and radius variance (*feature nb.16*), another differences are in variance of dimensions (*features nb.312, 354*). The values for size parameters are much higher for FP and it seems to be tubular shaped and much bigger than all the TPs. There is simple explanation for this – we only check the one (axial) plane and have no visual

information about z-axis. This finding resulted in need of multiplanar imaging (3D) in further investigation (6.2).

Checking these 5 structures we can also notice, that they are all situated in perifery of the left lung not far from the dense rib structure. It is clear that the localization and surroundings of the nodule candidate is not as important as its density, shape and other features describing the nodule itself. This finding can lead to the same conclusion as in chapter (**Chyba! Nenalezen zdroj odkazů.**), to concentrate to nodule itself and don´t care for urroundings.
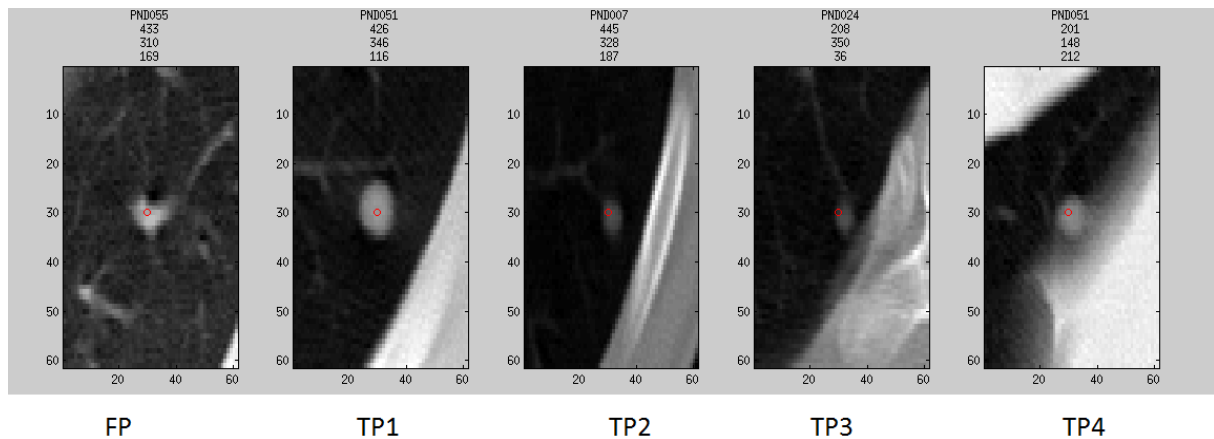


FP     TP1     TP2     TP3     TP4

**Picture 25 - randomly selected FP and the four nearest TPs (example 4)**

| Feature nb. | FP | TP1 | TP2 | TP3 | TP4 |
|---|---|---|---|---|---|
| 10 | **0.1006** | 0.0252 | 0.0099 | 0.0125 | 0.0435 |
| 13 | **0.3182** | 0.0452 | 0.0230 | 0.0302 | 0.1096 |
| 16 | **0.4079** | 0.0231 | 0.0267 | 0.0716 | 0.3003 |
| 312 | **0.6181** | 0.0989 | 0.1155 | 0.5223 | 0.6182 |
| 354 | **0.6893** | 0.0481 | 0.3839 | 0.3242 | 0.6391 |

**Table 6 - chosen feature values (example 4)**

## *Visual feature range*

Last visual analysis included sorting the detections in right order depending on the selected characteristics – volume (Picture 26, Picture 27), surface (Picture 28, Picture 29), mean intensity of the whole structure (Picture 30, Picture 31) and the intensity just in the detection point (Picture 32, Picture 33). Randomly chosen structures from the whole candidate nodule set and only from TP set were used. The results are shown in following pictures.
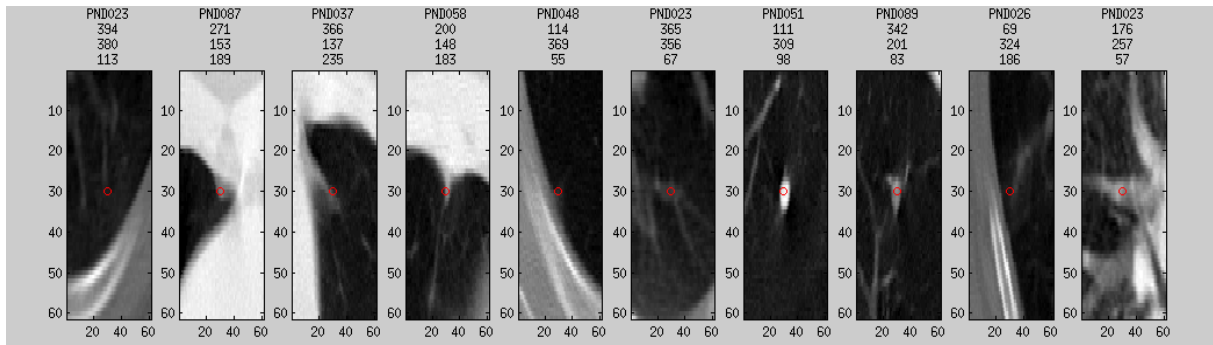
**Picture 26 – nodule candidates sorted by the volume**
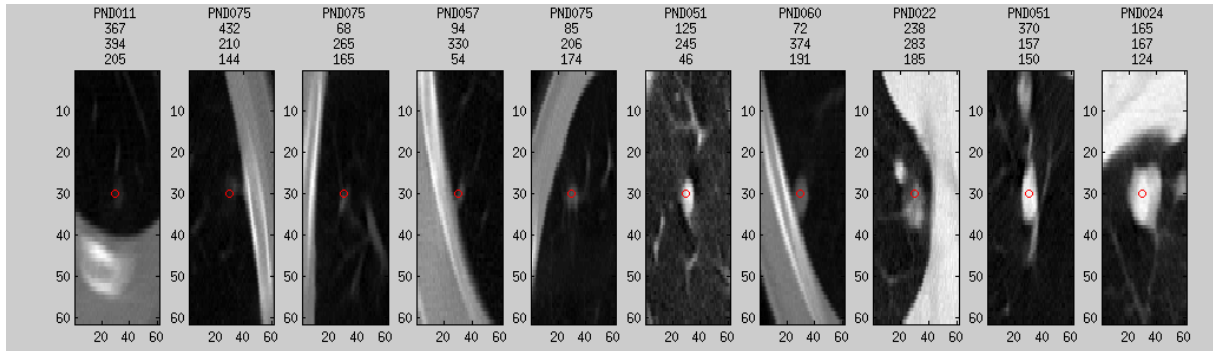


**Picture 27 – true nodules sorted by the volume**
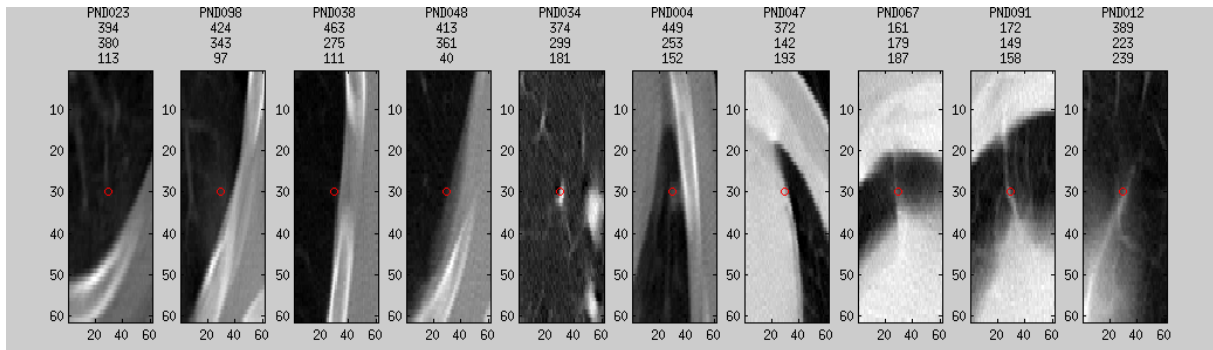


**Picture 28 - nodule candidates sorted by the surface**
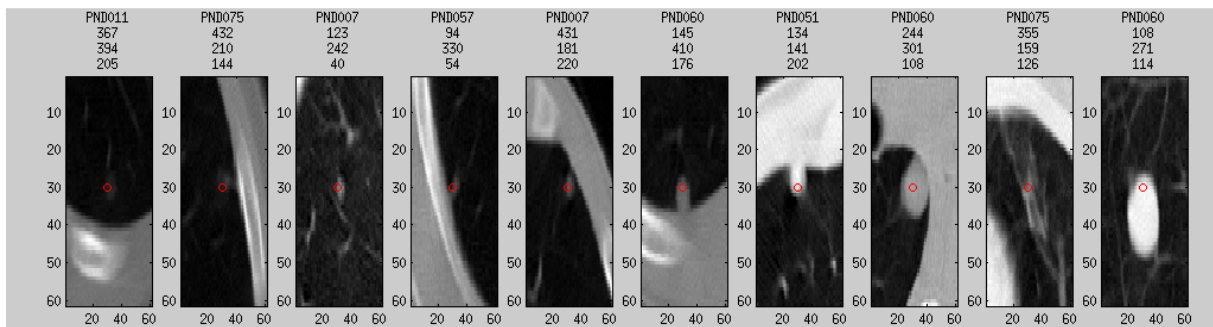


**Picture 29 – true nodules sorted by the surface**

41
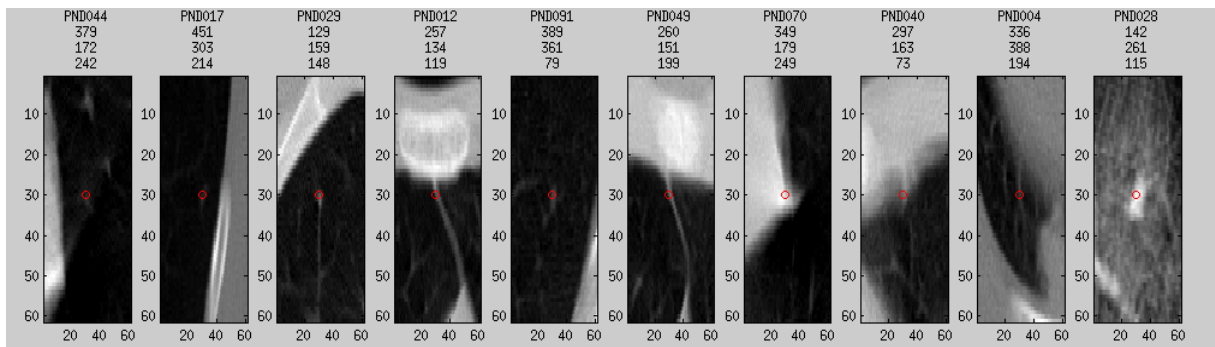
**Picture 30 – nodule candidates sorted by the mean intensity value**
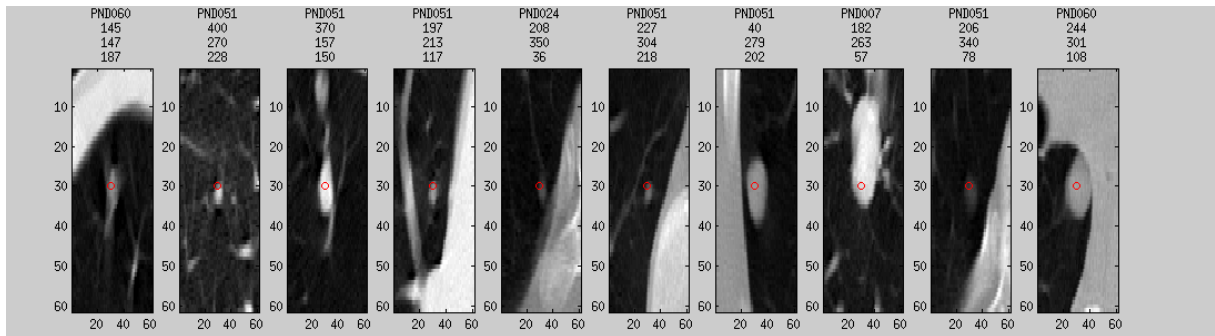


**Picture 31 – true nodules sorted by the mean intensity value**
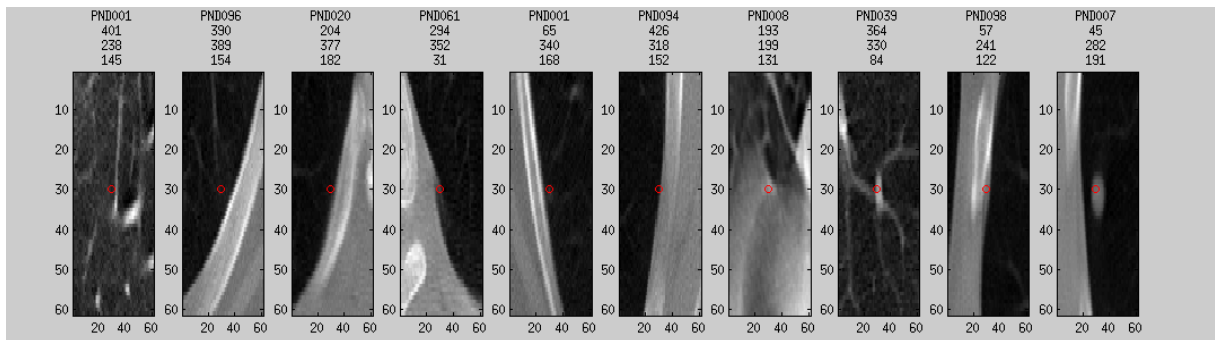


**Picture 32 – nodule candidates sorted by the intensity in the detection point**
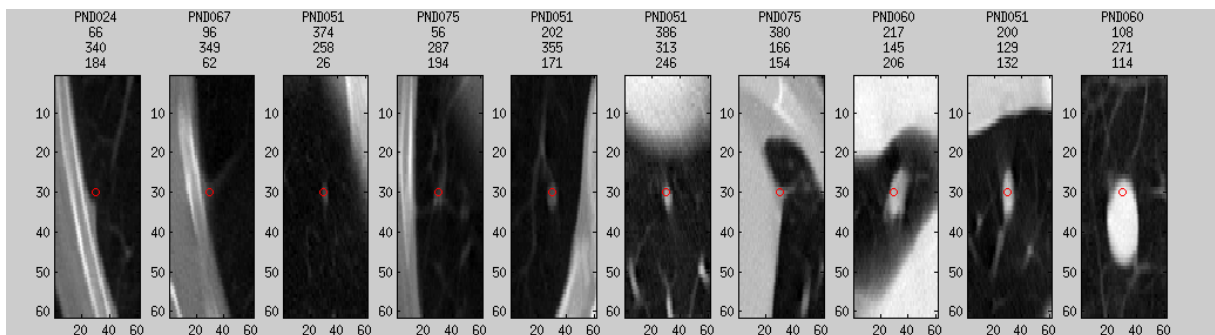


**Picture 33 – true nodules sorted by the intensity in the detection point**

It can be mentioned, that the intensity in the detection point describes the intensity value better then the mean intensity in the whole segmented structure.

### 4.4.2 Visual analysis results

Visual analysis verified, that most of chosen feature values are consistent with visual assessment, so the features are counted well. The „nearest" structures look similar, but differencies can be found. There are characteristics that can separate these examples and assign them to the correct class. The reason why the classificator did not use these „useful" characteristics may be that there are many other features which can´t differ the structures (for example candidate localization or measurements in structure surroundings).

In future we need to train only with effective descriptors to prevent misclassifications.

It is evident that only axial planes are not enought for visual valuating and we need multiplanar imaging.

### 4.4.3 Computed feature analysis

Another part of characteristics analysis is checking how they can estimate the inter-class diferences. I separated the detected structures into two classes *(nodule/non-nodule)* and for each feature I calculated histograms for both classes. Then I applied a simple thresholding method and for each feature found the ideal threshold (boundary), that maximizes the class separability. Once the threshold is found we can use it for predicting the class of new points by checking on which side of the boundary it falls. For each threshold the classification error *(err)* is computed *Separability =1-error*. It determines the success of data separation.

$$1 - err = \frac{\sum(nodules > threshold) + \sum(nonnodules < threshold)}{\sum nodules + \sum nonnodules} \qquad (1.7)$$

#### *Histograms examples*

Three characteristics well separating both classes are chosen in following pictures. Each picture consists of two charts presenting histograms (above) and classification error (below) in the whole feature range.

Feature value interval range is divided into 10 subintervals and histograms of both classes are displayed (nodules are in red and non-nodules in blue), the total number of all detections in each class is considered to be 100%. X axis shows intervals, y axis shows how much of the total sum is represented by this subinterval group (see above charts).

Line-plot representing the classification error in y axis for each feature value in x axis was created (see below charts).

Numbers between the charts show the minimal found classification error and the feature value that fits this error.
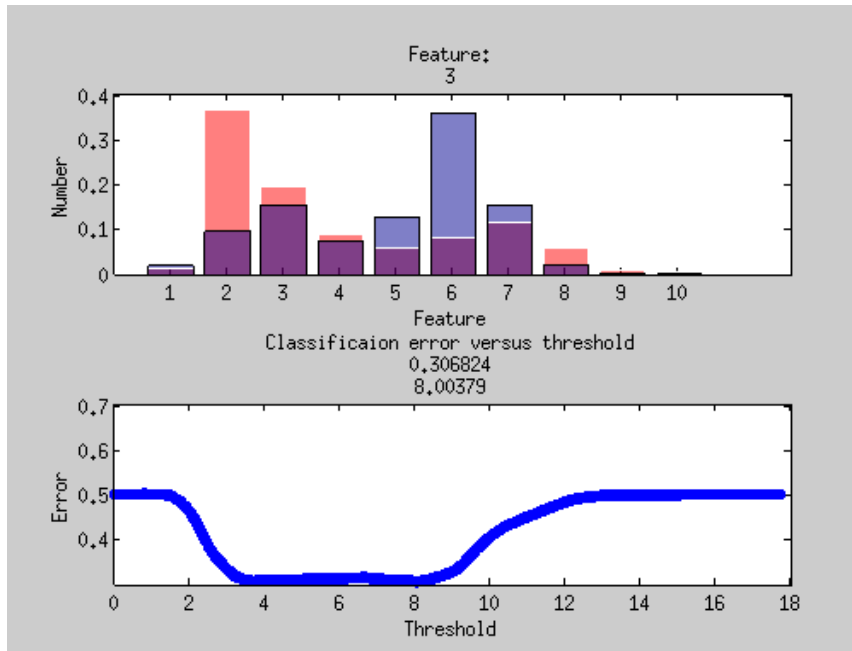
**Picture 34 – data distribution (nodules=red, non-nodules=blue) for the feature nb. 3 (maximal eigenvalue for segmented nodule candidate), minimal classification error err=0.31 for threshold value thr=8.00 was found**
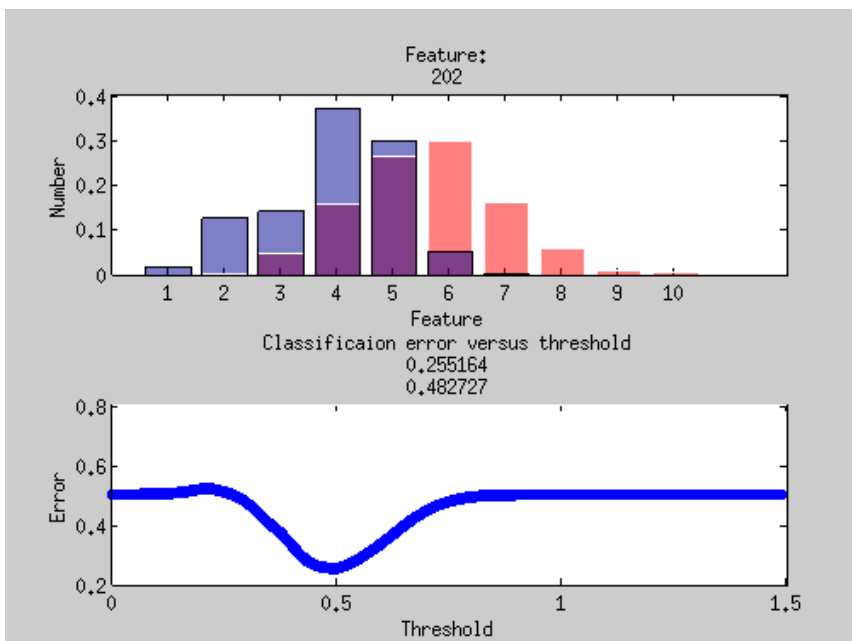


**Picture 35 - data distribution (nodules=red, non-nodules=blue) for the feature nb. 202 (mean shape index for ballmask with diameter d=4 mm), minimal classification error err=0.26 for threshold value thr=0.48 was found**
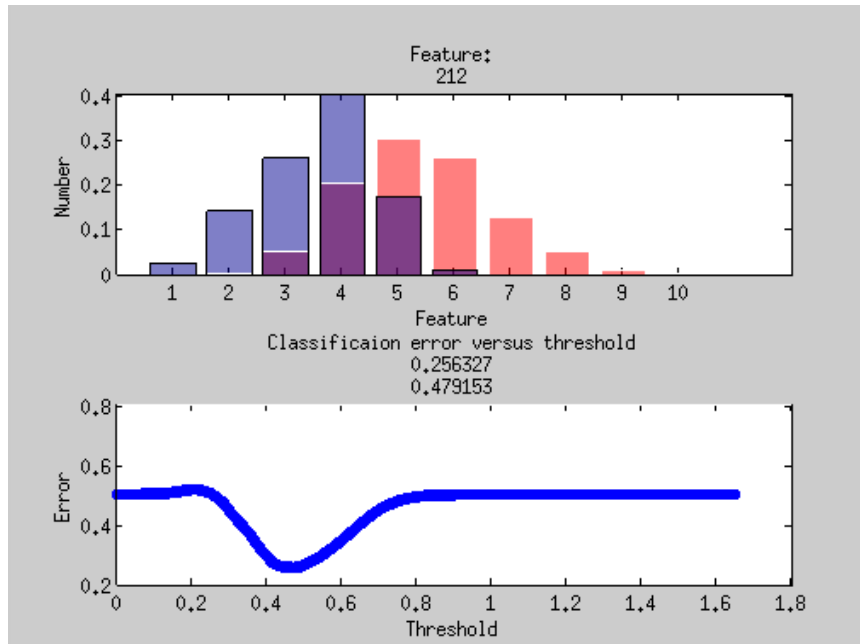
**Picture 36 - data distribution (nodules=red, non-nodules=blue) for the feature nb. 212 (mean shape index in annulus with diameter d=4-8 mm with the centre in the detection point), minimal classification error err=0.26 for threshold value thr=0.48 was found**

This analysis helped us find out, how the single feature can separate the data. We can use this method later to select required characteristics and to remove the redundant.

## *Selecting features*

Table 7 shows how many features remain in the feature matrix when different thresholds for the separation error is set. There is only one feature with classification error *err<0.25*, this is the feature *nb.229* which describes the shape index in the annulus ring mask of diameter d=4-8 mm.

For *err<0.3*, there are 16 features describing mainly mean and maximal shape index values (mean and maximal shape index value for the whole segmented structure, mean and maximal shape index for the segmented ball with diameter d=2 mm and d=4 mm and the same characteristics for segmented annulus ring with dimeter=4-8 mm and d=2-4 mm) and maximal dimension for the whole segmented structure on all 3 scales.

For *err<0.33*, features describing size are added (eigenvalues for segmented nodule, effective radius, volume, surface, variance of diameter.

| Class. Err | <0.25 | <0.3 | <0.33 | <0.35 | <0.4 | <0.45 | <0.5 | ≤0.5 |
|---|---|---|---|---|---|---|---|---|
| Features | 1 | 16 | 60 | 81 | 145 | 207 | 452 | 482 |

**Table 7 – number of features in the feature matrix for various classificatio error value**

There are 30 features with the highest classifacation error *(err=0.5).* They are only the features describing the curvature values in balls or annuli and local scale (LSC) values.

45

Displaying histograms of these characteristics, the data for both classes overlaps heavily. These features are not able to separate data well.

The results correspond to the segmentation analysis, there are group of features separating the data well (shape index, variance of radius, size) and the group of features separating the data bad (LSC values, or features counted using octant masks).

# 5   PROPOSED METHOD

The whole algorithm consists of two main parts, nodule candidates detection and nodule candidates classification respectively. The process of candidates detection is completely adopted from baseline method (25) with no changes in algorithm. The detection process was only analysed (see previous chapters). My own work is the data classification part. It was processed independently on the baseline method. The classification process consists of three steps: data preparation, data dimensionality reduction and finall classification.

## 5.1   Dimensionality reduction

Reducing the dimensionality leads to noise reduction and removing the redundant attributies. Dimensionality reduction methods select features or creates new features based on the existing ones. Both methods solve the problem of large number of variables, which could lead to overfitting and lack of memory space.

### 5.1.1   Feature selection

Selecting the subset of relevant features is called the „feature selection" method. It sets a subset of features that are capable to place the sample to the right class. In supervised learning (learning on the labeled dataset) it is easy to find these features. Feature selection shorten the training time and could prevent overfitting. To select features we can use wrapper, filter, embedded or hybrid models (29). I chose the filter model, which selects the features with the highest score. Each feature was evaluated independently (univariate feature evaluation).

### *Chosen features*

The features were chosen based on the value of the classification error described in (4.4.3). Training dataset of nodule containing scans was created and sensitivity and specificity rates were couned for all error thresholds ($err = 0.25 \rightarrow 0.5$) in this dataset. The threshold with best results (err<0.3) was chosen. Using this threshold, 16 features remain in the feature matrix. For counting the sensitivity and specificity values for comparing the feature subsets, the kNN classifier with *k=100* was used.

Picture 39 shows the sensitivity and specificity rates for 7 used classification thresholds [*0.250, 0.275, 0.300, 0.325, 0.350, 0.375, 0.400*]. X axis shows the threshold value, y axis the performance rate. The experiment was performed on all nodule containing scans (n=18). Specificity rates are stable, reaching 80-90%, but there are differences in sensitivity rates. It is evident, that best sensitivities are reached for err=0.3, with no sensitivity drop under 45%.

Next two pictures are showing the same comparison of performances, but only for one chosen scan (PND060). The threshold values are of various value range.
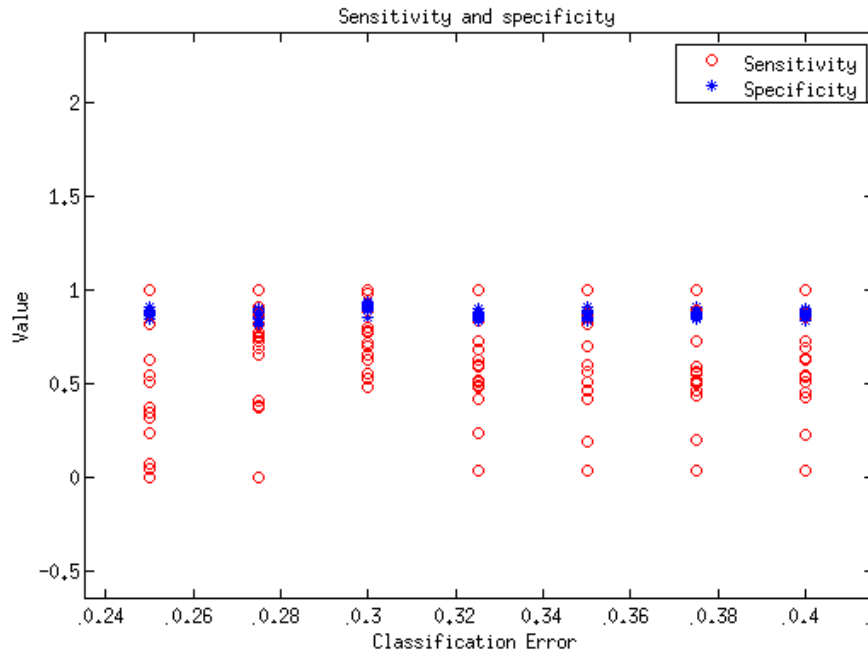
Picture 37 – sensitivity and specificity rates achieved when using different values of classification thresholds (at interval 0.25-0.4), performed on each of 18 nodule containing scans, kNN classifier with k=100 applied, threshold set at 0.3 has significantly the highest sensitivity values than the others
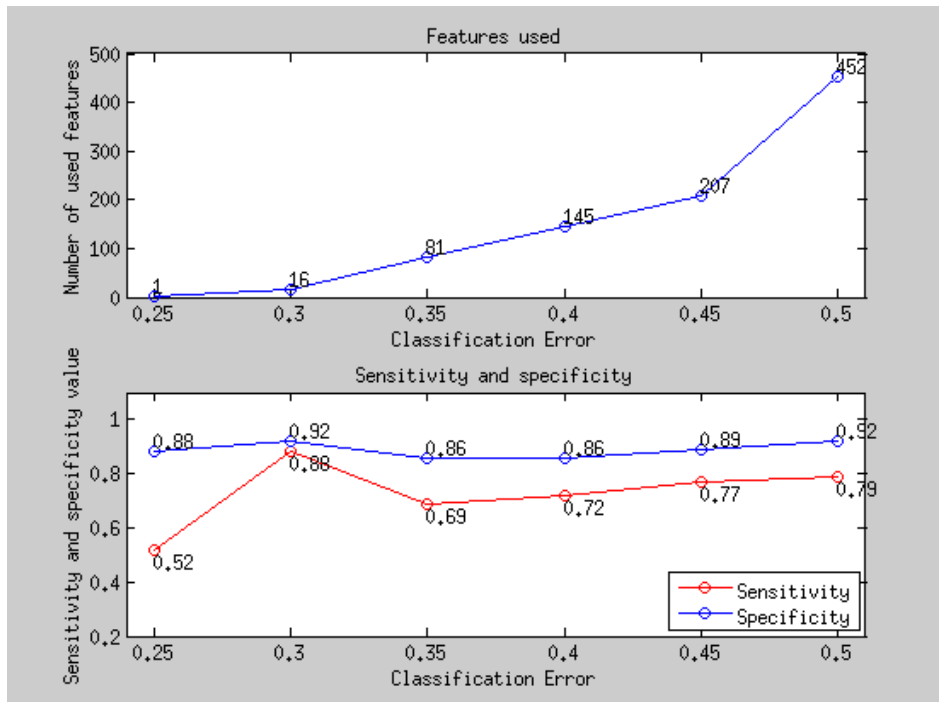


Picture 38 – number of features used for each classification error rate in the range 0.25-0.5 (above), comparison of sensitivity and specificity rates (under), data for PND060
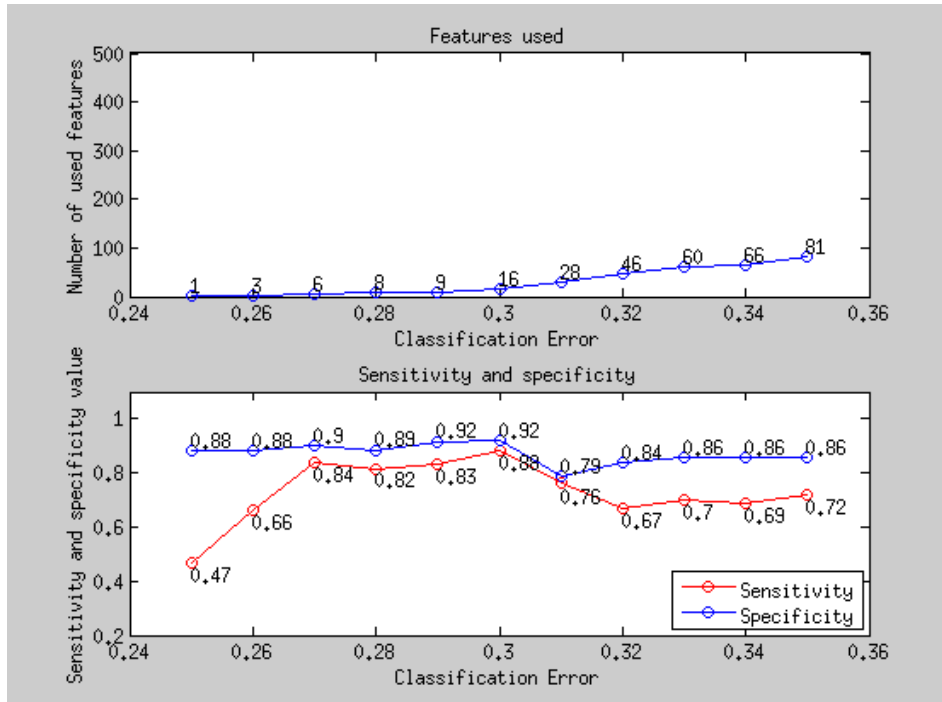
**Picture 39 - number of features used for each classification error rate in the range 0.25-0.36 (above), comparison of sensitivity and specificity rates (under), data for PND060**

### *Independent significance feature test*

Significance feature test is a simple test, which removes obviously useless features. Sometimes it is used as pre-processing phase to reduce the number of characteristics. This method was invented by Weiss and Indurkhya (30) and the Fisher´s discriminant ratio is used to count the significance power of each feature. Features are sorted according to their power and only the most powerful ones are chosen (31).

$$Sig = \frac{(mean_1 - mean_2)^2}{var_1 - var_2} \tag{1.8}$$

## 5.1.2 Principal component analysis

Principal Component Analysis (PCA) is a method used for reducing the dimension of characteristics and when we want to get rid of relationships among variables. It creates a new set of features based on the existing one.

The principal components can be found by calculating the eigenvectors and eigenvalues of the data covariance matrix after removing the mean from each sample *Xi* (it is supposed the sum of the samples is zero). *Zi* is the approximation of *Xi*. Residuum *res* should be minimal.

$$\sum_i X_{i=1...k} = 0; \quad X_i \in R \tag{1.9}$$

49

$$Z_i = \sum_j w_{ij} Y_j ; \quad Y_i \in R$$

$$res = \sum_i \|X_i - Z_i\|^2$$

The eigenvector with the largest eigenvalue corresponds to the direction of largest variation (32). The total variance is defined as a sum of the variances of all components and could be expressed as a sum of all the eigenvalues.

PCA finds a direction that corresponds to maximal variance among the data points, then it rotates the original data to the new coordinate space (the space is given by this principal direction) and it generates new feature matrix with the same size as the original one where the features are reordered based on the decreasing variance (33).

### Chosen features

Features were chosen based on their data variance description. The variance threshold was set as 1% and all features describing more than 1% of the data variance were chosen into the new reduced feature matrix. Using this criterium, 15 features create the new feature data. In Picture 40 features are sorted by their ability to describe the variance of the data (in %). Best 50 features are shown.
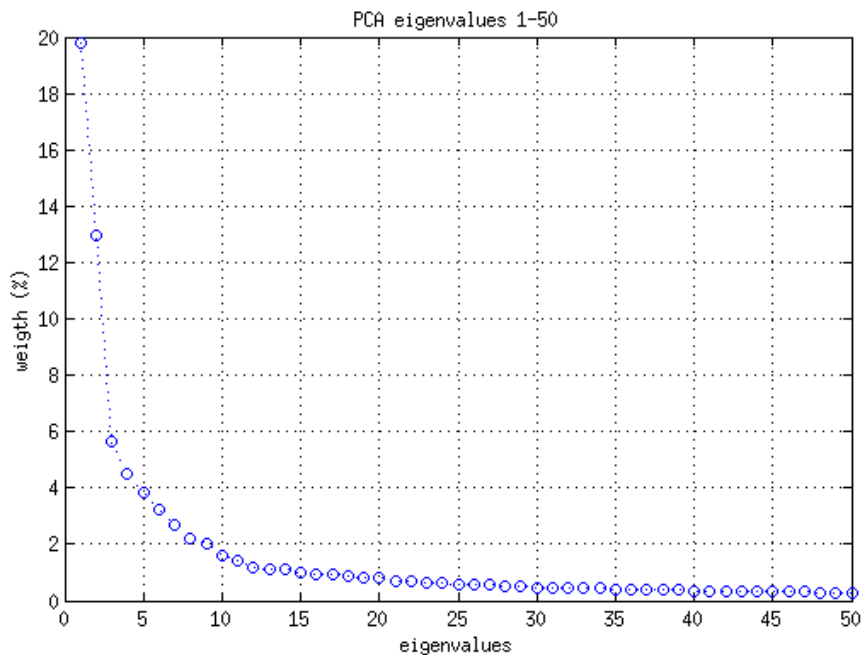


**Picture 40 – features sorted by their ability to describe the data variance, each „o" point represents one feature**

## 5.2 Classifiers

The last part of the data evaluation is to apply the classification method. Classifier is a device with $n$ inputs and $y=1$ output, which consists of two classes $y \in \{1,0\}$. We are using a binary classifier.

In baseline method the Fisher Linear Discriminant and Multithreshold method were applied (3.1.1), in this work classifiers based on Nearest Neighbours and Decision Trees were chosen.

### 5.2.1 Nearest Neighbours

K-nearest-neighbour classification is one of the most simple and fundamental „supervised" classification methods (34). It is commonly based on the Euclidean distance between the training samples and the test samples. The Euclidean distance between two samples *(a,b)* with *f1,f2...fn* features is defined as:

$$d(a,b) = \sqrt{\left(a_{f1} - b_{f1}\right)^2 + \left(a_{f2} - b_{f2}\right)^2 + ... + \left(a_{fn} - b_{fn}\right)^2} \qquad (1.10)$$

The predicted class of the test sample *b* is set equal to the class of its nearest neighbour training sample *a*. The distance of the nearest samples is defined as:

$$d(a_i,b) = \min_n \{d(a_n,b)\} \qquad (1.11)$$

### 5.2.2 Decision trees

Decision tree learning method creates a tree-shaped model based on provided train data for predicting values of the test data. Having the data of two classes, each tree is represented by several binary splits with two possible results „1" or „0". Classification trees are adaptive and robust, but weak and do not generalize well. To enhance their performance, we can perform bagging or boosting (both improving prediction and reducing the variance of unstable procedures). Bagging uses simple averaging and combines results of based classifiers by voting, the final predicted class is the class with most „votes". Boosting uses weighted averages and traines the base classifiers on weighted data. Weights are based on the performance of previous classifier, with increasing the weights for misclassified data (35). Boosting is typically aplied to weak learners that don´t have to be much better than random guess (36). Bagging usually construct deep trees while boosting shallow trees.

Random forest is a method using many decision trees as classifiers and random sample of features (typically $\sqrt{features}$) at each tree split.

### 5.2.3 ROC curve

ROC curve (Receiver Operating Characteristic Curve) describes the performance of a model across the entire range of classification threshold. It is an excellent tool for assessing class separation. It shows the relationship between sensitivity and specificity values - the true positive rate (sensitivity) versus false positive rate (1-specificity).

### *ROC curve in binary classifiers*

Binary classifier (such as kNN) separates data into 2 classes (true/false = 0/1), it produces discrete outputs. For plotting the ROC curve, these outputs have to be scored (else the ROC curve produces only one-point). The score is a value that corresponds to the probability that the data point belongs to the chosen class. Simply, the score value express how certain the algorithm is. A high score signifies that the instance is more likely from the positive data class, low score signifies that it is more likely from the negative class.

When using kNN, I defined scores based on the ratio of classes in the neighbourhood. I counted the score values as a sum of positive data class in all found neighbours.

### 5.2.4 AUC

AUC (Area Under Curve) is the area under the ROC curve. It is a single number, which sets the performance. *AUC=1* indicates perfect performance when all positive examples *xn={1}* are classified as positive ones, any deviation from this ranking decreases the AUC value, *AUC=0.5* indicates the random performance.

Let *n* is the number of positive examples, and *m* is the number of negative examples. Then *x1...xn={1}* is the output of positive examples of the classifier and *y1...ym={0}* is the output of negative examples of the classifier and *fx* is the indicator function. Then AUC is given by (37):

$$AUC = \frac{\sum_{i=1}^{n} \sum_{j=1}^{m} f_{yj<xi}}{nm}$$

(1.12)

In MATLAB ROC and AUC could be easily received using *perfcurve* command.

# 6 IMPLEMENTATION

The data analysis and the whole algorithm was implemented in MATLAB 2011b run under the Unix system. It is divided into several scripts and functions, all available in provided archive.

## 6.1 Archive description

The principal algorithm scripts, functions and data are attached. This data is provided on CD and consists of these files:

*Nodule* - root folder contaning all the data

    *Nodule/data* – subfolder containing all the input data

    *Nodule/results* – subfolder containing examples of output data

    *Nodule/code* – subfolder containing used scripts and functions

*Readme_code.txt* - operating instructions to all scripts

*Readme_data.txt* – description of provided input and output data

*Dipl.pdf* – diploma thesis text

## 6.2 3D visualization

First it was possible to view the data only in 2D space, in transversal (axial) plane. In praxis, it is routine to examine CT data in 3 orthogonal planes: axial, coronal and sagittal (Picture 41). For providing a cutout view in all 3 planes a simple script *zobraz3.m* was made. The resolution in *x* and *y* axes (the axial plane) is higher than the resolution in *z* axis (coronal and sagittal planes), this is influenced by the slice thickness and the voxel size.
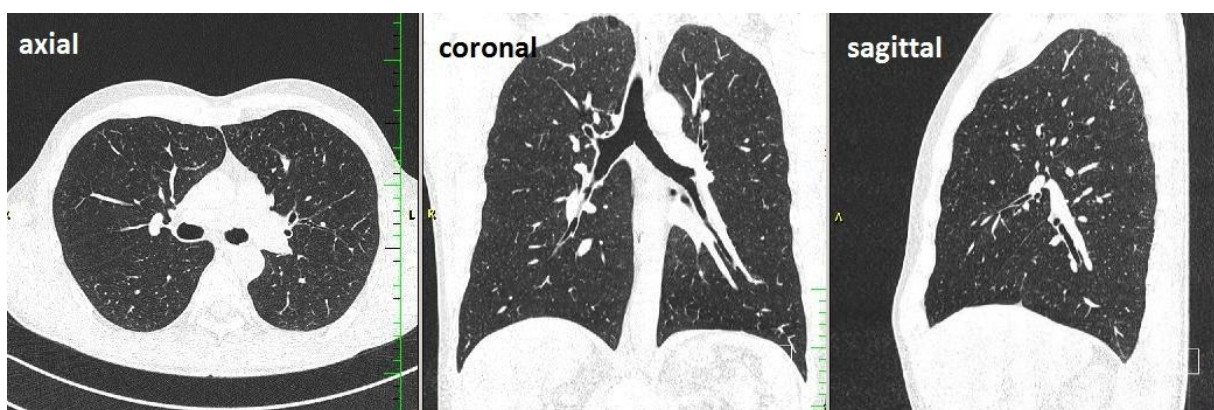


**Picture 41 – 3 fundamental planes of sectioning in medicine imaging**

Using all 3 planes we can easily get a grasp of structure appearance (Picture 42, Picture 43). Both pictures are divided into 4 parts, we can see the whole axial slice (with scan number and xyz coordinates above) and the three cutout views (in axial, coronal and sagittal plane). Red circle is always placed at the detection point.
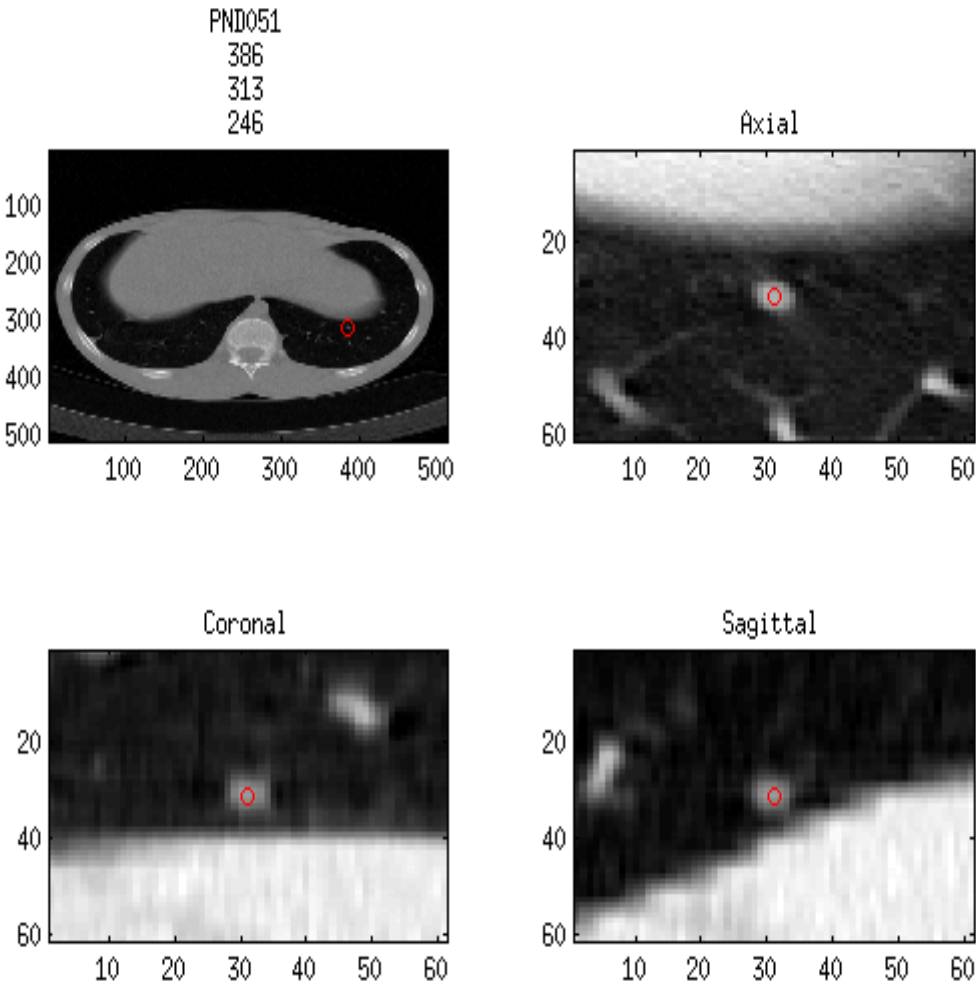


**Picture 42 - Example of a nodule structure visualized by zobraz3.m**
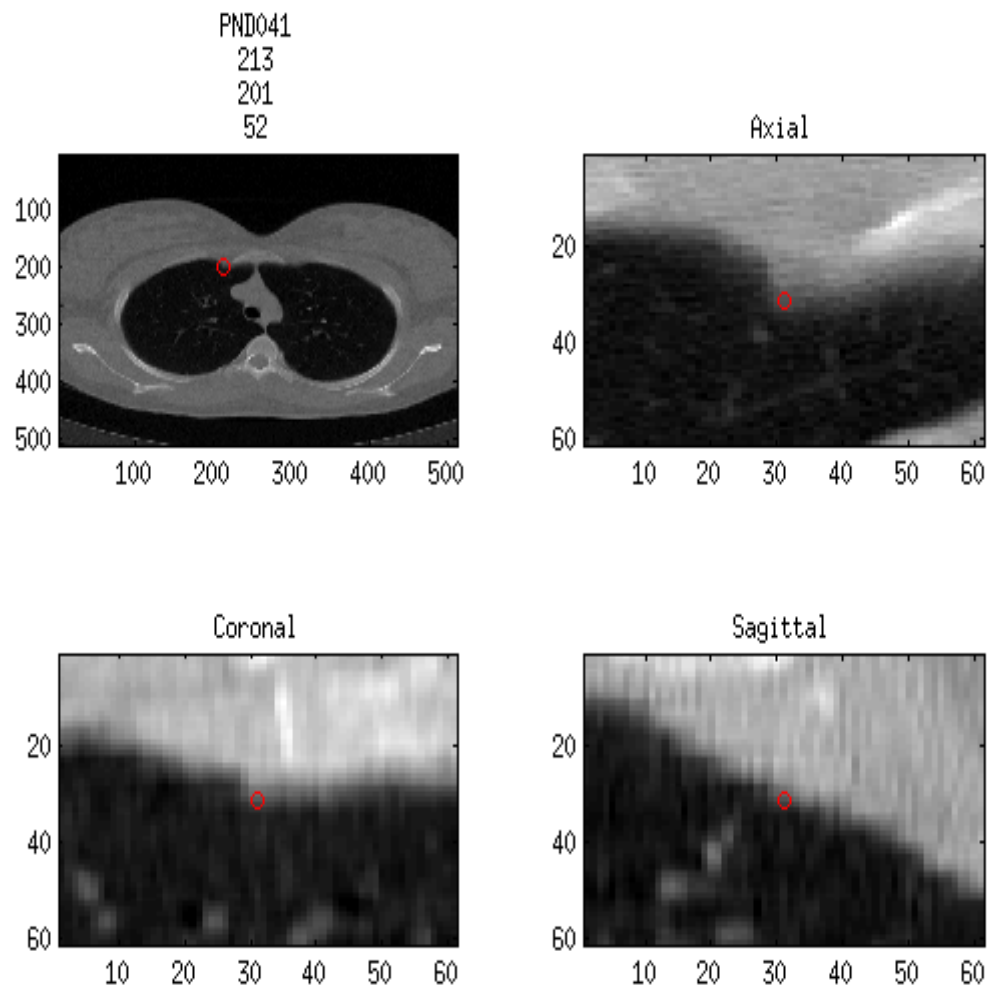
**Picture 43 - example of a non-nodule structure (margin of a sternocostal junction) visualized by zobraz3.m**

# 7 RESULTS / EXPERIMENTS

## 7.1 Data preparation

Data containing all the detected nodule candidates are available on cmp.server as: *scratch/dolejm1/nodclassdata.* Script *zparudomatice.m* converts this data from struct to one well-arranged matrix used for further computation. Created input data matrix is saved in enclosed CD archive as: *Nodule/data/detdata.mat*. There are two matrices, the first is containing normalized data and the second original data without normalization (see *readme_data.txt* for more details). Finally I used the normalized data only for visual feature analysis (4.4.1) to make the feature values clear and easy comparable. The original data was used for the rest of my work, not to misrepresent the computed values.

Each column of input „*detdata*" matrix represents one detected point, rows *row(1)* represent the number of set *row(1)*, scan *row(2)*, 3 coordinates *row(3,4,5)*, true class label *row(6)* and 482 features *row(7-489)*. The whole data consists of 2 sets and the overall number of detected nodule candidates is 410994.

For further experiments only data for the first set counting 98 patients was used *row(1)==1*.

### 7.1.1 Data normalization

Normalization is a process rescaling the numeric variables in the range $y \in \langle 0,1 \rangle$. *Xmax* and *Xmin* are the maximal and minimal values for variable *x*.

$$x_{normalized} \quad = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1.13}$$

## 7.2 Ground truth correction

First of all the GT data was corrected due to shortcomings discussed in (4.1.6). Script for excluding non-nodule structures and including new nodules (nodules missed by the supervisor) was constructed. All the included and excluded structures were visually checked using *zobraz3.m*.

GT data will never be the same and perfectly correct, when more than one radiologist is examining the scans – they can interpret the same observations differently . Some experts can target the sensitivity and mark better all of the suspicious structures and the others can target on specificity and notice only the real pathologies. This different judgement is caused by different experiences, praxis length and other parameters. This is one of the reason, why the results of supervised learning differ (supervisors are the inconsiderable reason) and why it is necessary to provide the chance to change the GT data again.

For preserving the original GT information, the existing GT data was not totally changed. Functions for changing GT data (*readinouts.m, upravGT.m, upravkres.m, velikost_pixelu.m*) were inserted into the existing script *prepareREGanalysisNaseData.m*. For further individual changes it is possible to use new GT, the original GT data or to create your own new re-corrected data.

### 7.2.1 Original and corrected GT comparison

According to the original ground truth, nodules were present in 18 of the patients, each counting 1 - 72 nodules. After the GT data correction 34 new nodules were inserted into the GT data and 37 nodules were removed. As a whole there are now 232 nodules in 18 (different) patients, each counting 1 - 78 nodules.

For comparing the results before and after GT correction, the classificaton process described in 7.5 was used. PCA dimensionality reduction method followed by kNN classification was applied. This test was performed on the whole dataset containing 98 scans. Table 8 shows the increase of resulted sensitivity values and number of false positive detections per scan.

| Method (PCA followed by kNN, k=100) | Sensitivity (%) | FPs/scan |
|---|---|---|
| Results before GT correction | 87.8 | 148 |
| Results after GT correction | 91.4 | 151 |

**Table 8 – comparison of original and new GT data classification process**

## 7.3 Data validation

For ensuring that the process operates on correct and clean data, we have to use some model validation technique. This divides data to the training set (data used to train the classifier) and the testing set (unknown data used to estimate effect and error rate of the trained classifier).

### 7.3.1 Cross-validation

Cross-validation method divides data randomly into $k$ equal size subsamples. Due to random division, results of our classification process using k-fold-cross-validation method is influenced by multiple detections of the same structure. The same structure can be found in training and testing data at the same time and leads to false low error rate (38). The error rate $E$ is estimated as the average of separated error rates $E_i$.

$$E = \frac{1}{k}\sum_{i=1}^{k} E_i \qquad (1.14)$$

 To avoid this it is necessary to split the data into disjoint subsets.

### 7.3.2 Per-patient division

I applied per patient division using one patient scan for testing and the rest 97 scans for training. The training data were then partly reduced by removing the randomly chosen 1/10 of the data (leave-part-out method).

$$data = set_1\{s_1, s_2 \ldots s_{98}\}$$
$$test = s_x; \quad s_x \in set_1$$
$$train = s_y; \quad s_y \in set_1 - s_x - \frac{1}{10} * set_1$$

(1.15)

## 7.4 Evaluation criteria

In the input data (*detdata.mat*), each detection of a real nodule structure is marked as „1", that means it is a nodule. Each detection of a non-nodule structure is marked as „0". All detections, even the multiple detections of the same structure are used (this raises the number of real nodule structures in training group). Results of the classification process are compared with the real class of the testing data.

Nodules found by the algorithm are considered to be the true positive detections (TPs), missed nodules are considered to be the false negative detections (FNs). False positives (FPs) are all the non-nodule structures detected as nodules.

To measure the performance of the classifier, sensitivity and specificity rates are counted based on the number of true positives, false positives and false negatives.

$$Sensitivity = \frac{TPs}{(TPs + FPs)}$$

(1.16)

$$Specificity = \frac{TNs}{(FPs + TNs)}$$

(1.17)

To make my results comparable with the results from baseline algorithm, I used a modified original script *prepareREGanalysisNaseData.m.* It computes the overall sensitivity rate and number of FPs for the whole used dataset and gives results in the same form as the baseline algorithm. Nodule is considered to be a TP if exists any detection point $DET_i$ from the set of detections $DET_n$ that matches any nodule point $GT_i$ in the nodule area $GT_n$. If it does not exist the nodule is marked as FN. All out-lying detections are considered as FPs. $GT_n$ area is delimitated by the elipse with x,y,z half axes lenghts enlarged of 10% with the centre in the *x=0,y=0,z=0*.

$$TP = \{DET_i \in GT_n\}$$
$$FP = \{DET_i \notin GT_n\}$$
$$FN = \{GT_i \notin DET_n\}$$

(1.18)

## 7.5  Experiment 1 – dimensionality reduction methods

As a first experiment I compared dimensionality reduction methods. Training and testing data were prepared as mentioned in 7.3. For classification, the kNN (k=100) classifier was performed. Sensitivity rate, number of FPs and AUC were counted for each of the 98 scans, the final results are reported as a mean. From feature selection methods I chose two. Selecting features based on the classification error threshold and feature selection method based on Fisher´s discriminant ratio (both described in 5.1.1). The third dimensionality reduction method is PCA.

### 7.5.1  Experiment 1 - results

For threshold based feature selection method only features with classification error<0.3 were used (16 features used).

For Fisher´s discriminant method the significance power threshold for choosing features was set at median value of all significances. It means that only features of higher significancies than a median of significancy for all features are used (146 features used).

$$sig_f > median(sig_F); \ \ f \in F \hspace{2cm} (1.19)$$

PCA is a part of Statistics Toolbox and could be computed using various commands in MATLAB: pca, pcacov, princomp. „*Princomp*" was chosen and features that each interpret more than 1% of the data variability were included (15 features included using this criterium).

| Method | Sensitivity (%) | FPs/scan | AUC |
|---|---|---|---|
| Threshold based feature selection method[4] „fselection" | 88.8 | 367 | 0.87 |
| Fisher´s discriminant test for selecting features[5] „signiffeat" | 83.2 | 361 | 0.77 |
| nPCA[6] | 91.4 | 151 | 0.94 |

Table 9 - results for comparing dimensionality reduction methods (experiment 1)

From all three methods PCA has the highest sensitivity and the lowest number of FPs per scan. Method based on Fisher´s discriminant was the worst, one possible reason can be the huge number of used nodule features (146 used features, while the other two methods have comparable number of used features). This method was rejected and it is not used in further experiments.

---

[4] Results saved as: vysledkyfselectionkNN100
[5] Results saved as: vysledkysigniffeatkNN100
[6] Results saved as: vysledkynPCA15kNN100

## 7.6   Experiment 2 – classification methods

For classification step classifiers based on Nearest Neighbours search and DecisionTrees were performed. I applied quick and available MATLAB *knnclassify*, *TreeBagger* and *fitensemble* commands. For this experiment only scans containing nodules were included ($n_{scans}$=18) into the testing set, the training data remains the same.

### 7.6.1   Finding the optimal number of nearest neighbours (k)

To find the ideal number of nearest neighbours I computed the equal error rate for first 1000 nearest neighbours for each test sample. The mean lowest error rate falls into interval $k \in (100,200)$ with the lowest *k=130* value. Therefore kNN classifier with *k=100* nearest neighbours were applied in further tests. Equal error rate was counted as:

$$err = (1 - sensitivity) + (1 - specificity) \qquad (1.20)$$

Picture 44 shows that the error rate first decreases very fast, then it starts to increase slowly. When the number of nearest neighbours is raising, values of sensitivity and false positive rate are increasing and the value of specificity is decreasing (Picture 45).
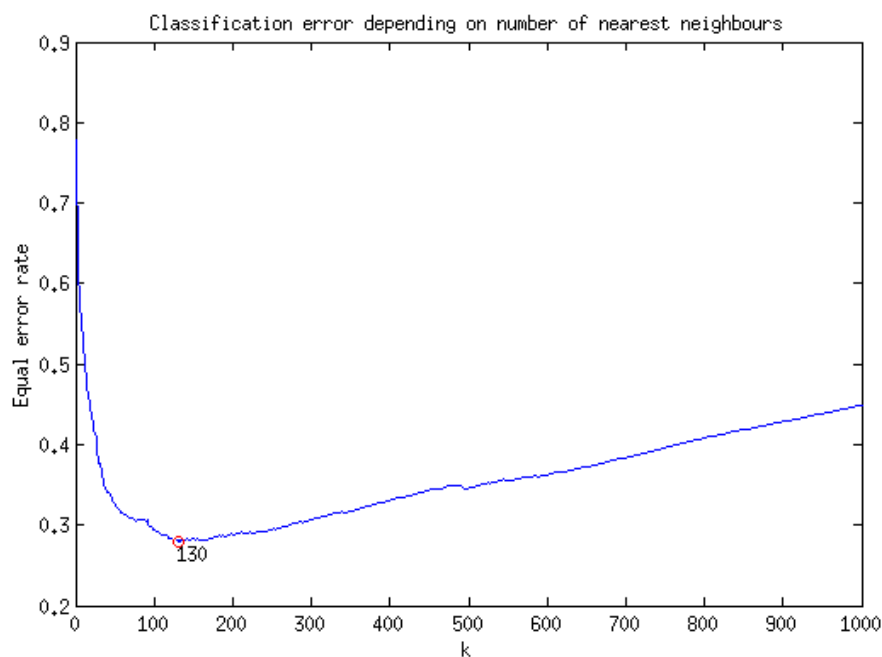


**Picture 44 – error rate depending on number of used k-nearest neighbours, performed on the dataset of 18 scans**
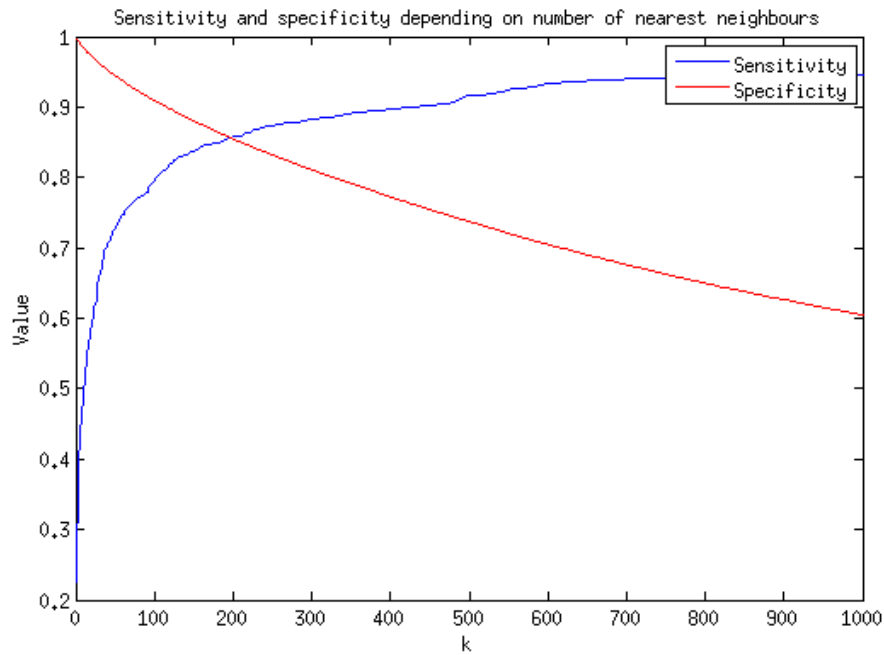
**Picture 45 – sensitivity and specificity rate depending on number of used k-nearest neighbours**

## 7.6.2 Bagging method

For creating bagged decision trees, command *TreeBagger* was used. For finding the optimal number of grown trees I applied the algorithm with 100 trees for the dataset. Then I displayed the out-of-bag-error (error of unused trees) for each tree. The out-of-bag error is decreasing with increasing number of used decision trees. The optimal number of used trees was set as 30, because there is no steep decrease of error rate when more than 30 trees are used.
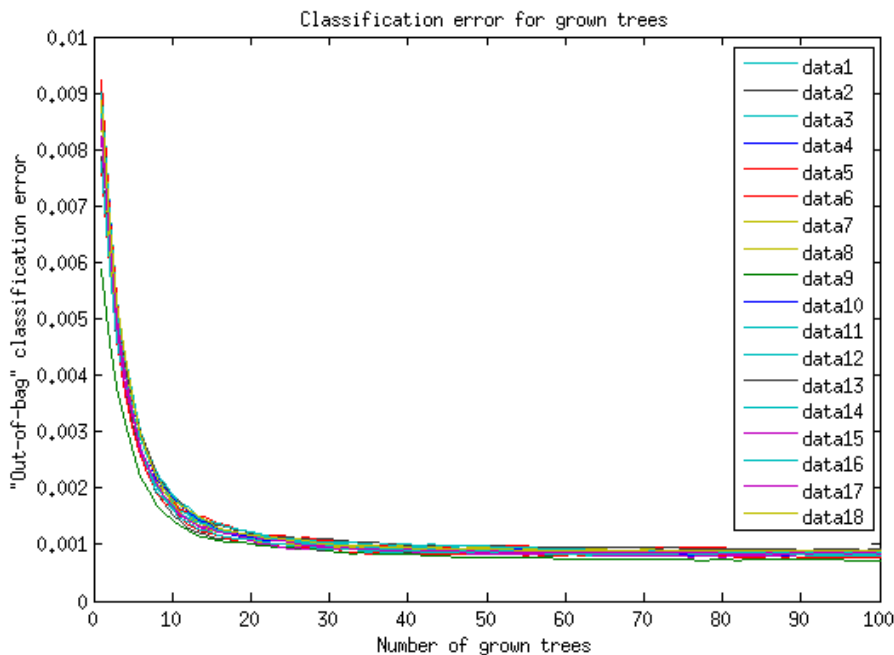


**Figure 7 – out-of-bag error for each of 18 scans, TreeBagger classifier performed**

When the input data is not balanced or we need to favour one of the classes, misclassification cost have to be set. Cost matrix is a square matrix *[0 a;b 0]* which sets the cost of placing an observation into bad class. The classifier prefers „costly" class which is oversampled, while the „less costly" class is undersampled by the computer (39).

In our dataset cost matrix did not work well and even the high cost of „nodule" class did not change the results. To increase the cost of this class I added *duplic.m* script for re-sampling the input data manually. User can manually modify the size of training data seting the parameters for oversampling the „nodule" trains and undersampling the „nonnodule" trains.

### 7.6.3 Boosting method

In MATLAB it is possible to perform many methods using a number of weak learners into one high-quality predictor (40). As a third method I applied *fitensemble* command with *decision tree* used as a weak learner.

For binary classification AdaBoostM1, GentleBoost and LogitBoost method is required (40). For choosing the best one, resubstitution error and ROC curves were drawn up for each (Picture 46, Picture 47). Resubstitution is testing the data on the provided set already used for training, the error rate received from this testing is called the resubstitution error. It says how good the results are when applied on training data. Based on this parameter Logit Boost was chosen for the further testing.
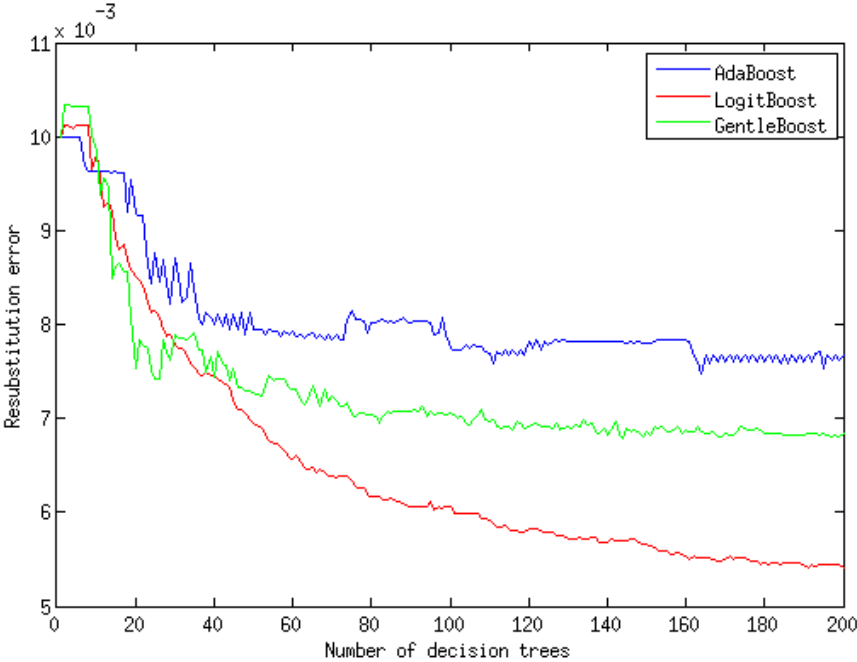


**Picture 46 – resubstitution error values for 3 tested boosting algorithms**

62
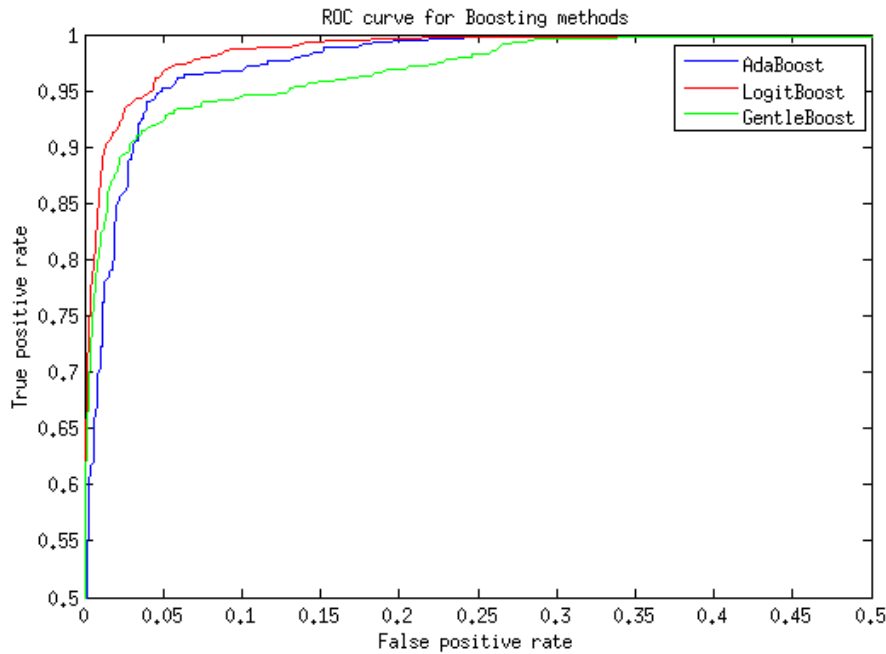
**Picture 47 – ROC curves for 3 tested boosting algorithms**

### 7.6.4 Experiment 2 - results

For classification I chose classifier based on nearest neighbours and classifiers based on decision trees (enhanced by bagging and boosting method).

Results of the classifier performances are shown in following tables. Method (performed classifier), dim. Reduction (performed dimensionality reduction method), Sensitivity, FPs/scan and AUC results are presented in the columns. Column „Number" presents the number of used nearest neighbours or decision trees. Methods with „*duplic*" in the title are applied on the modified training data (see 7.6.2) received undersampling and oversampling the classes.

Tables are showing the first results and results recalculated into the same sensitivity rate. Charts are displayed for every method to compare the results.

Some of the resulted data used for creating the ROC curves are available in data archive in subfolder *Nodule/results*.

## kNN classifier

| Method | Number of kNN | Dimensionality reduction | Sensitivity [%] | Specificity [%] | FPs/scan | AUC |
|--------|---------------|--------------------------|-----------------|-----------------|----------|-----|
| kNN | 100 | Fselection | 77.0 | 91.3 | 426 | 0.87 |
| kNN | 1000 | Fselection | 93.1 | 61.4 | 1909 | 0.93 |
| kNN | 100 | nPCA | 88.8 | 96.0 | 186 | 0.94 |
| kNN[7] | 1000 | nPCA | 96.0 | 82.8 | 838 | 0.97 |

**Table 10 – classification results (test data=18 scans), kNN method used**

| Method | Number of kNN | Dimensionality reduction | Specificity rates for chosen sensitivities | | |
|--------|---------------|--------------------------|---------|---------|---------|
| | | | Sens=77% | Sens=90% | Sens=95% |
| kNN | 100 | Fselection | 91.3 | - | - |
| kNN | 1000 | Fselection | 94.3 | 73.9 | - |
| kNN | 100 | nPCA | 99.3 | - | - |
| kNN | 1000 | nPCA | 99.2 | 95.1 | 86.9 |

The best results are reached using kNN classifier with 100 or 1000 nearest neighbours. PCA is better dimensionality reduction method, than feature selection.
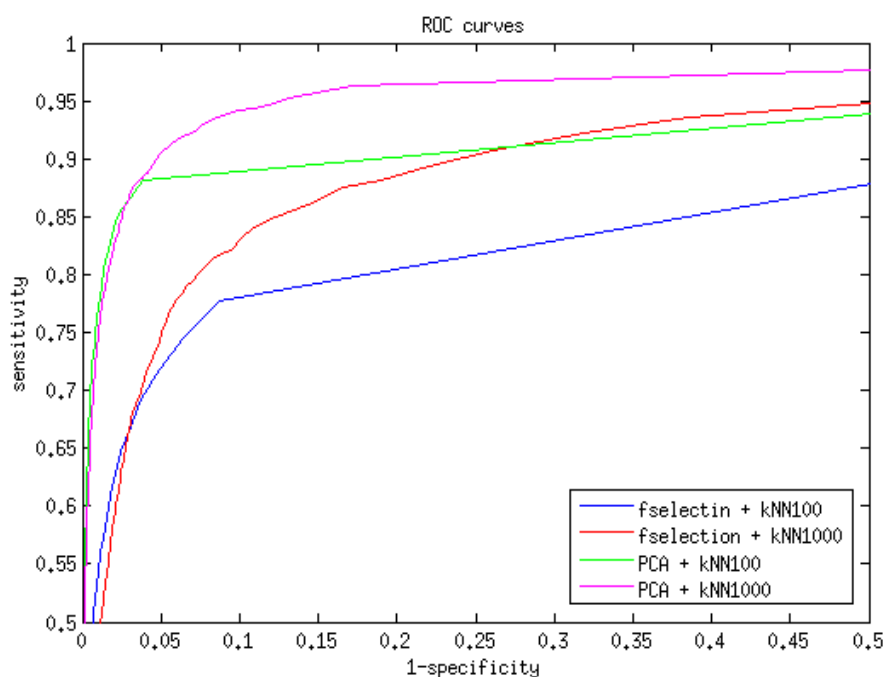


**Picture 48 – performance curves for kNN classifier**

---

[7] Results saved as: vysledkynPCAkNN1000

## Bagged trees

| Method | Number of trees | Dimensionality reduction | Sensitivity [%] | FPs/scan | AUC |
|---|---|---|---|---|---|
| TreeBagger | 100 | Fselection | 55.5 | 2 | 0.94 |
| TreeBagger-duplic | 30 | Fselection | 79.4 | 107 | 0.94 |
| TreeBagger-duplic | 100 | Fselection | 63.0 | 127 | 0.93 |
| TreeBagger-duplic | 30 | nPCA | 74.1 | 30 | 0,96 |
| TreeBagger[8] | 100 | nPCA | 46.8 | 2 | 0,96 |

**Table 11 - classification results (test data=18 scans), bagged trees method used**

| Method | Number of trees | Dimensionality reduction | Specificity rates for chosen sensitivities | | |
|---|---|---|---|---|---|
| | | | Sens=77% | Sens=90% | Sens=95% |
| TreeBagger | 100 | Fselection | 98.7 | 90.5 | - |
| TreeBagger-duplic | 30 | Fselection | 96.2 | 85.1 | 59.0 |
| TreeBagger-duplic | 100 | Fselection | 94.4 | 78.1 | 57.7 |
| TreeBagger-duplic | 30 | nPCA | 98.5 | 91.1 | 87.7 |
| TreeBagger | 100 | nPCA | 98.7 | 90.0 | 78.0 |

The best results are reached using both TreeBagger classification methods with PCA dimensionality reduction.
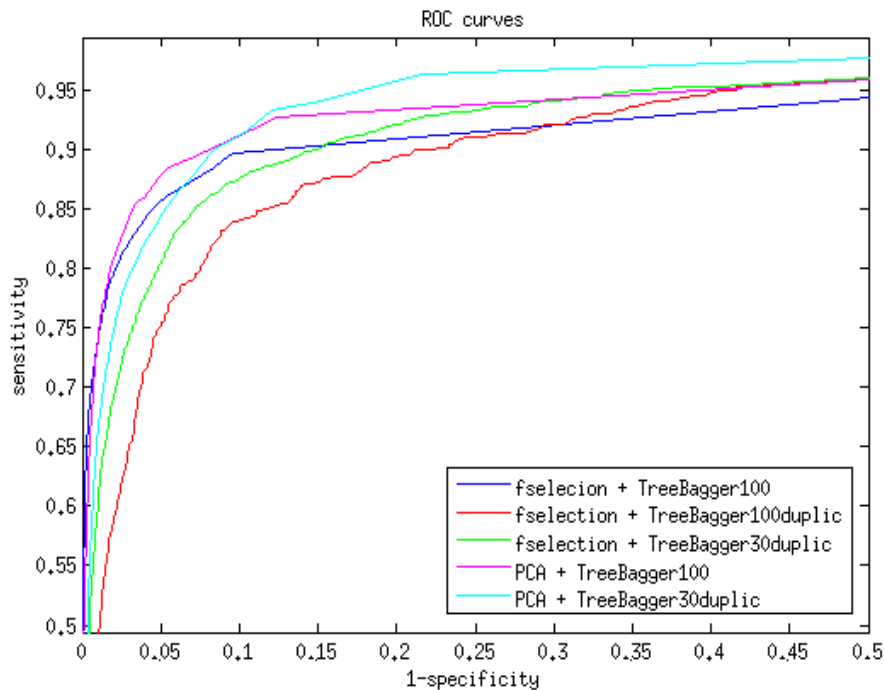


**Picture 49 – performance curves for TreeBagger classifier**

---

[8] Results saved as: vysledkynPCATreeBagger100

### *Boosting method*

| Method | Number of trees | Dim.reduction | Sensitivity [%] | FPs/scan | AUC |
|---|---|---|---|---|---|
| LogitBoost[9] | 100 | Fselection | 58.2 | 6 | 0.96 |
| LogitBoost[10] | 100 | nPCA | 54.3 | 4 | 0.96 |
| LogitBoost-duplic | 100 | nPCA | 94.4 | 1169 | 0.96 |
| LogitBoost-duplic | 200 | nPCA | 94.4 | 946 | 0.96 |

**Table 12 - classification results (test data=18 scans), LogitBoost method used with decision trees as a weak lerners**

| Method | Number of trees | Dim.reduction | Specificity rates for chosen sensitivities | | |
|---|---|---|---|---|---|
| | | | Sens=77% | Sens=90% | Sens=95% |
| LogitBoost | 100 | Fselection | 97.0 | 88.2 | 73.6 |
| LogitBoost | 100 | nPCA | 95.2 | 87.5 | 78.5 |
| LogitBoost-duplic | 100 | nPCA | 94.7 | 87.5 | 80.0 |
| LogitBoost-duplic | 200 | nPCA | 94.6 | 86.6 | 78.3 |

There is no significant difference in performance rates among all presented LogitBoost methods. The only difference is in ROC curve shape. Methods based on PCA dimensionality reduction have better results when used in higher sensitivities.
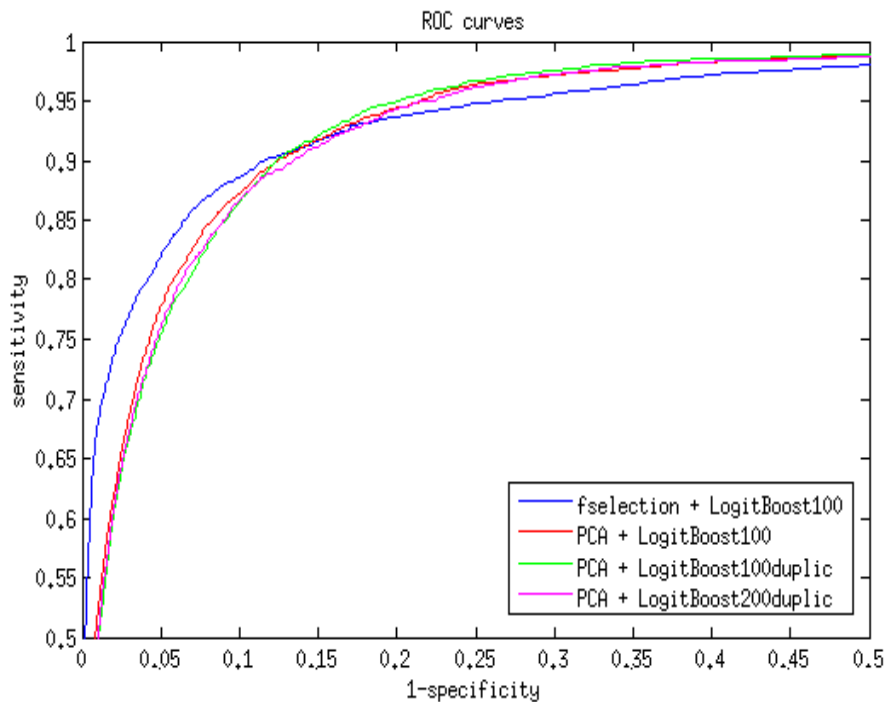


**Picture 50 - performance curves for LogitBoost classifier**

---

[9] Results saved as vysledkyfselectionfitensemble100
[10] Results saved as: vysledkynPCA15fitensemble100

## Classifiers comparison

The best results are reached when bagging or boosting classifier with PCA as a dimensionality reduction method is used. For comparison of the performed methods see Picture 51.
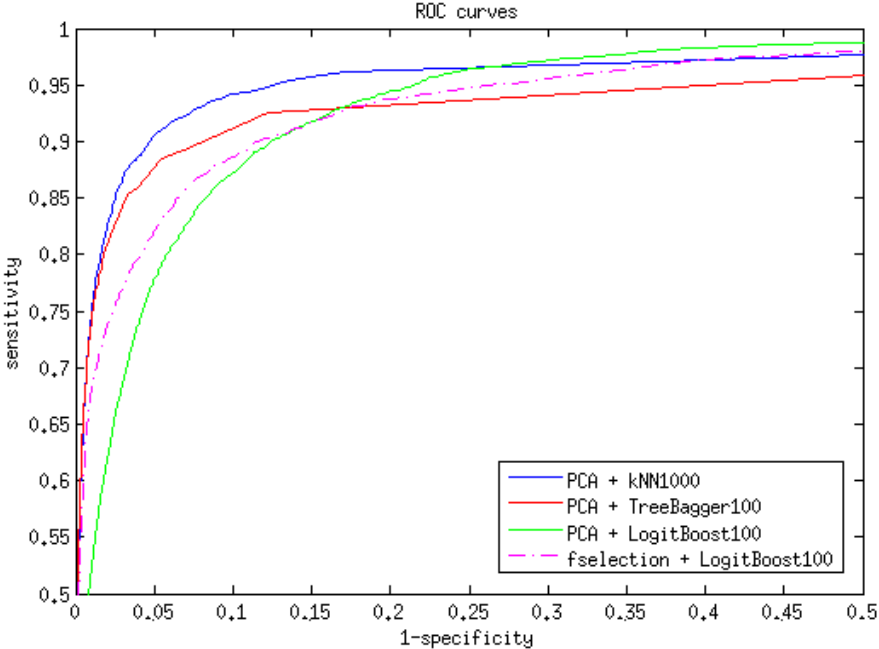


**Picture 51 – comparison of the performed methods**

## 7.7 Comparison with other works

There is an evidence, that performances of CAD systems signifficantly differ when tested on different datasets. Based on Table 13 some results of the authors seem to be outstanding, even unreal. For example Hirose (18) reports 71% sensitivity with less than 0.95 FPs per scan and Choi (22) only 2.27 FPs per scan at more than 95% sensitivity rate. For better evaluation, the resulted number of false positives should be reported in slices, because the number of slides per one scan has a significant influence on the resulting number of FPs. Zhao (14) reports maximum 80 slices per scan, our average number of slices is more than 4 times higher (slices=350), numbers of our FPs/slide is in that case much smaller than 0.5. The results also depend on the input data set and the nodule size. Some of our nodules (>1mm) are much smaller than the nodule size reported by the others, even nodule>5mm for Gori et al. (20).

| Author | Sensitivity [%] | FPs/scan |
|---|---|---|
| Bisheng Zhao et al. (14) | 84.2 | 5 |
| Golosio et al. (15) | 79 | 4 |
| Kyongtae et al. (16) | 95.1 | 6.9 |
| Tan et al. (17) | 87.5 | 4 |
| Hirose et al. (18) | 71.4 | 0.95 |
| El-Baz et al. (19) | 82.3 | 12 |
| Gori et al. (20) | 85.2 | 6 |
| Yuan et al. (21) | 72.6 | 3.19 |
| Choi (22) | 95.28 | 2.27 |
| Cascio et al. (23) | 88 | 2.5 |
| Dolejší et al. (24) | 94.03 | 5.46 FPs/slice |
| Dolejší et al. (25) | 74.3 | 2.6 FPs/slice |
|  | 89.6 | 9 FPs/slice |

**Table 13 – performance rates presented by other authors**

Results of the baseline algorithm my work was based on were 95.9% sensitivity with 12 FPs/slice for the detector only, 74.3% sensitivity with 2.6 FPs/slice for FLD classifier and 89.6% sensitivity with 9 FPs/slice for the multi-threshold classifier.Knowing the average number of slices per one scan FPs/slice can be recounted to FPs/scan.

For comparison with the existing algorithm and other works see Picture 52.
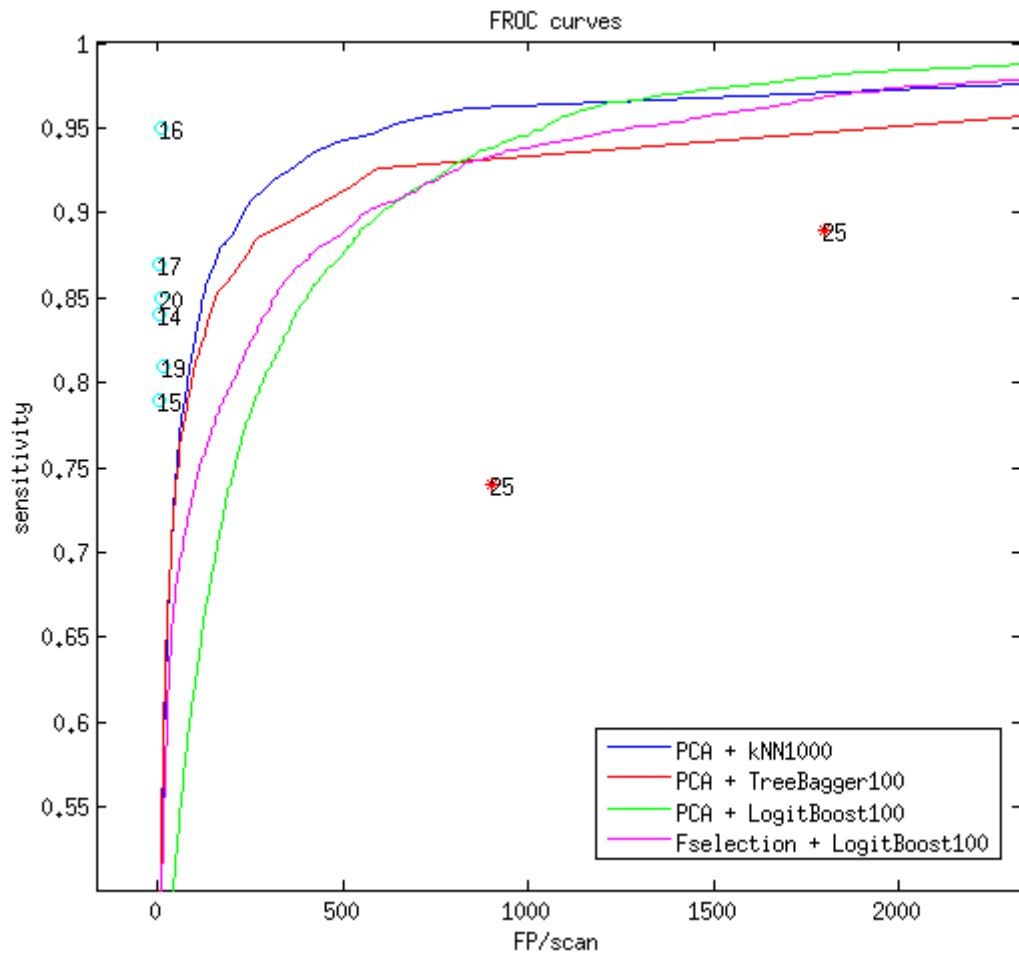
**Picture 52 – out algorithm compared to the baseline algorithm results (25) and the results reported in other works (14,15,16,17,19,20 – the numbers correspond to the reference in bibliography)**

# 8 CONCLUSION

Analysing the results of baseline algorithm carefully I found shortcomings in the process of nodule candidate detection, in feature computation and also in the ground truth information. All the findings were used for further work.

First of all the shortcomings in GT data was corrected manually. Data correction improved the sensitivity rate, the specificity rate remains the same. The resulted difference between original and corrected data classification are 87,8% sensitivity rate with 148 FPs per scan for the original GT and 91,4% sensitivity rate with 151 FPs per scan for the corrected data.

Then the input data dimension was reduced, redundant features were removed and only effective features were chosen using 3 independent dimensionality reduction methods (Feature Selection method, Fisher´s Discriminant analysis and Principal Component Analysis). The best results were saved and combined with 3 suitable classification methods (based on nearest neighbour search, bagged decision trees method and boosting method with decision trees used as a weak learners). The classification parameters were set to receive as good sensitivity and as low number of false positive structures as possible, methods were compared on the same sensitivity levels.

The highest sensitivity rate together with low FPs were observed in kNN classifier (96% sensitivity with 838 FPs/scan, which corresponds to 2.4 FPs/slice, AUC=0.97). Both other methods (Treebagger and LogitBoost) reached AUC=0.96 and the results were more ballanced compared to kNN. The best results were achieved with dimensionality reduction process based on PCA.

When comparing our results to other studies, we reached similar performance values, but not better, our number of FPs is still very high. However we have to consider that some of the authors have less slices per one scan, bigger nodules and different datasets, this all can lead to better results. When compared to the baseline algorithm the resulted performance values improved a lot.

# 9 FUTURE WORK

Results of the algorithm are still not perfect, in future it is possible to improve it by:

*Improving the detection process* to reduce definite non-nodule detections

*Improving the segmentation process* to guarantee the precise contours of the structure and following feature computation

*Reducing the number of FPs.* Performed pre-classification method to reduce the number of FPs (for example double-threshold cut)

*Categorization the detected data* into several groups (not only „nodule" and „non-nodule") to provide information about possible etiology of detected structure (like „sure nodules", „possible nodules", „sure non-nodules", „other non-nodule pathologies" etc)

*Individualization*. User can set the required sensitivity value based on the individual CT findings. For example individual performance rates depending on the final number of found nodules.

# BIBLIOGRAPHY

1. **WHO.** Cancer. *World health Organisation.* [Online] WHO. [Cited: 2 1, 2013.] http://www.who.int/mediacentre/factsheets/fs297/en/.

2. **Prokop, Mathias.** *Spiral and Multislice Computed Tomography of the Body.* 2001. ISBN:0865778701.

3. **MacMahon, Heber.** Guidelines for Management of Small Pulmonary Nodules Detected on CT Scans: A Statement from the Fleischner Society. *Radiology.* 2005, Sv. Volume 237, Issue 2, stránky 295-400.

4. **Miller, Janet Cochrane.** Evaluating Pulmonary Nodules. *Radiology Rounds.* 4, 2006, Sv. 8.

5. **Shah, Vaibhav Ramesh.** *Radiodiagnosis, Nuclear Medicine, Radiotherapy and Radiation Oncology .* ISBN 9789350257128.

6. **Gurney, Jud W.** *Specialty Imaging: HRCT of the Lung: Anatomic Basis, Imaging Features, Differential Diagnosis.* místo neznámé : Amirsys, 2009. ISBN 1931884161.

7. **Balekian, Alex A.** The Lung Nodule That Refused To Grow. *Agency for Healthcare Research and Quality.* [Online] 2012. http://webmm.ahrq.gov/case.aspx?caseID=287.

8. **Weerakkody, Dr Yuranga.** Angioinvesive aspergillosis. *Radiopaedia.* [Online] http://radiopaedia.org/articles/angioinvasive-aspergillosis.

9. **Gill, Ritu R.** Cavities in the Lung in Oncology Patients: Imaging Overview and Differential Diagnoses. *Appl Radiol.* 2010, Sv. 39(6), stránky 10-21.

10. Cancer. *2nd Radiology Opinion.* [Online] [Citace: 1. 2 2014.] http://www.2ndradiologyopinion.com/symptoms/cancer/.

11. **Leung, Ann.** Solitary pulmonary nodule: benign versus malignant. *Radiology Assistant.* [Online] http://www.radiologyassistant.nl/en/p460f9fcd50637/solitary-pulmonary-nodule-benign-versus-malignant.html.

12. **Brandman, Scott a P.Ko, Jane.** Pulmonary Nodule Detection, Characterization, and Management With Multidetector Computed Tomography. *J Thorac Imaging.* 26, 2011, stránky 90-105.

13. **Biegelman-Aubry, C., Hill, C. a Boulanger, X.** Evaluation of a computer aided detection system for lung nodules with ground glass opacity component on multidetector-row CT. *Journals of Radiology.* 90, 2009, stránky 1843-49.

14. **Zhao, Bisheng.** Automatic detection of small lung nodules on CT utilizing a local density maximum algorithm. *Journal f applied clinical medical physics.* 4, 2003, Sv. 3.

15. **Golosio, B.** A novel multithreshold method for nodule detection in lung CT. *Medical Physics.* 26, 2009, stránky 3607-18.

16. **Kyongtae, Bae T.** Pulmonary Nodules: Automated Detection on CT Images with Mirphologiec Matching Algorithm - Preliminary Results. *Radiology.* 236, 2005, pp. 286-294.

17. **Tan, Maxine.** A novel computer-aided lung nodule detection system for CT images. *Medical Physics.* 38, 2011.

18. **Hirose, T., Nitta, N. a Shiraishi, J.** Evaluation of computer-aided diagnosis (CAD) software for the detection of lung nodules on multidetector row computed tomography (MDCT): JAFROC study for the improvement in radiologists' diagnostic accuracy. *Academic Radiology.* 15(12), 2008, stránky 1505-12.

19. **El-Baz, Ayman.** Automatic Detection of 2D and 3D Lung Nodules in Chest Spiral CT Scans. *International Journal of Biomedical Imaging.* vol. 2013, 2013.

20. **Gori, I., a další.** An automated system for lung nodule detection in low-dose computed tomography. *Proceedings of SPIE.* 6514, 2007.

21. **Yuan, Ren, Vos, Patrick M. a Cooperberg, Peter L.** Computer-Aided Detection in Screening CT for Pulmonary Nodules. *American Journals of Radiology.* 186, 2007, stránky 1280-87.

22. **Choi, Wook-Jin a Choi, Tae-Sun.** Automated Pulmonary Nodule Detection System in Computed Tomography Images: A Hierarchical Block Classification Approach. *Entropy.* 15, 2013, stránky 507-523.

23. **Cascio, D., a další.** Automatic detection of lung nodules in CT datasets based on stable 3D mass–spring models. *Computers in Biology and Medicine.* 42, 2012, stránky 1098-1109.

24. **Dolejší, Martin, a další.** The Lung TIME - Annotated Lung Nodule Dataset and Nodule Detection Framework. *In Proceedings of SPIE.* February 2009.

25. **Dolejší, Martin.** Detection of Pulmonary Nodules from CT Scans. *Master's thesis.* Prague : FEE Czech Technical University, 2007.

26. **Wiemker, R.** Computer-aided detection (CAD) and volumetry of pulmanary nodules od high-resolution CT daa. *MedicaMundi.* 47/3, 2003, stránky 37-44.

27. **Nietert, Paul J., a další.** Influence of nodule Detection Software on Radiologists Confidence in Identifying Pulmonary Nodules With Computed Tomograpgy. *J Thoracic Imaging.* 26(1), 2011, stránky 48-53.

28. **Dolejší, Martin a Kybic, Jan.** *Comparison of Nodule Candidatte Detecting Algorithms.* 2012.

29. **Alelyani, Salem, Tang, Jiliang a Liu, Huan.** Feature Selection for Clustering: A Review. [Online] [Citace: 24. 4 2014.] http://www.public.asu.edu/~salelyan/MyPub/FS4clustering_chapter.pdf.

30. **Weiss, Sholom a Indurkhya, Nitin.** *Predictive Data Mining; A Practical Guide.* místo neznámé : Morgan Kaufmann, 1997. 978-1-55860-403-2.

31. **Dwinnel, Will.** Feature slection, Phase 1: Eliminate the Chaff. *Data mining in MATLAB.* [Online] [Citace: 25. 3 2014.] http://matlabdatamining.blogspot.cz/2006/12/feature-selection-phase-1-eliminate.html.

32. **Šochman, Jan.** Principal Component Analysis (PCA). *Cvičení z předmětu Rozpozávání (X33RPZ).* [Online] Center for Machine Perception. [Citace: 24. 4 2014.] http://cmp.felk.cvut.cz/cmp/courses/recognition/Labs/pca/index.html.

33. Principal Component Analysis. *Lecture 15.* [Online] http://www.doc.ic.ac.uk/~dfg/ProbabilisticInference/IDAPILecture15.pdf.

34. **Peterson, Leif E.** K-nearest Neighbor. *Scholarpedia.* [Online] 2009. http://www.scholarpedia.org/article/K-nearest_neighbor.

35. **Srinet, Amit a Snyder, Dave.** Bagging and Boosting. [Online] http://www.cs.rit.edu/~rlaz/prec20092/slides/Bagging_and_Boosting.pdf.

36. **Sutton, D. Clifton.** Classification and Regression Trees, Bagging and Boosting. *Handbook of Statistics.* 24, 2005.

37. *Confidence Intervals for the Area under the ROC Curve.* **Cortes, Corrina a Mohri, Mehryar.** NIPS.

38. **Gutierrez-Osuna, Ricardo.** Validation. [Online] http://research.cs.tamu.edu/prism/lectures/iss/iss_l13.pdf.

39. TreeBagger. [Online] MathWorks. [Citace: 24. 4 2014.] http://www.mathworks.com/help/stats/treebagger.html.

40. Ensemble Methods. *MathWorks.* [Online] 30. 4 2014. http://www.mathworks.com/help/stats/ensemble-methods.html#bsw8au3.

41. **Doi, Kunio.** Computer-aided diagnosis in medical imaging: Historical rewiev, current status and future potentials. *Computerized Medical Imaging and graphics.* 2007, Vol. 31, pp. 197-211.

42. **Swensen, Stephen J.** Screening for Lung Cancer with Low-Dose Spiral Computed Tomography. *Am. J. Respir. Crit. Care Med.* 2002, vol.165, pp. 508-513.

43. **Park, EA, a další.** Efficacy of computer aided-detection system and this-slab maximum intensity projection technique in the detection of pulmanary nodules in patients with resected metastases. *Inverst Radiol.* 44, 2009, Sv. 2, stránky 150-13.

The materials were also searched in question and answer sites for people interested in statistics and data mining (Cross Validated[11]) and programming (stackoverflow[12])

---

[11] Available on: http://stats.stackexchange.com/
[12] Available on. http://stackoverflow.com/