# A novel spherical shell filter for reducing false positives in automatic detection of pulmonary nodules in thoracic CT scans

Sil van de Leemput[a], Frank Dorssers[b] and Babak Ehteshami Bejnordi[c]

[a]Department of Artificial Intelligence, Radboud University Nijmegen, The Netherlands;
[b]Department of Computing Science, Radboud University Nijmegen, The Netherlands;
[c]Department of Radiology, Radboud University Medical Center, Nijmegen, The Netherlands

## ABSTRACT

Early detection of pulmonary nodules is crucial for improving prognosis of patients with lung cancer. Computer-aided detection of lung nodules in thoracic computed tomography (CT) scans has a great potential to enhance the performance of the radiologist in detecting nodules. In this paper we present a computer-aided lung nodule detection system for computed tomography (CT) scans that works in three steps. The system first segments the lung using thresholding and hole filling. From this segmentation the system extracts candidate nodules using Laplacian of Gaussian. To reject false positives among the detected candidate nodules, multiple established features are calculated. We propose a novel feature based on a spherical shell filter, which is specifically designed to distinguish between vascular structures and nodular structures. The performance of the proposed CAD system was evaluated by partaking in the ANODE09 challenge, which presents a platform for comparing automatic nodule detection programs. The results from the challenge show that our CAD system ranks third among the submitted works, demonstrating the efficacy of our proposed CAD system. The results also show that our proposed spherical shell filter in combination with conventional features can significantly reduce the number of false positives from the detected candidate nodules.

**Keywords:** Computer-aided diagnosis, Lung nodule detection, Spherical shell filter, ANODE09 challenge

## 1. INTRODUCTION

Lung cancer is the leading cause of cancer death worldwide. In 2012, lung cancer accounted for more than 13% of all cancer cases and 27% of all cancer deaths in the United States.[1] Early detection of lung cancer is crucial for improving the survival-rate of the patients. Traditionally, detection of nodules for a patient is done using a computerized tomography (CT) scan which is evaluated by a radiologist. The radiologist has to go through all the slices of the CT scan to detect nodules. The task of detecting the nodules is, however, laborious and prone to subjectivity. Computer aided diagnosis (CAD) has a great potential in helping the radiologist to detect candidate nodules by highlighting certain areas which might be interesting for the radiologist.[2] CAD systems do not suffer from habituation and present an objective assessment of a detected candidate nodule, removing any subjectivity from diagnosis.

A vast body of work exists on the topic of lung nodule detection[3–9] that can provide a useful pre-selection of possible pulmonary nodules, which reduce the time required for the radiologists to make their diagnosis and improve their performance.[2] However, among the potential nodules detected by such systems a majority are false positives resulting from lung vessels, lung boundaries, and CT scanner noise. The performance of the CAD systems can be significantly enhanced by developing features that can discriminate between true nodules and artifacts. Within the literature, the latter step is referred to as false positive reduction.[10]

For false positive reduction usually a machine learning model is trained using a set of annotated candidate nodules as ground truth, to recognize true pulmonary nodules from false positives. One of the key components for training such a model is to select features for each of the candidate nodules that capture characteristic differences between false positives and nodules. There are a plethora of features within the literature.[4–6,8,11] Features can be based on CT scan intensities, morphometrics (e.g. sphericity), or contextual information (e.g. the distance of the nodule to the boundary of the lung).

Further author information: (Send correspondence to B. Ehteshami Bejnordi)
B. Ehteshami Bejnordi: E-mail: Babak.EhteshamiBejnordi@Radboudumc.nl

In this paper we present a CAD system for detecting lung nodules in CT scanned images. Our CAD system initially segments the lung and detects candidate nodules for further analysis. The proposed system makes use of a novel blobness feature called spherical shell filter in combination with other conventional features described in the literature to discriminate between true detected candidate nodules and false positives. Our proposed feature is specifically designed to discriminate between pulmonary vascular structures and spherical structures in preprocessed CT imagery. The performance of our CAD system is evaluated using the dataset from the ANODE09 challenge.[12] *

## 2. METHODS

### 2.1 Lung segmentation

Lung segmentation is an essential initial step in the development of a computer-aided lung nodule detection system, as it restricts the space for candidate nodule detection and classification which consequently helps to reduce false positives occurring outside of the lung area. An overview of the algorithm used for segmentation of the lungs is shown in Figure 1. First two image masks were generated from the raw image by applying two thresholds of respectively all CT intensities above -268 and all CT intensities above -948. These threshold values were empirically determined using some training images. The resulting masks were then used to segment the outer boundaries of the lung. This was done by taking the inverse of the former mask and then using a logical AND-operation with the latter mask and consequently only taking the biggest connected chunk of voxels from the resulting image. The resulting image gives the outlines of the lungs, but contains many holes that were filled using a morphological closing operation with a kernel of size $15 \times 15 \times 15$.

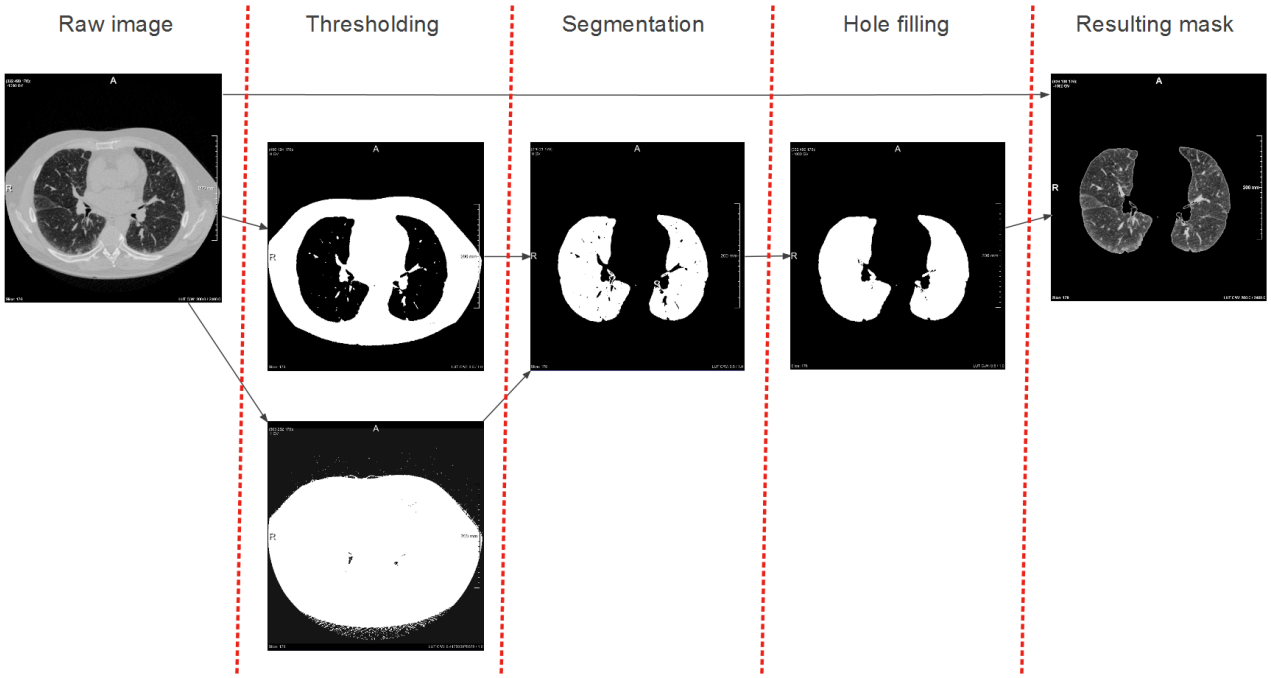



Figure 1: Overview of the lung segmentation

### 2.2 Candidate nodule detection

Laplacian of Guassian (LoG) filter is a circular-symmetric filter which is widely used for blob detection. This filter is in particular useful for detecting bright blobs surrounded with dark pixels or dark blobs surrounded by bright pixels. Therefore, these filters appear useful for candidate nodule detection,[4] since nodules are generally blob-like even when bordering lung vessels or lung boundaries. However, LoG filters have a maximum response to a particular blob size. Because nodules can be different in size we combined two differently sized LoG filters by using normalized kernels for

*The ANODE09 grand challenge for automatic nodule detection `http://anode09.grand-challenge.org/`).

both filters with the following parameters: $(\sigma = 3, \text{kernel extent} = 7)$ and $(\sigma = 5, \text{kernel extent} = 11)$. Next the voxel-wise maximum between both filtered images was calculated. Then using an empirically established threshold, only the pixels with intensities above $1.725e + 19$ were kept.

## 2.3 Candidate feature extraction

Table 1 presents an overview of the features used in this study, which were extracted from CT images. These features can be broadly divided into 4 classes: morphometric features, intensity based features, contextual features and blobness features. Morphometric features describe the geometry of the object. Intensity based features are derived from the statistics of the histogram of the gray values inside the candidate object. Contextual features describe the position of the object inside the lung and its distance to the lung boundary. Our proposed blobness feature called spherical shell filter is designed to identify isolated lung nodules and is described in the next section.

Some features require determining the centre-point or centroid of the three-dimensional candidate objects. However, determining those centroids can be challenging, since the shape of candidate nodules can be highly irregular. Using a simple approach, one could draw a bounding box around the pixels of the candidate nodule and take the exact centre of the box as a centroid. However, in this case for example a curve-shaped nodule might miss the body of the nodule. That is not desirable, since then the feature will not reflect the property of the nodule very well and thus will very likely reduce the classification performance. To address this problem, for each of the candidate nodules two different centroids were established: the bounding box centroid and the center of mass. The first is defined as the center of the bounding box (i.e. minimal box which contains all of the candidate voxels), whereas the second is calculated by taking the average over the x, y and z dimensions for the non-background voxels of a candidate. The later centroid is thus influenced by the mass distribution. Both centroids were used to calculate various features from Table 1.

| | | | | | |
|---|---|---|---|---|---|
| | | **Features** | | | |
| $M$ | Volume [6] | $M$ | Major axislength (xy, yz, xz) [11] | $I$ | Contrast region vs background [4] |
| $M$ | Surface area (xy, yz, xz) [6*] | $M$ | Minor axislength (xy, yz, xz) [11] | $I$ | Contrast region vs bounding box [4] |
| $M$ | Circularity [7] | $M$ | Cube compactness [8] | $I$ | Standard deviation CT value [6] |
| $M$ | Eccentricity [11] | $M$ | Compactness [7] | $I$ | Average CT value [6*] |
| $M$ | Diameter [7*] | $M$ | Elongation [8] | $I$ | CT value in $3 \times 3 \times 3$ kernel [*] |
| $M$ | Slice area [8] | $C$ | Minimum distance to lung [3] | $B$ | Spherical shell avg. & max. |

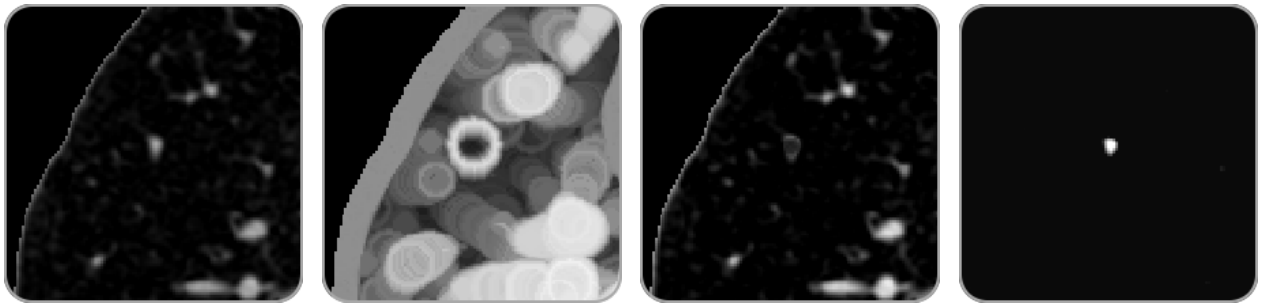
Table 1: Candidate nodule features and their types. * indicates that a feature was calculated using both the center of mass and the bounding box centroid of the candidate nodule. (xy, yz, xz) indicates that the feature is calculated for three 2D slices in the xy, yz and xz planes. Feature types are indicated by the initial letters, $M$ denotes Morphometric features, $I$ denotes Intensity based features, $C$ denotes Contextual features and $B$ blobness features.

### 2.3.1 Spherical shell filter

The majority of the false positives detected by lung nodule CAD systems are caused by pulmonary blood vessels. Blood vessels and vessel junctions are usually differentiated from lung nodules by having multiple attachment to other vessels. For this reason, we chose to design a feature which could discriminate between a true nodule and these false positives. Our proposed feature is inspired by the 2D nucleus detection filter designed by Moshavegh et al.[13] for detecting free-lying cell nuclei in Pap smear images. The algorithm described in the paper makes use of gray-scale annular closing operation to identify free-lying cell nuclei. The effect of this operator is to remove isolated dark spots in the gray-scale image.

The same idea for distinguishing candidate cell nuclei can be adapted for candidate lung nodules identification within CT scans. This was achieved by adapting the filter for use in 3D volumes, which effectively promotes the annular structuring element to a spherical shell volume, hence the name of the filter. Furthermore, instead of detecting dark cell nuclei, we are interested in detecting nodules that have bright intensities, therefore the gray-scale closing operation was changed to an opening operation.

Figure 2 shows the different steps for applying the filter. First the original segmented lung image is filtered using a gray-scale dilation with a spherical shell structuring element, then the voxel-wise minimum is calculated between the filtered image and the original lung segmentation. Finally by taking the difference between the resulted image shown in



(a) (b) (c) (d)

Figure 2: Example use of the spherical shell filter. (a) the segmented lung. (b) gray-scale dilation with a spherical shell structuring element. (c) Pixel-wise minimum of *(a)* and *(b)*. (d) arithmetic difference between *(c)* and *(a)*.

Figure 2c and the original image in Figure 2a isolated nodules are identified. For our CAD we used a single spherical shell filter with an inner and outer radii of respectively $8$ and $9$ voxels. For each candidate nodule mask, the maximum and average values of the resulting image are computed to serve as features. For the average case, zero valued voxels were excluded from the calculation.

### 2.4 Nodule classification

For each candidate nodule the set of features presented in Table 1 were computed. A random forest classifier with $100$ decision trees was then trained on the extracted features from the annotated cases. Candidates were labeled as positive if their mass coincided with an annotated nodule and as negative if this was not the case.

## 3. EMPIRICAL EVALUATION

To evaluate the performance of our system we took part in the ANODE09 challenge,[12] that was created to compare CAD systems for the automatic detection of pulmonary nodules in chest CT scans. The data available for this challenge are 5 annotated cases (CT scans) for system training and 50 cases for evaluation. One scan could have multiple annotations where each annotation corresponds to a specific nodule indicating its position and the category it belongs to. There are two major types of annotations in this cohort: 1) a true nodule, 2) a finding that is not a relevant nodule according to the protocol used in the screening study. For this challenge only nodules of the first category are considered relevant, which represent "relevant" or "actionable" nodules. These nodules can be described as the nodules that a CAD system should detect, since they may represent lung cancer. Irrelevant findings may indicate nodules that are too small to be relevant, nodules with benign characteristics, or findings which mimic a nodule but are not considered as nodules according to the observer. The markings around irrelevant findings by the CAD system are ignored (i.e. they are neither counted as false positives (FP) nor true positives (TP)).

The ANODE09 evaluation system uses free-response receiver operating characteristic (FROC) analysis for comparing different CAD systems. FROC analysis is a tool to characterize the performance of a detection system at different decision thresholds. FROC curve plots the sensitivity of the detection system as a function of the average number of false positives per CT scan.

### 3.1 Results

Figure 3 shows the FROC curve of our CAD system on the ANODE09 evaluation dataset. Table 2 presents the sensitivity values of our system at different false positive rates per scan on this dataset for different types of nodules. Two conditions were evaluated: with and without employing the spherical shell filter feature. Our overall system score with the shell filter was $0.336$, which is determined by calculating the average sensitivity of the system over several predefined false positives rates: 1/8, 1/4, 1/2, 1, 2, 4 and 8. This result places our CAD system third on the ANODE09 leader-board. The results show that our CAD system works remarkably well for detecting isolated and peri-fissural nodules. Furthermore, it shows that our system has satisfactory performance for detecting vascular and big nodules too. In comparison, the results without the spherical shell filter had an average sensitivity over all nodules of $0.309$ and had lower performance on average on

the majority of the nodule types demonstrating that the spherical shell filter provides a considerable increase in the overall performance of the proposed CAD system.

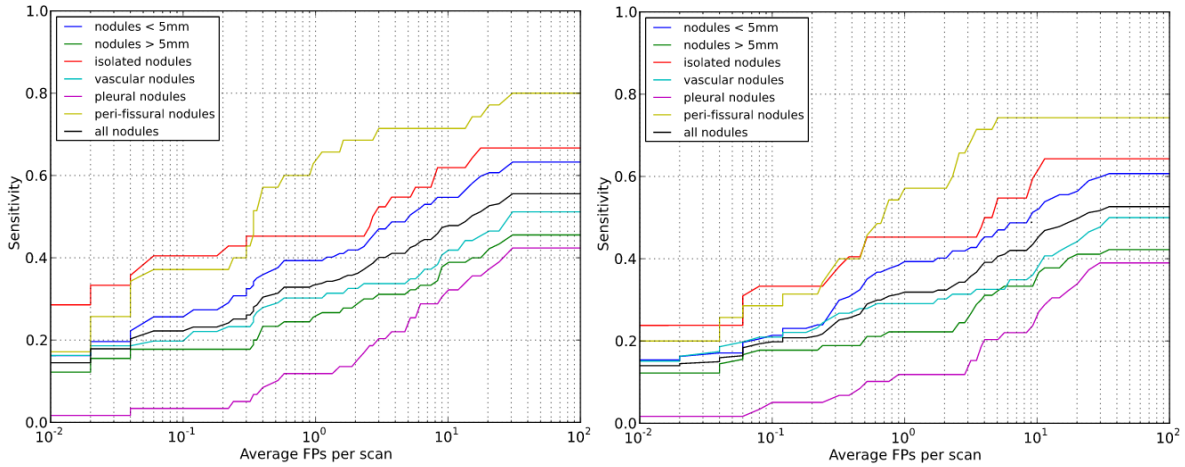



Figure 3: FROC plots of the sensitivity versus the number of false positives per scan for different nodule types on the ANODE test dataset. The left plot shows the FROC for the complete CAD system. The right plot shows the FROC for the CAD system without the spherical shell filter. Images were generated and obtained using the ANODE09 online system.[12]

| **FPs/scan** | **1/8** | **1/4** | **1/2** | **1** | **2** | **4** | **8** | **average** |
|---|---|---|---|---|---|---|---|---|
| all nodules | 0.232 | 0.251 | 0.312 | 0.335 | 0.360 | 0.411 | 0.454 | *0.336* |
| isolated nodules | 0.405 | 0.429 | 0.452 | 0.452 | 0.452 | 0.548 | 0.601 | 0.477 |
| large nodules | 0.178 | 0.178 | 0.233 | 0.258 | 0.284 | 0.311 | 0.340 | 0.255 |
| peri-fissural nodules | 0.371 | 0.400 | 0.571 | 0.636 | 0.686 | 0.714 | 0.714 | 0.585 |
| pleural nodules | 0.034 | 0.051 | 0.099 | 0.119 | 0.146 | 0.220 | 0.288 | 0.137 |
| small nodules | 0.274 | 0.308 | 0.373 | 0.393 | 0.419 | 0.487 | 0.541 | 0.399 |
| vascular nodules | 0.221 | 0.233 | 0.289 | 0.302 | 0.326 | 0.337 | 0.379 | 0.298 |

| **FPs/scan** | **1/8** | **1/4** | **1/2** | **1** | **2** | **4** | **8** | **average** |
|---|---|---|---|---|---|---|---|---|
| all nodules | 0.208 | 0.222 | 0.282 | 0.319 | 0.324 | 0.390 | 0.420 | *0.309* |
| isolated nodules | 0.333 | 0.339 | 0.437 | 0.452 | 0.452 | 0.498 | 0.548 | 0.437 |
| large nodules | 0.178 | 0.189 | 0.204 | 0.222 | 0.222 | 0.310 | 0.333 | 0.237 |
| peri-fissural nodules | 0.314 | 0.350 | 0.438 | 0.571 | 0.571 | 0.714 | 0.743 | 0.529 |
| pleural nodules | 0.051 | 0.053 | 0.096 | 0.119 | 0.119 | 0.202 | 0.220 | 0.123 |
| small nodules | 0.231 | 0.247 | 0.342 | 0.393 | 0.402 | 0.452 | 0.487 | 0.365 |
| vascular nodules | 0.221 | 0.247 | 0.279 | 0.291 | 0.302 | 0.326 | 0.349 | 0.288 |

Table 2: The reported sensitivity values on the ANODE09 test dataset for different numbers of false positives per scan. The top table contains the values for the complete CAD system, the bottom table contains the values for the CAD system without the spherical shell filter. Tables were generated and obtained using the ANODE09 online system.[12]

## 4. CONCLUSIONS AND DISCUSSION

In this paper we presented a CAD system for the detection of nodules in chest CT scans. Our proposed CAD system makes use of a novel spherical shell filter specifically designed to distinguish between small spherical shaped regions and vessel-like structures. The results of the empirical evaluation obtained from the ANODE09 challenge shows that our system achieves a high score for detecting relevant nodules in lung CT scans. It is noteworthy to mention that, the current result of the proposed CAD system was obtained by using solely 5 training cases. For comparison, the two better scoring

systems ISI-CAD[10] $(0.632)$ and FlyerScan[8] $(0.552)$ were both trained on over more than $50$ training cases. In addition, in comparison with the competing systems, ISI-CAD has a distinct advantage since it was trained using a larger training set from the same lung cancer trial. The performance of the proposed CAD system, whilst still quite high, could be improved. A review of the missed nodules indicate that, due to the rather rough lung segmentation several pleural nodules, i.e. the nodules bordering the outline of the lungs, are missed. More advanced methods will be used to enhance the lung segmentation result, e.g. using a rolling ball algorithms.[14] In addition, the CAD system was trained on only 5 chest CT scans (with 2 cases containing no nodules). In the future we intend to train and test the proposed CAD system on a larger dataset.

## REFERENCES


[1] Siegel, R., Naishadham, D., and Jemal, A., "Cancer statistics, 2012," *CA: a cancer journal for clinicians* **62**(1), 10–29 (2012).

[2] De Dombal, F., Leaper, D., Staniland, J. R., McCann, A., and Horrocks, J. C., "Computer-aided diagnosis of acute abdominal pain," *British medical journal* **2**(5804), 9 (1972).

[3] Jacobs, C., van Rikxoort, E. M., Twellmann, T., Scholten, E. T., de Jong, P. A., Kuhnigk, J.-M., Oudkerk, M., de Koning, H. J., Prokop, M., Schaefer-Prokop, C., et al., "Automatic detection of subsolid pulmonary nodules in thoracic computed tomography images," *Medical image analysis* **18**(2), 374–384 (2014).

[4] Fotin, S. V., Reeves, A. P., Biancardi, A. M., Yankelevitz, D. F., and Henschke, C. I., "A multiscale laplacian of gaussian filtering approach to automated pulmonary nodule detection from whole-lung low-dose ct scans," in [*SPIE Medical Imaging*], 72601Q–72601Q, International Society for Optics and Photonics (2009).

[5] Choi, W.-J. and Choi, T.-S., "Automated pulmonary nodule detection system in computed tomography images: A hierarchical block classification approach," *Entropy* **15**(2), 507–523 (2013).

[6] Teramoto, A. and Fujita, H., "Fast lung nodule detection in chest ct images using cylindrical nodule-enhancement filter.," *Int. J. Computer Assisted Radiology and Surgery* **8**(2), 193–205 (2013).

[7] Tartar, A., Kilic, N., and Akan, A., "A new method for pulmonary nodule detection using decision trees," in [*Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*], 7355–7359, IEEE (2013).

[8] Messay, T., Hardie, R. C., and Rogers, S. K., "A new computationally efficient cad system for pulmonary nodule detection in ct imagery.," *Medical Image Analysis* **14**(3), 390–406 (2010).

[9] Li, Q., Li, F., and Doi, K., "Computerized detection of lung nodules in thin-section ct images by use of selective enhancement filters and an automated rule-based classifier," *Academic radiology* **15**(2), 165–175 (2008).

[10] Murphy, K., van Ginneken, B., Schilham, A. M., De Hoop, B., Gietema, H., and Prokop, M., "A large-scale evaluation of automatic pulmonary nodule detection in chest ct using local image features and k-nearest-neighbour classification," *Medical Image Analysis* **13**(5), 757–770 (2009).

[11] Yang, M., Kpalma, K., Ronsin, J., et al., "A survey of shape feature extraction techniques," *Pattern recognition* , 43–90 (2008).

[12] van Ginneken, B., Armato III, S. G., de Hoop, B., van Amelsvoort-van de Vorst, S., Duindam, T., Niemeijer, M., Murphy, K., Schilham, A., Retico, A., Fantacci, M. E., et al., "Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: the anode09 study," *Medical image analysis* **14**(6), 707–722 (2010).

[13] Moshavegh, R., Bejnordi, B., Mehnert, A., Sujathan, K., Malm, P., and Bengtsson, E., "Automated segmentation of free-lying cell nuclei in pap smears for malignancy-associated change analysis," in [*Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*], 5372–5375, IEEE (2012).

[14] Armato III, S. G. and Sensakovic, W. F., "Automated lung segmentation for thoracic ct: Impact on computer-aided diagnosis1," *Academic Radiology* **11**(9), 1011–1021 (2004).