

International Journal of Intelligent Computing and Cybernetics
© Emerald Group Publishing Limited

Algorithm fusion to improve detection of lung cancer on chest radiographs

Gergely ORBÁN*

*Department of Measurement and Information Systems, Budapest University of Technology and Economics, Magyar tudosok korutja 2.
Budapest, 1117, Hungary
orbanger@mit.bme.hu
<http://home.mit.bme.hu/~orbanger>*

Gábor HORVÁTH

*Intelligent Systems Research Group, Magyar tudosok korutja 2.
Budapest, 1117, Hungary
horvath@mit.bme.hu
<http://home.mit.bme.hu/~horvath>*

Received (08-May-2011)

Revised (23-June-2011; 10-July-2011; 14-November-2011)

Accepted (21-November-2011)

Purpose - The purpose of this paper is to show an efficient method for the detection of signs of early lung cancer. Various image processing algorithms are presented for different types of lesions, and a scheme is proposed for the combination of results.

Design/methodology/approach - A computer aided detection (CAD) scheme was developed for detection of lung cancer. It enables different lesion enhancer algorithms, sensitive to specific lesion subtypes, to be used simultaneously. Three image processing algorithms are presented for the detection of small nodules, large ones, and infiltrated areas. The outputs are merged, the false detection rate is reduced with four separated support vector machine (SVM) classifiers. The classifier input comes from a feature selection algorithm selecting from various textural and geometric features. A total of 761 images were used for testing, including the database of the Japanese Society of Radiological Technology (JSRT).

Findings - The fusion of algorithms reduced false positives on average by 0.6 per image, while the sensitivity remained 80%. On the JSRT database the system managed to find 60.2% of lesions at an average of 2.0 false positives per image. The effect of using different result evaluation criteria was tested and a difference as high as 4 percentage points in sensitivity was measured. The system was compared to other published methods.

Originality/value - This study proves the usefulness of lesion enhancement decomposition, while proposing a scheme for the fusion of algorithms. Furthermore, a new algorithm is introduced for the detection of infiltrated areas, possible signs of lung cancer, neglected by previous solutions.

*Marczibányi tér 5., Budapest, 1022, Hungary

Keywords: lung nodule; infiltrated area; chest radiograph; CAD; lung cancer; early detection.

Paper type: **Research paper**

1. Introduction

Lung cancer is one of the most common causes of cancer death. Survival rate is extremely low, as many cures are only effective in the early and symptomless stage of the disease. Screening can help early diagnosis, but a cheap, side effect-free method, with high sensitivity to early signs, has to be used to enable mass usage. Standard chest radiography meets these requirements, except that current methods have moderate sensitivity. Research results of Muhm *et al.* [1983], Quekel *et al.* [1999], Shah *et al.* [2003] and Doi [2007] showed that many of the tumors are overlooked by radiologists in the early stage. The aim of computer aided detection – referred to as CAdE to prevent confusion with computer aided diagnosis – is to automatically detect lesions, helping the examiner. According to recent studies of Kobayashi *et al.* [1996], MacMahon *et al.* [1999] and De Boo *et al.* [2009] efficiency of lung screening can be improved by using a CAdE system; however, Hoop *et al.* [2010] warns that further training of the users may be necessary, as they often find it hard to differentiate between true and false findings of the machine. The goals of these systems are to find as many true lesions as possible – have high sensitivity – and produce only a small number of false detections – to give the examiner less extra work.

Cancerous tumors can appear on chest X-rays – on those where denser structures are darker – as approximately circular, homogeneous and dark shadows called lung nodules, but also as irregular, dark structures called infiltrated areas. However, these shadows can be signs of other diseases – like tuberculosis (TB) or other interstitial lung diseases (ILD) –, and in most of the cases, the exact cause cannot be determined based only on the radiograph. Thus all suspicious patients should be marked and referred for further examination. Considering lesion size, structures with largest dimension smaller than 30 mm are usually treated differently. According to a lung tumor staging method reported by Mountain [1997], these objects can sign early stage (IA) tumors with a relatively high, 67% five-year survival rate. However, larger tumors cannot be dismissed, as under certain circumstances they still indicate a stage (IB) with a five-year survival rate of approximately 57%.

In the past decades many efforts have been made to automatically detect the lesions by various algorithms. All of the proposed methods have a more-or-less common scheme. The authors usually assume a model that describes the targeted objects, and apply image processing algorithms to enhance nodules based on this model. The models are designed to include all the possible variations of the targeted lesions to minimize the number of false negative cases. This usually implies that many false positive regions are also enhanced. After extracting the enhanced, lesion-model-like areas, the number of false findings is reduced with a classifier. The input

of the classifier consists of various features that can ideally help to distinguish true and false positive findings.

Most of the studies focus on small, hard-to-detect lung nodules, pursuing cancerous tumors. They almost always describe the lesion with a single model. Although measurements are present for every system, these are rarely comparable, as some details of the test methods are missing or the test database is not available. Next, we will briefly introduce some of these existing solutions.

Giger *et al.* [1990] used the so-called difference image technique that creates a nodule enhanced and a nodule suppressed image by image processing algorithms, and outputs the difference which ideally eliminates everything but the shadow of nodule-like objects. They utilized various geometry features to leave out false findings; however, they used less powerful methods for classification. Lo *et al.* [1995] applied a matched filter and sphere template matching to enhance round shaped objects, and filtered false results with a convolutional neural network (CNN). They measured good performance on a small set of images. Yoshida *et al.* [1995] used two enhancers: a difference image algorithm and wavelet transform sensitive to subtle nodules. Here, the authors integrated two different – though overlapping – lesion models, used them in combination and got superior results to the single algorithm solution. However, they did not incorporate any kind of false positive reduction technique. Xu *et al.* [1997] employed another difference image algorithm. For false positive reduction, they created a hybrid system using an artificial neural network (ANN) and a rule-based technique. Carreira *et al.* [1998] took normalized cross-correlation to enhance areas similar to a nodule model and classified nodule candidates in curvature space. Penedo *et al.* [1998] investigated the usefulness of neural network for nodule enhancement. They utilized another ANN to select only true findings. They presented the results on a moderate-sized, private database extended by simulated nodules. Mao *et al.* [1998] evaluated the nodule enhancing capabilities of fragmentary window filtering. The authors presented preliminary results on a few generated radiographs. Li *et al.* [2001] introduced a technique for false positive reduction by comparing candidates to a nodule and a non-nodule template database. Wei *et al.* [2002] used a nodule model based on radial gradient vectors by applying adaptive ring filter. A statistical classifier based on Mahalanobis distance separated true and false findings. The classifier input was a result of a feature selection algorithm choosing relevant ones from 210 features. They evaluated the system on a radiograph database created by the JSRT. This database later became a standard for publishing comparable nodule finder results. The presented system showed high sensitivity, but the large number of included features suggests a risk of over-training. Keserci and Yoshida [2002] used the common difference-image scheme. In this case, they applied matched filter for nodule enhancement, while smooth-ring filter for suppression. Wavelet snake based geometric features and morphological operators helped an ANN to classify candidates. They tested the algorithm on a subset of the JSRT database. Coppini *et al.* [2003] integrated a

multi-scale Laplacian of Gaussian (LoG) filter and a Gabor-filter to raise suspicious areas, while false findings were eliminated by an ANN using fuzzy coding. They published comparable results on the JSRT database; however, the number of false positives was high, relative to sensitivity. Yoshida [2004] developed an algorithm for false positive reduction exploiting the quasi symmetry of the lung. They proposed a wavelet based image registration algorithm that allows contralateral-subtraction. The performance of the whole detection process was not published there. Suzuki *et al.* [2005] used the difference-image technique to enhance nodules, and multiple methods to eliminate false positives. The latter involves linear discriminant analysis and rule-based methods followed by a multiple massive-training artificial neural network. Low false positives were measured but on a relatively small, private database. Shiraishi *et al.* [2006] enhanced nodules with average radial gradient filtering. The filter worked adaptively based on anatomical classification. Three types of ANN, and symmetry analysis helped to classify candidates. They experienced a relatively high number of false positives; however, the numbers are not comparable because of the private database used. Schilham *et al.* [2006] searched for Gaussian scale space maxima in multi-scale to detect different sized nodules and a k -nearest neighbor (k -NN) classifier. They tested the algorithms on the JSRT database. Their solution performed well only at higher false positive rates. Campadelli *et al.* [2006] applied a filter searching for round-shaped objects in multi-scale and an SVM to classify initial results. They included opaque areas – lying outside of the viewable lung area – in the search and achieved good results on the JSRT database; however, some details of the test methods were not described. Pereira *et al.* [2007a] introduced a new filter based on the convergence of radial gradients called the sliding band filter (SBF). They published results on the JSRT database without a false positive reduction step. Shiraishi *et al.* [2007] involved lateral images to improve accuracy while using edge gradients to enhance nodules, and an ANN combined with a rule-based technique to filter them. Although involving lateral images improved sensitivity, these extra views are usually not available at routine screening. Acharyya *et al.* [2008] proposed a feature characterizing nodule smoothness using a random process-based fractal image processing technique. Although classification performance was promising, they validated the method only on a small number of images. Hardie *et al.* [2008] introduced a modified convergence index (CI) filter to find initial candidates and a Fischer linear discriminant classifier to select only positive findings. Although a linear classifier has many advantages, it bottlenecks the system, as the separation problem is almost certainly nonlinear. The authors tested their solution on the JSRT database while training on a large set of images from another database. Weng *et al.* [2009] proposed a solution based on a difference image technique and SVM classification. They used the statistics of gradients as features. Although they included the JSRT in the testing, they published results only on a merged database. Snoeren *et al.* [2010] proposed a method to train a system with simulated nodules. For the extraction of nodule candidates they utilized a

neural network based on gradient convergence and location features. They did not employ a false positive reduction step. Tests were carried out on a subset of the JSRT database. Chen *et al.* [2011] used two-stage nodule enhancement based on morphological operations and radial gradient analysis. False positive reduction was done by a nonlinear SVM; however, hyperparameter selection was done heuristically. Results were presented in detail using the JSRT database.

Although a tremendous amount of work has been invested in all of these systems, they still suffer from a high number of false positive findings, justifying new attempts to contribute alternative methods for the improvement of results.

Another common property of recently published systems is that they only target circular and mostly small – <30 mm – lung nodules, neglecting various shaped infiltrated areas and large lesions. This narrow scope may be inadequate for a CAD at screening. According to MacMahon *et al.* [1990] infiltrated areas are by far the most frequent findings on chest radiographs, which remains true even if we take only subtle, hard-to-detect ones. Furthermore, infiltrated areas can indicate other serious diseases like pneumonia, TB or other ILDs, which are also useful to mark at screening. A very few attempts were made to detect these lesions. Monnier-Cholley *et al.* [1998] used Fourier transform to calculate features of the power spectrum of texture. They omitted perihilar regions to keep false positive rates low. Ginneken *et al.* [2002] proposed a system to detect signs of TB. They used a k -NN classifier based on textural and symmetry features to separate abnormal regions. However, they did not segment the lesions allowing further analysis. They conducted tests on patients with TB and ILD. Sundararajan *et al.* [2010] targeted the detection of pneumoconiosis. They used various textural features and separate SVM classifiers for disjoint segments of the lung, while a single classifier to label the image as a whole. They omitted the lesion segmentation step and only focused on a subset of lung disorders. Le [2011] developed a complex system for the detection of various lung disorders but the nodule detector employs only a simple image processing step without false positive reduction.

Due to the fact that almost all solutions tried to detect only small nodules, Ginneken *et al.* [2009] suggested a change of focus, claiming that finding only nodules are not enough to detect signs of mortal diseases. Our goal is to make a step towards a more comprehensive result set by including infiltrated areas and large nodules while not allowing false positive rates to increase.

Although large lesions are less likely to appear at an early stage of lung cancer, we are convinced that detecting them too – besides small nodules and minor infiltrated areas –, has several advantages. First, they complement the findings of a CADE system to be more comprehensive, which is a necessity if the goal is to partly automate the screening procedure – to prevent the examiner from dealing with obviously negative cases. Second, a system finding a comprehensive set of lesions can speed up analysis by requiring the examiner to draw less new markers, instead only to correct CADE findings. Last, the trust of the users – the physicians – can

only be gained if the system does not miss cases that are relatively easy for the human reader to detect. This trust, however, is essential to get the physicians to use the systems and provide feedback helping further development.

Finding a diverse set of lesions may require a new approach. The common solution is to create a single nodule model and use only one – or a few serially connected – filter(s) to enhance shapes matching the model. This methodology is present in all of the previously introduced systems, except for [Yoshida *et al.* (1995)], where the authors used two enhancers sensitive to different subtypes of nodules. In the current study, we generalized the idea, and created a framework that enables an arbitrary number of enhancer algorithms to be used simultaneously. We claim that the problem is complex enough to justify the application of the divide and conquer paradigm. Although algorithms exist for very general lesion detection, like symmetry based methods described in [Katsuragawa and Doi (2007)], because of their generality, they cannot be so efficient for relatively well-defined circular nodules. To get the desired, more comprehensive set of findings while keeping the sensitivity to nodules, we developed three lesion enhancer filters based on different lesion models. We used separate models specific to small nodules, to larger ones, and a more general model to infiltrated areas. When designing our lesion models and enhancers, our goal was to develop accurate algorithms while making their response as different as possible. This can ensure that a committee machine – like the system we designed – performs the best for a wide range of input cases [Haykin (1994)].

Looking at the publications in the field, one can notice that measurements and results are hardly comparable. This is mostly due to the different radiograph databases used, but minor differences in methodology can also cause an undesirable bias. A frequent one, the criterion used for accepting a candidate, is neglected in many publications. This problem was already emphasized [Li *et al.* (2010); Chen *et al.* (2011)] but a detailed measurement of difference on a public database is still missing. In this study we analyzed three commonly used criteria on the JSRT database. We also collected existing results that are more-or-less comparable, and attempted to carry out a fair comparison.

To sum up, our goals are threefold. First, we try to prove the utility of decomposing nodule enhancement along nodule type. Second, we introduce three algorithms for lesion enhancement, one of them using a completely new approach to detect infiltrated areas. Third, we compare our system to other state-of-the art nodule finders on a publicly available database and reflect on some problems preventing wider and more accurate comparison.

In the next section we will introduce the two radiograph databases used for the development and testing of our algorithms, and then describe the methods used in the CADe system.

2. Radiograph databases

We used two separate databases throughout the study. The first one is the work of the JSRT. It is publicly available, enabling us to publish comparable results. The second one is a larger and more diverse database, not available to the public due to privacy issues.

2.1. *The database of the JSRT*

The JSRT database created by the cooperation of several Japanese clinics contains 247 digitized – originally analogue – radiographs [Shiraishi *et al.* (2000)]. The resolution of images is 2048×2048 pixels, where the size of a pixel is 0.175 mm. 154 of the cases contain exactly one marked lung nodule validated by CT. Nodule diameters range from 5 mm to 60 mm, the average is 17 mm and the median is 15 mm. 140 of the nodules are smaller than 30 mm. The nodules are classified as obvious, relatively obvious, subtle, very subtle and extremely subtle expressing the difficulty of finding them for a typical radiologist. 14 of the cases are located in the opaque regions of the lung, which makes them impossible for some specific algorithms to find.

2.2. *The private database*

The JSRT database includes only a subset of signs that can indicate lung cancer. It mainly contains circular nodules smaller than 30 mm. To experiment with other types of lesions, we involved another chest X-ray database originating from the Pulmonological Clinic of Semmelweis University (PCSU) located in Hungary. The database consists of images of 514 patients, 364 of them containing at least one malignant lung nodule or infiltrated area. Only one radiograph is included for each patient. A sum of 210 nodules are present on 136 images while 348 infiltrated areas are marked on 232 images as some radiographs contain both types of lesions. The images were acquired by a digital X-ray machine in a resolution of approximately 6 megapixels and pixel size of 0.16 mm. Lesions were validated by ultrasound or CT. Nodule diameters range from 2 mm to 97 mm, the average and median is 22 mm and 16 mm respectively. 161 of the nodules are smaller than 30 mm. An effective diameter was calculated for every infiltrated region: $d_e = 2\sqrt{A/\pi}$, where A is the area of the region. The minimum, maximum, average and median of this effective diameter is 10 mm, 146 mm, 62 mm and 60 mm respectively. 58 cases are smaller than 30 mm. This database shows more diversity than the JSRT, as it contains several nodules larger than 30 mm, many infiltrated areas and patients with other disorders. Furthermore, the images are less aligned than the ones in the JSRT. On the other hand, in terms of details and contrast, the PCSU database is superior owing to direct digital image acquisition.

3. The lesion finder system

For our lesion finder solution we used the following three-step scheme. The first step segments the viewable area of the lung and frees the image from shadows of bones. The viewable area reduces the possible set of locations for lesions, as we only target the ones partly or completely inside this area. Bone shadow removal helps to see the lesions easier and eliminates frequent false positive findings due to bone shadow crossings. The algorithms used in this step are described in [Juhász *et al.* (2010)].

The second step, called lesion enhancement, highlights suspicious structures like the target lung nodules and infiltrated areas by using image processing algorithms. After normalization and resizing, the processing splits up to three threads based on the targeted nodule type. Constrained sliding band filter (CSBF) is used for the enhancement of smaller and subtle nodules, another method called large nodule filter (LNF) for larger ones with high contrast, while the novel outlier area filter (OAF) for infiltrated areas. Afterwards, the candidate collection converts the enhanced images into separate and segmented nodule candidates.

The last step reduces the number of false positive findings on the enhanced image with the help of a classifier. It begins with the calculation of features of candidates serving as an input for the classifier. Classification is done for each thread followed by a merging step that eliminates duplicate and highly overlapping results. A final classification is carried out on the merged set of candidates. The second and third major steps are illustrated in Figure 1 and described in detail in the sequel.

Figure 1 here

3.1. Preprocessing

Preprocessing of images involves simple resampling and normalization. The target size of images is approximately a quarter megapixel – corresponding to around 0.7 mm / pixel –, containing enough detail to find small nodules but keeping running times low. Normalization is to ensure that intensity values in images originating from different sources are comparable. Setting the mean and standard deviation to predefined values – by a linear transformation in intensity space – is a commonly used technique. Instead, our solution adjusts median value inside the viewable lung area and the deviation from the median, as the median showed more robust results. By choosing the median and the deviation to be 0.35 and 0.18 respectively, we ensure that only a few values are mapped outside the $[0, 1]$ interval, but most of the dynamic range is exploited. To prevent values lying outside the allowed intensity range of $[0, 1]$, the following mapping (S) is used.

$$S(x) = \begin{cases} \tanh \frac{x-l}{l} + l & x < l \\ x & l \leq x \leq u \\ (1-u) \tanh \frac{x-u}{1-u} + u & x > u \end{cases}, \quad (1)$$

where $l = 0.1$ and $u = 0.9$ are the lower and upper bounds of the linear mapping

interval. Values outside this interval are mapped by a sigmoid function. In the next sections the three different nodule enhancers are described.

3.2. The Constrained Sliding Band Filter

The first group of targeted lesions is small lung nodules. These are approximately circular and dark structures with a diameter of 5 mm to 30 mm. According to our experiences, at this level shape information can provide a better clue for finding them because they tend to have low contrast, making usual intensity based filters less effective. A commonly used filter family called Convergence Index (CI) approximates object borders and enhances a structure if its shape is approximately circular. A common property of round shaped objects is the radial direction of gradient vectors along their border. A general CI filter considers the surroundings of each pixel. The output depends on the angles of the vectors connecting the center and the surrounding points and the gradient vectors at these surrounding points. One of the most successful realizations is the Sliding Band Filter [Pereira *et al.* (2007b)].

The SBF algorithm considers each pixel of the image as a potential center of an object. For each center – as shown in Figure 2 – it slides a band in different directions within given bounds, while the band has a fixed width. For each band position the algorithm takes the points inside the band and sums the cosine of the angles of radial vectors – the vector connecting the center and the given point – and gradient vectors at the points. The final position of the band for a direction will be the one with the highest sum. More precisely, it assigns each pixel the following value.

$$SBF(x, y) = \frac{1}{N} \sum_{i=1}^N Cmax_i, \quad (2)$$

$$Cmax_i = \max_{R_{min} \leq r \leq R_{max}} \frac{1}{d} \sum_{m=n}^{n+d} \cos \theta_{im}, \quad (3)$$

where R_{min} , R_{max} are the bounds of the target object radius, N is the number of directions concerned, d is the width of the band, θ_{im} is the angle of the m^{th} gradient vector along the i^{th} radial direction and the corresponding radial vector.

Note that a high sum is caused by the convergence of negative gradient vectors towards the center and can sign an object border. For a round shaped, dark object, the negative gradients are convergent in every direction if the starting point was the center of the object. Using this idea, the algorithm sums the maximal band values in each direction, and a high final sum indicates a nodule.

Figure 2 here

One weakness of the algorithm is the independence of the bands in each direction, enhancing very spiculated and distorted objects. An intuitive solution is to apply a constraint on the bands in different directions. The CSBF – our modification of the

SBF – links the position of the bands, allowing smaller distortion. It ensures that the final band positions satisfy a circularity constraint controlled by a coefficient. The coefficient forces an upper bound to the ratio of the distance of the farthest and the closest band from the center.

Another modification we carried out reduces the sensitivity of the filter to random noise. Using only the direction of gradient vectors, the original SBF is able to enhance fairly dimmed objects; however, vectors of negligible length caused by image noise can mislead the filter. To reduce sensitivity to noise, we used vector length, but rescaled it using a piecewise linear approximation of a shifted sigmoid function denoted by r .

$$r(x, a, b) = \begin{cases} 0 & x < a \\ x - a & a \leq x < b \\ b - a & b \leq x \end{cases}, \quad (4)$$

where x is the original vector length, a and b are parameters. This eliminates the effect of some very small gradients, but won't let high contrast objects suppress dimmed nodules. Integrating the modifications, the enhanced pixel values can be calculated with the following formula.

$$CSBF(x, y) = \max_{R_{min} \leq r \leq \frac{R_{max}}{c}} \frac{1}{N} \sum_{i=1}^N Cmax_{ir}, \quad (5)$$

$$Cmax_{ir} = \max_{r \leq n \leq r * c} \frac{1}{d} \sum_{m=n}^{n+d} \cos \theta_{im} r(|g_{m,i}|, a, b), \quad (6)$$

where R_{min} , R_{max} are the bounds of the target object radius, c is the shape constraint coefficient, N is the number of directions concerned, d is the width of the band, $g_{m,i}$ is the m^{th} gradient vector along the i^{th} radial direction, θ_{im} is the angle of $g_{m,i}$ and the corresponding radial vector and a and b are the parameters of the length scaling function. An output identical to CSBF can be achieved by running several SBF filters with different bounds (R_{min} , R_{max}) and taking the minimum for each center; however, the execution times would be much greater. Due to the implementation, the CSBF takes only moderately longer than one SBF run. For properly large c values the CSBF works as a standard SBF, while for $c = 1$ it is identical to the Iris filter, another realization of the CI family [Pereira *et al.* (2007b)].

Parameter selection for CSBF was made heuristically. R_{min} and R_{max} were set to match the smallest and largest nodules to be found. The parameter d affects the sensitivity to noise and was set to 5.6 mm. N was set to 16, providing a good compromise between precision and speed. The circularity parameter was optimal around 1.2. The parameters of the length scaling function are $a = 0$ and $b = \frac{1}{400}$.

3.3. The Large Nodule Filter

Enabling the CSBF to detect large nodules – by increasing R_{max} and N – notably increases its run time. Furthermore, larger structures and intensity variations can considerably alter gradient directions. On the other hand, large nodules have higher contrast that can help detection and enable the usage of a simpler, more reliable algorithm. The reason for this higher contrast is the following. Consider a simple model for X-ray imaging, where image intensity is the fraction of X-rays that can pass through the matter. Assume that a nodule is homogeneous and its matter absorbs half of penetrating X-rays at length l . Image intensity at a given point ($F(x, y)$) can be calculated by the following formula.

$$F(x, y) = c \frac{1}{2^{d(x, y)/l}} \quad (7)$$

where c is the background intensity ($0 < c \leq 1$), and $d(x, y)$ is the depth of the nodule at the given position. Using this formula, we can approximate the intensity profile of a smaller and a larger nodule, as it can be seen on Figure 3. Note that the profile of the large nodule – due to the logarithmic nature of X-ray images – is more rectangular, resulting in a sharper transition, while the profile of the small nodule remains circular, causing a dimmer transition. The rectangular profile, together with a larger difference in intensity, results in a higher contrast.

Figure 3 here

The LNF aims to enhance nodules with a diameter between 30 mm and 75 mm and high contrast, but allows them to lie almost completely outside the viewable lung area. The basic idea behind the algorithm is a local contrast enhancement (LCE) followed by a top-hat filter. Our LCE algorithm is based on the method described by Lee [1980]. The modified output - $G(x, y)$ - is given by the following equations.

$$G(x, y) = \frac{1}{1 - \exp\left[-\left(F(x, y) - \frac{1}{|R(x, y)|} \sum_{(u, v) \in R(x, y)} F(u, v)\right)\right]}, \quad (8)$$

$$R(x, y) = \begin{cases} \{(u, v) | (u - x)^2 + (v - y)^2 < 2r^2\} \cap L & (x, y) \in L \\ \{(u, v) | (u - x)^2 + (v - y)^2 < 2r^2\} \cap /L & \text{otherwise} \end{cases}, \quad (9)$$

where F is the original image, L is the viewable lung area and r is the radius of the targeted nodule. The trimming of R with L ensures we have a homogeneous area completely inside or outside the lung. The rationale behind the logistic function is to get a result in between local normalization and local thresholding. An example output can be seen in Figure 4.

Top-hat filtering is a simple convolution by a cylinder shaped kernel C with radius r . The side of the cylinder is normally vertical, which maximizes filter output for perfectly circular and sharp shapes; however, if the target shape is somewhat distorted, the filter output heavily decreases. To keep a high filter output for slightly

12 *Gergely Orbán, Gábor Horváth*

distorted or blurred objects, the side of the cylinder is tilted. This gives the following kernel.

$$C(x, y) = \begin{cases} \frac{1}{P} & \sqrt{(x^2 + y^2)} < r - \frac{1}{d} \\ \frac{d}{P}(r - \sqrt{(x^2 + y^2)}) & r - \frac{1}{d} \leq \sqrt{(x^2 + y^2)} < r \\ \frac{d}{N}(r - \sqrt{(x^2 + y^2)}) & r \leq \sqrt{(x^2 + y^2)} < r + \frac{1}{d} \\ \frac{-1}{N} & r + \frac{1}{d} \leq \sqrt{(x^2 + y^2)} < \sqrt{2}r - \frac{1}{d} \\ \frac{d}{N}(\sqrt{(x^2 + y^2)} - \sqrt{2}r) & \sqrt{2}r - \frac{1}{d} \leq \sqrt{(x^2 + y^2)} < \sqrt{2}r \end{cases}, \quad (10)$$

$$P = \pi\left(r^2 - \frac{r}{d} + \frac{1}{3d^2}\right), \quad (11)$$

$$N = \pi r\left(r - \frac{\sqrt{2} + 1}{d}\right), \quad (12)$$

where d determines the steepness of the sides. $d \rightarrow \infty$ gives the vertical sided case. Note that P and N are to normalize the integral of the positive part to one and the negative part to minus one.

This method would also enhance – besides nodules – other dark structures like remainders of rib shadows or areas filled with vessels. According to our observations, nodules tend to be more homogeneous than these misleading areas, so to suppress them the filter output is weighted with the smoothness of the area. For smoothness the standard deviation of smoothed nodule pixels inside the viewable lung is calculated. An example result is shown in Figure 4. The method uses a multi-scale framework to detect different sized nodules.

Figure 4 here

To find nodules lying outside the viewable lung area, the top-hat filter is run for the entire image. This would enhance structures like the vertebra so the filter output is kept only where both negative and positive parts of the cylinder overlap with the viewable lung, and the area of intersection for both parts is greater than 15% of the filter part area. As a post processing step, the areas where the predicted nodule would lie outside the whole lung are suppressed. This requires an overestimation of the whole lung area, for which we use the following algorithm. The binary masks of the left and right viewable lung parts are dilated towards the centre and then eroded with the estimated nodule radius. Then the union is taken for the two parts. The resulting mask is consistent with our assumption that nodules can reside under the shadow of the heart, aortic arch or hemidiaphragm, but cannot hang out towards the side of the body. An example is shown in Figure 5.

Figure 5 here

3.4. The Outlier Area Filter

To detect infiltrated areas, neither of our previous approaches can be used, as the model of approximately round shaped objects will not fit these lesions. A typical infiltrated area is shown in Figure 6. Hence we decided to use mostly intensity information and make very little constraint on the shape. As infiltrated areas usually

do not create a definite border, we also rejected the usual approach of comparing the object to its surroundings. Instead, we used only absolute intensity. The basic principle of our algorithm is the following. For each location in the lung we estimate the distribution of intensity values for healthy cases and determine an intensity interval in which they lie with high probability. If the value for a specific patient lies outside the interval, we consider it less probable for a healthy case, so we mark it as suspicious. This algorithm can also be interpreted as an intensity based novelty detector.

Figure 6 here

The method is based on the assumption that intensity values of a lesion are distinctly darker at a fixed location, in other words, they come from a different distribution having little overlap with the intensity distribution of healthy cases. To support our assumption we analyzed these distributions for a few random locations, one of them is plotted in Figure 7.

Figure 7 here

To get comparable locations on different radiographs, we need a robust coordinate system. For this we decided to use a polar-like system – we will refer to it as lung polar coordinate system (LPCS) – illustrated in Figure 8. The origin (O) here is the center of gravity of the lung masks marking the viewable lung areas. For a given point p the first coordinate φ is the angle of the vector \vec{Op} with the vertical axis, mapped to $[-\pi, \pi]$. The second coordinate is a normalized distance $d(p, \varphi)$ depending on the location and the first coordinate.

$$d(p, \varphi) = \frac{\|p - I_1\|}{\|I_2 - I_1\|}, \quad (13)$$

where I_1 is the closest intersection of the half-line starting from the origin and crossing p with the lung mask. I_2 is the farthest intersection and $\|x\|$ denotes the Euclidean distance. To reduce running time, φ and d are quantized to 200-200 discrete values.

Figure 8 here

The implemented infiltrated area finder algorithm consists of two main phases. The first one is a preprocessing or training stage only run once in design time; the next one is the recall step, which runs for every analyzed image. Training is done on a large set of radiographs of healthy patients. First a median filter is applied with a disk shaped kernel of 4.2 mm radius to eliminate noises of a few millimeters. As infiltrated areas tend to be larger than this size, we will not lose true positives here. The next step maps every training image to LPCS and the mapped intensity values are used to create a histogram for each coordinate. The last step calculates the threshold below which an intensity value is to be considered suspicious. An upper threshold is not needed, as the lesions are darker than the usual.

To determine the lower threshold, we interpreted the algorithm as a pixel-wise hypothesis tester, where for each location the null hypothesis is:

$$H_0 : F(x, y) \text{ is produced by a healthy tissue,} \quad (14)$$

where $F(x, y)$ is the image intensity. This way we can control, the type I. error – its probability $P((x, y) \text{ is marked as unhealthy} | (x, y) \text{ is healthy})$ – by setting the significance level. According to our observations, marking less than 3% of image pixels will result in a very few false outputs as most of the markers are eliminated by the morphological operations in the recall phase. For this reason, we have chosen the desired threshold – the significance level – to be the three percentile of the estimated distributions. This means the algorithm will mark a healthy region – falsely – with a probability of approximately 0.03 if our database is representative enough. An alternative way would be to tune this parameter together with the classifier, but this would largely increase parameter space, causing learning to be slower and more difficult. Note that by setting the threshold with the described method, only the analysis of healthy images is necessary.

The algorithm of the recall phase is very similar to the learning phase. It is shown in the block for Algorithm 1. Here \ominus and \oplus denotes morphological erosion and dilation respectively. *LungMask* is a binary image indicating the viewable lung area. The last few steps carry out a morphological opening while discarding small hits close to the lung border. This is to eliminate false results caused by small structures and lung segmentation errors. An example result is shown in Figure 9.

Algorithm 1 here

Algorithm 1 OAF recall phase (Inputs: *Radiograph*, *LungMask*; Output: *Outliers*)

```

for all  $(x, y)$  do
   $Radiograph(x, y) \leftarrow \text{MedianFilter}(Radiograph, (x, y), r = 4.2 \text{ mm})$ 
end for
for all  $(\varphi, d)$  do
   $Radiograph_{polar}(\varphi, d) \leftarrow Radiograph(\text{LPSC2Cartesian}(\varphi, d))$ 
   $Outliers_{polar}(\varphi, d) \leftarrow Radiograph_{polar}(\varphi, d) \leq \text{PercOfHistogramAt}(\varphi, d, 3)$ 
end for
for all  $(x, y)$  do
   $Outliers(x, y) \leftarrow Outliers_{polar}(\text{Cartesian2LPSC}(x, y))$ 
end for
 $K \leftarrow \text{Disc}(r = 7 \text{ mm})$ 
 $Outliers_e \leftarrow Outliers \ominus K$ 
 $LungMask_e \leftarrow LungMask \ominus K$ 
for all  $(x, y)$  do
   $Outliers_e(x, y) \leftarrow Outliers_e(x, y) \text{ AND } LungMask_e(x, y)$ 
end for
 $Outliers \leftarrow Outliers_e \oplus K$ 

```

Figure 9 here

3.5. Candidate collection

Each nodule enhancer produces an output where greater intensity means a higher probability of a suspicious object, enabling the use of a unified candidate collector algorithm. The collector – run separately for each filter output – is described in Algorithm 2. Here c is a coefficient determining the proportion of total area covered by the collected candidates. This method proved to be more robust than the global thresholding of *EnhancerOutput*, as the detail and noisiness of images affect filter output especially for the CSBF – which for example, may cause more results on a smoother image. Another possibility is collecting a fixed number of candidates per image, but this may fail if the small structures near the mediastinum produce a lot of small false candidates.

Algorithm 2 here

Algorithm 2 Candidate Collector (Inputs: *Radiograph*, *EnhancerOutput*, *LungMask*; Output: *Candidates*)

```

TotalArea ← AreaOf(LungMask)
CoveredArea ← 0
while CoveredArea < c * TotalArea do
  (x, y) ← arg max(x,y) EnhancerOutput(x, y)
  LesionMask ← LesionSegmenter(Radiograph, (x, y))
  for all (x, y) ∈ LesionMask do
    EnhancerOutput(x, y) ← 0
  end for
  CoveredArea ← CoveredArea + AreaOf(LesionMask)
  Candidates.Add(LesionMask)
end while

```

The segmentation of lesions (*LesionSegmenter*) is different for each enhancer output. The border of infiltrated areas is the side product of the OAF, as connected regions will be candidates. For the remaining two filter outputs, nodule segmentation is done by an optimization algorithm trying to keep a circular shape while adjusting the border to high contrast discontinuities in several directions.

3.6. False Positive Reduction

The candidate collector finds approximately 20-30 candidates on the CSBF enhanced image, five on the LNF enhanced image and one on the OAF output. This number of markers – especially for the first two sets – has to be reduced, for which we use a support vector machine classifier (SVM or SVC) [Cortes and Vapnik (1995)]. The efficient application of an SVM requires good quality training data, a careful

choice of kernel function, an appropriate input vector, and hyperparameter tuning of the generalization coefficient (often denoted by C) and the kernel parameters.

Our training data comes from the two databases mentioned above – the JSRT and the PCSU database – using cross-validation to prevent overfitting. Positive samples of lesions are the validated findings of pulmonologists and radiologists, while negative samples are false findings of the candidate collector. We noticed that physicians do not always mark all the lesions on an image just the most important ones. These unmarked nodules could mislead our system if we use them as negative training samples. To overcome this issue, we decided to use false findings only in healthy images for training. As 30 times more negative samples were available than positive ones, we used a cost-sensitive version of the SVM. Cost sensitivity is achieved by using different C values for positive and for negative samples.

The capability of an SVM largely depends on the choice of the kernel function, for which we use the widespread isotropic Gaussian kernel. The input vector of the kernel consists of various features describing texture, geometry and location. To eliminate irrelevant features, we ran a simple feature selection algorithm. As we have 168 implemented features to choose from and expected only a few to contain valuable and irredundant information for classification, we applied a forward selection method. This simply observes the performance of the SVM while adding features to the base set – being empty at start – one-by-one. The feature providing the best performance is added to the base set and the procedure is restarted. We made the method more robust by adding feature removal phases and increased its speed by adding more features in each iteration simultaneously.

As the relevant features – described later in section 4.1 – turned out to be different for the output of the three nodule enhancing algorithms we decided to use separate classifiers for each result set. Using one classifier requires the union of relevant features, which would increase the number of dimensions reducing overall performance. Furthermore, tests showed that using multiple classifiers can save run time.

The SVM with isotropic Gaussian kernel requires two hyperparameters to be chosen by the user. These are C and the width parameter (σ) of the kernel. For tuning, we use an iteratively refining grid search being robust against local maxima, but having acceptable speed. According to our observations, the two hyperparameters are strongly dependent and the optimal solution lies in a one-dimensional subspace, described by a new parameter d .

$$\log C + \log \sigma = d, \tag{15}$$

This allows us to reduce the two-dimensional problem into two separate one-dimensional ones – the tuning of the d parameter and the searching inside the one-dimensional subspace. In other words, with proper hyperparameter transformation we managed to cut the complexity of the search.

4. Experimental results and discussions

To evaluate our system, we first introduce the relevant features given by the feature selection algorithm. Then we show the utility of algorithm fusion by comparing the hybrid system to simpler, serial versions. Afterwards, we demonstrate comparable results on the JSRT database and finally, we try to compare our system to some other published solutions.

For all of the tests where it is not explicitly specified we consider a result to be true positive if and only if its center of gravity is inside a nodule marked by a physician. Else we label it as false positive. We refer to this method as 'centroid criterion'. For the demonstration of our results we used free-response receiver operating characteristic (FROC) analysis [Chakraborty and Berbaum (2004)]. We calculate sensitivity as the fraction of true positive outputs divided by all lesions in the database, while the average number of false positives is the sum of all outputs that are not true positive divided by the total number of images. For the entire test we used fourfold cross-validation by randomly separating the images into four equally sized sets. A test run is the average of four iterations; where in each iteration we used the candidates from three sets for training and the remaining set for testing. We repeated each test at least 30 times with different random separations to reduce variance introduced by the random distribution of images.

4.1. Results of feature selection

Our feature selection technique showed that different features are relevant for the classification of different enhancer outputs. First, we shortly describe the ones for the output of the CSBF.

- (1) Mean of the average fraction under the minimum (AFUM) filter inside nodule border [Heath and Bowyer (2000)].
- (2) The maximum value of the AFUM filter inside the nodule.
- (3) The linearity of nodule border transition. This considers the intensity profiles perpendicular to the nodule border and calculates the error to the best fitting linear approximation.
- (4) The contrast close to the determined nodule border (C_b).

$$C_b = \frac{\bar{I}}{\bar{O}}, \quad (16)$$

$$I = \{F(x, y) | d((x, y), \delta B) < \frac{d_e}{8}, (x, y) \in B\}, \quad (17)$$

$$O = \{F(x, y) | d((x, y), \delta B) < \frac{d_e}{8}, (x, y) \notin B\}, \quad (18)$$

where \bar{X} means the average of X , F is the image intensity, d is the Euclidean distance, B is the set of nodule points, δB is the set of nodule border points, d_e is the effective diameter.

18 *Gergely Orbán, Gábor Horváth*

- (5) The contrast between the nodule parts inside and outside the viewable lung area (C_l).

$$C_l = \frac{\bar{I}_l}{\bar{I}_o}, \quad (19)$$

$$I_l = \{F(x, y) | (x, y) \in B, (x, y) \in L\}, \quad (20)$$

$$I_o = \{F(x, y) | (x, y) \in B, (x, y) \notin L\}, \quad (21)$$

where L is the set of points inside the viewable lung area. Otherwise, the notation is similar to the one used for the previous feature.

- (6) Standard deviation of a derivative of Gaussian filter output in different directions [Campadelli *et al.* (2006)]. The width parameter of the Gaussian function (σ_G) was 3 mm.
- (7) Mean of a Laplacian of Gaussian (LoG) filter for the center of the nodule ($\sigma_G = 6mm$) [Campadelli *et al.* (2006)].
- (8) LoG filter value at nodule center ($\sigma_G = 3mm$).
- (9) LoG filter value at nodule center ($\sigma_G = 6mm$).
- (10) LoG filter mean value inside nodule border ($\sigma_G = 1.5mm$).
- (11) LoG filter mean value inside nodule border ($\sigma_G = 3mm$).
- (12) LoG filter mean value inside nodule border ($\sigma_G = 6mm$).
- (13) LoG filter minimum value inside nodule border ($\sigma_G = 1.5mm$).
- (14) LoG filter minimum value inside nodule border ($\sigma_G = 6mm$).
- (15) LoG filter standard deviation inside nodule border ($\sigma_G = 3mm$).
- (16) LPCS distance coordinate described in section 3.4.
- (17) Difference of minimum values inside the nodule and in the surroundings of the nodule normalized by the intensity range in the lung field (D_m).

$$D_m = \frac{\min \{F(x, y) | (x, y) \in B\} - \min \{F(x, y) | (x, y) \in S\}}{\min \{F(x, y) | (x, y) \in L\} - \max \{F(x, y) | (x, y) \in L\}}, \quad (22)$$

$$S = \{(x, y) | d((x, y), \delta B) < \frac{d_e}{\sqrt{2}} - 1, (x, y) \notin B\}. \quad (23)$$

- (18) The rank of the candidate in a list sorted by enhancer output intensity.
- (19) Horizontal coordinate measured from the closest point of the lung field to the spine.
- (20) Output of the LNF enhancer using the nodule border as a kernel for the top-hat filter instead of the original round shaped one.
- (21) Vertical coordinate measured from the topmost point of the viewable lung area.

On the other hand, for the large nodules we observed the following features to be relevant.

- (1) The maximum value of the AFUM filter inside the nodule.
- (2) Average of contrast. Described by Haralick *et al.* [1973] as the 2nd feature (and referenced as H2).
- (3) Average of difference variance (H10).

- (4) Average of an information measure of correlation (H13).
- (5) Average of sum entropy (H8).
- (6) Joint entropy version one (H12-H13).
- (7) Joint entropy version two (H12-H13).
- (8) Mean of an LoG filter for the center of the nodule ($\sigma_G = 1.5mm$).
- (9) LPCS distance coordinate described in section 3.4.
- (10) Nodule enhanced image (LNF output) value.
- (11) Horizontal coordinate measured from the closest point of the lung field to the spine.
- (12) Difference of nodule and lung field intensity mean (D_a).

$$D_a = \overline{\{F(x, y) | (x, y) \in B\}} - \overline{\{F(x, y) | (x, y) \in L\}}. \quad (24)$$

- (13) Ratio of nodule and lung field mean intensity.

$$D_a = \frac{\overline{\{F(x, y) | (x, y) \in B\}}}{\overline{\{F(x, y) | (x, y) \in L\}}}. \quad (25)$$

- (14) A feature describing robustness of nodule border. It restarts nodule segmentation from centers near the original one and compares the resulting borders.
- (15) Similarity of histograms inside and outside nodule, near the border. Intensity values are sampled perpendicular to the nodule border, starting from several border points. For a given point, the number of bins of the inner histogram is counted where the difference between the corresponding bin-count of the outer histogram is greater than a threshold.

We also ran the algorithm for the OAF output, but we faced serious overfitting issues. Due to the fact that the OAF produces roughly one candidate per image resulting in a training set of approximately 300 images we could use only a very few sample points. As the number of sample points is in the same order of magnitude as of the features, while of the available combinations – determining the search space for the feature selection algorithm – is much higher, we experienced no generalization capability of the feature selection algorithm. For this reason, we rather used the relevant features of the LNF also for the OAF because the LNF finds more similar candidates to the OAF than the CSBF, and we hoped it includes some universally relevant features like coordinates and contrast.

Although the selection algorithm seemed to be sensitive to noise and we cannot guarantee that the selected features are truly the most relevant, some observations can be made. Both result sets contain global coordinates (CSBF: 16, 19, 21 and LNF: 9, 11) consistent with the fact that different lung segments contain a nodule with different probability. Interestingly, coordinates provide the highest classification accuracy when used on their own. For smaller nodules the AFUM and LoG filter outputs serve as important clues (CSBF: 1, 2, 7-15). These filters are tuned to enhance small and circular objects, which can be a reason for their utility. Rank and nodule enhanced image value (CSBF: 18, LNF: 10) are both related to the

nodule enhancer output thus helpful when deciding about a candidate. Some features describing nodule contrast (CSBF: 4, 5, 17, 20 and LNF: 2, 12, 13) turned out to be relevant probably because true lesions usually have lower contrast than false findings like vessels and remainders of rib shadows. The derivative of Gaussian filter enhances the edges of mid-sized nodules, making it useful for classification (CSBF 6). The features analyzing the nodule border and its surroundings (CSBF: 3 and LNF: 14, 15) measure the robustness of nodule segmentation, which is usually better for a true positive lesion. Information-theory-based and intensity-variance-related features for the LNF output (3-7) measure the homogeneity of the nodule. This can contribute to classification accuracy as false findings of LNF are usually dense but inhomogeneous areas of non-circular structures. The AFUM and LoG output for the LNF (1, 8) probably got selected due to noise as the scale of the filter is very different from the targeted nodule size.

4.2. *Effect of algorithm fusion*

To support our thesis that the combination of the three algorithms outperforms single algorithm solutions, we first show the performance of the three separate algorithms on their own compared to the hybrid system. The resulting four FROC curves can be seen in Figure 10.

Figure 10 here

We expected the poor performance of the LNF on its own. Despite finding most of the circular and large lesions, we measured a very low sensitivity as the radiographs contained mostly small nodules. Meanwhile, the OAF shows relatively high sensitivity at very low false positive rates but it is only capable of finding 48% of lesions in the candidate collection step, setting an upper bound for the classifier output. However, integrating their results with the CSBF improves sensitivity without adding many false positives. At constant 80% sensitivity the number of false positives can be reduced from 2.44 to 1.84 per image. Alternatively, with a false positive rate of 2.0 the sensitivity can be increased from 78.8% to 81.2%.

Basically, the LNF and OAF only complement the CSBF in finding larger lesions as the latter is originally parameterized to find nodules smaller than 30 mm. To make the CSBF versus overall system comparison fair, we measured the performance of a modified CSBF enable to find larger circular objects of up to 75 mm diameter. To do this, we increased the R_{max} and N parameter of the original algorithm. The results can be seen in Figure 11.

Figure 11 here

The results showed that the CSBF parameterized for a wider range of nodules is not just worse than the hybrid system but cannot even outperform the original CSBF. The reason behind this effect is the increased number of false findings if we make the algorithm sensitive to more structures. On the other hand, keeping the false positive number the same, we start losing smaller true positives as we include large areas.

Next, we modified our system to use only one classifier instead of the hierarchy of four SVM-s and compared the results in Figure 12. Here the input of the single classifier was the union of relevant features for the original four, plus we added an integer feature (ranging from one to three) identifying the source nodule enhancer.

Figure 12 here

The single classifier version showed worse results than the original, probably due to the effect of irrelevant features and an increase in dimensionality. Furthermore, the required run time for training increased by 10% compared to the cumulated training time of the four classifiers.

We also tested our algorithms on subsets of the PCSU database containing only specific types of lesions. Our aim was to check if the specific algorithms developed for particular lesions are truly more sensitive in those cases. For this measurement we used a more precise criterion for accepting a candidate as true positive. We accepted a candidate if it satisfied the 'centroid criterion' and for its area A_c , $0.5A_r < A_c < 2A_r$ where A_r is the area of the physician's marker. We plotted the results as FROC curves in Figure 13 for smaller nodules, larger nodules and infiltrated areas.

Figure 13 here

According to our expectations, the CSBF performed best for smaller nodules and only found a few smaller infiltrated areas, mostly circular ones. The 0% sensitivity for large nodules is not surprising because even if it enhanced some inner structures in a larger nodule, the diameter of the candidate was not large enough to be accepted as true positive. We can also see that the CSBF outperforms the other algorithms on small nodules. The same applies for the LNF but for large nodules. It shows the best performance for this type of lesions and it outperforms the others on this dataset. For the OAF we had slightly different experience. For the database containing only infiltrated areas it can only overcome the LNF for false positive rates less than 0.45; however, it performs better on large nodules under 0.07 false positives. This controversy may be caused by the fact that the OAF produces a very few false positives making it more effective in the regions with low false positive rate. Another reason why the LNF and the OAF do not produce many complementary results is the fact that infiltrated areas often contain dense and more-or-less circular regions while large nodules satisfy our model created for infiltrated areas. However, we should note that the OAF still performs better on infiltrated areas than large nodules relative to the LNF as the ratio of sensitivities is not more than 1.37 in the former and 1.76 in the latter case for large false positive ratios.

4.3. Results for the JSRT database

To produce comparable results with other systems we evaluated the algorithms on the publicly available JSRT database. For sensitivity calculation, we used all the 154 images containing a nodule. However, 14 of the cases contained a nodule in the opaque regions of the lung, omitted by our current algorithms. The first experiment, in Figure 14, is the same as in Figure 10 for the PCSU database.

Figure 14 here

Here the hybrid system cannot outperform the CSBF as the JSRT database contained only a few nodules larger than 30 mm; moreover these were all marked by the CSBF.

Considering possible working points for clinical applicability, the sensitivity values of 54.8%, 60.2% and 63% probably include lesions the examiner may have overlooked while producing 1.0, 2.0 and 3.0 false positives respectively.

We also analyzed if the same cases are difficult for the physicians and for the CADe system, hoping that our solution can find some of the cases the examiner probably overlooks. We can say, if the detections of a CADe system are complementary to the findings of radiologists or pulmonologists, it can better improve sensitivity when the human examiner and a CADe work in cooperation. For the experiment we used the labels expressing subtlety – proportional to the difficulty for a human reader – given for the JSRT radiographs. We plotted five FROC curves showing overall performance for nodules of different difficulties. The results can be seen in Figure 15.

Figure 15 here

We can see that the algorithm performs clearly better for easier nodules. It can almost always find obvious nodules at false positive rates of 0.3; however it can only detect 28% of extremely subtle nodules even with 2.7 false positives. The steepness of the curves shows that increasing the false positive rate above one increases sensitivity for more difficult cases – especially for the extremely subtle set – while detection of easier cases saturates. This can suggest a higher number of false positives if the goal is to produce complementary results to radiologists' findings.

We checked the system's generalization capability by training it on the PCSU and testing on the JSRT database. The resulting curves are plotted in Figure 16.

Figure 16 here

We see a large deterioration in performance, despite using cross-validation, when training on the JSRT. This effect is probably a result of improper image normalization as the radiographs in the two databases show very different characteristics regarding contrast and detail. It should be also noted that the PCSU database contained only a few very subtle nodules, while these are present in large numbers in the JSRT.

4.4. Comparison with state-of-the-art algorithms

In this section we try to compare our algorithm with other published solutions using the JSRT database. First we must state that proper comparison is almost impossible as a unified methodology does not exist for evaluation. For example most of the publications use a different criterion to determine a candidate as true positive resulting in a great variation in performance. To prove this, we evaluated our system using three different and commonly used criteria.

Our original criterion labels a candidate as positive if its centroid is inside a real

lesion's border. Another possible criterion inspects the overlap between real lesions and the candidate. If the overlap is larger than a predefined proportion, for example 50% of the smaller object, it takes it as a true positive finding. The last and most widespread criterion only compares the distance of candidate and lesion center. If they are closer than for example 25 mm, the candidate is classified as true positive. It is obvious that neither of these criteria is perfect. The 'centroid criterion' does not take into account the size of the candidate. For example a small candidate inside a large nodule gets a true positive label in spite of being undesirable. The '>50% overlap criterion' suffers from the same issue. Furthermore, a candidate marking the whole lung will be always classified as true positive if a nodule exists. The '25 mm criterion' is insensitive to nodule size, making it the most permissive of the three as it can accept false outputs for small structures especially under 15 mm. It should be also noted that the 'overlap criterion' can only be used in a system incorporating nodule segmentation. If a candidate border is not present, the center can be regarded as a centroid, so the distance or the centroid based criterion can be used. The FROC curves for our system using different evaluation methods are shown in Figure 17.

Figure 17 here

It can be seen that while the 'centroid' and '>50% overlap' criterion showed almost identical results, using the '25 mm criterion' resulted in a sensitivity increase by approximately 4 percentage points (pp) at false positive rates of 1-3. The similar performance of the first two criteria is due to the fact that they usually assign the same label. Different labeling can be obtained only if the candidate is much larger than the lesion, so it can easily contain the area marked by the physician but it is badly centered so the centroid will not lay inside the marked area. In these cases the '>50% criterion' is more permissive, but this is only a rare situation. On the other hand the permissiveness of the '25 mm criterion' often leads to false results. An example scenario is shown in Figure 18.

Figure 18 here

Another source of differences in performance analysis is the partly different images of the test set, even if all these test sets come from the JSRT database. Although Keserci and Yoshida [2002] published results on the JSRT database, they used only a subset for testing as they did not take advantage of cross-validation. Some other results differ in including or excluding nodules in opaque areas, as most of the algorithms search only within the visible lung area. This can affect sensitivity as sometimes opaque areas like the region under the shadow of the heart or diaphragm contain malignant nodules. Luckily, this can be corrected for the JSRT database by a multiplication of the sensitivity by $\frac{140}{154}$ when the test results exclude opaque cases, as this database contains 14 such cases.

Taking into consideration the difficulties mentioned, we collected some published results that are somewhat comparable. We listed some operating points in Table 1. We also plotted these operating points in Figure 19.

Table 1. Comparison of current published CADe systems on the JSRT database.

CADe system	Avg. no. of FP-s	Sensitivity	Criterion used
[Chen <i>et al.</i> (2011)]	1.0	54.5	centroid
	5.0	69.5	
[Snoeren <i>et al.</i> (2010)]	1.0	51.3	25 mm
[Hardie <i>et al.</i> (2008)]	2.0	57.8	25 mm
	4.0	71.0	
	4.3	71.4	
	5.0	72.8	
	5.4	75.6	
[Campadelli <i>et al.</i> (2006)] ^a	1.45	70	unspecified
	2.55	77	
	3.1	84	
[Schilham <i>et al.</i> (2006)]	2.0	51	>0% overlap
	4.0	67	
[Coppini <i>et al.</i> (2003)]	4.3	60	centroid
[Wei <i>et al.</i> (2002)]	5.4	80	unspecified

^asearching also in opaque regions

Figure 19 here

Looking at the results, we may note that our system compares differently depending on the number of false positives. At higher rates it has a lower sensitivity than the average; however, it can outperform most systems at false positive rates under three. The system published by Campadelli *et al.* [2006] is still better at this range, but their algorithm also searches for lesions in opaque areas and they did not specify the criterion used for evaluation, making it especially difficult to compare. Previous studies confirmed that showing good results at lower false positive rates is more important for clinical applicability, as the human reader cannot use CADe results as an aid if too many false findings are present [Matsumoto *et al.* (1992)].

The reason why our system shows better results at lower false positive rates is the parameterization of the CSBF algorithm. We set the threshold of the algorithm to produce on average 20-30 candidates per image, while other solutions usually use around 100. This way we can unburden the classifier's task and further reduce the number of false positives, but as a compromise, we limit the maximum sensitivity of the system to the sensitivity of the enhancers. With perfect classification results this step would be unnecessary, but current classifier performance justifies it. Although choosing the right operating point of the system is a difficult task, our experiences show that physicians get frustrated if they get more than three false findings per image. This is the reason why we tried to optimize the performance in the region of one to three false positives.

5. Conclusions and future work

To summarize our study, we developed a complex framework for lesion detection in chest radiographs. Three different algorithms, enhancing specific types of lesions indicating lung cancer, are included. The results are merged and the number of false positives is reduced with a hierarchy of SVM classifiers. Classifier input is determined by a feature selection algorithm.

We analyzed relevant features to get a better understanding of the classification procedure. We ran experiments on a diverse radiograph database to compare the lesion enhancer algorithms on their own to the complex system. We tried to judge the usage of more enhancers working in parallel, and more, independently trained SVM classifiers. We also tested how the system performs on different types of lesions and on lesions of different subtlety. For comparison with other published systems, we ran experiments using different candidate evaluation criteria.

We found that the proposed OAF filter is able to enhance amorphous infiltrated areas effectively. It produces very few false positives, while finding a fair amount of true lesions. However, it is not suitable for independent usage, as it cannot fulfill on its own the sensitivity requirements of a CADe system, furthermore, it cannot detect small, subtle nodules. Finally, we found a good use of it complementing our existing enhancers.

By using a hybrid system involving the LNF and OAF we managed to expand the scope of the CADe system towards larger lung nodules and infiltrated areas. These types of lesions were rarely targeted by previous CADe systems, despite being important findings at lung screening. Our tests showed that the LNF and OAF algorithms truly give complementary results to the CSBF, a modification of a widely used radial gradient based filter. Running measurements on subsets of different nodule subtlety showed that detection accuracy is much worse for nodules classified as extremely subtle; however, increasing the number of false positives to around three amends the result set, as the proportion of subtler findings increase.

The hybrid solution can also be used for the reduction of false positive findings. By combining the results of the three algorithms, we were able to reduce it, compared to a single CSBF solution, by 25% at constant 80% sensitivity on a diverse radiograph database. However, we could not measure an improvement on the JSRT database, due to the fact, that it contains only small and circular nodules.

Feature selection results showed that different descriptors are necessary for different types of lesions; however, the majority of the features measure analogous properties. Our experiences showed that for different sources using separate classifiers is worth consideration because irrelevant features affect SVM performance negatively.

Comparing our results with recently published, state-of-the-art systems showed that our solution performs well for low (≤ 3) false positive rates. It can outperform all – more-or-less comparable – systems when tuned to produce less than three false positives. Although some systems showed better results at higher false positive

rates, these regions are less important when considering real medical applicability (clinical application or application in lung screening centres). Furthermore, we are convinced that the more comprehensive set of findings of our scheme is a definite advantage compared to previous CADe systems.

Our measurements also reflected on the importance in using standard methods and criterions for accepting true positive findings. Not doing so may result in slight differences in sensitivity. For our system, we measured a difference of approximately 4pp when using two, commonly used criteria. Another obvious source of differences is the use of different databases. We saw approximately 20pp difference in sensitivity between the JSRT and PCSU image sets. Furthermore, using the parameters learned for one database showed worse performance on the other. This fact draws attention to generalization capability when using only one database and cross-validation. Also the representativity of the two databases should be investigated in everyday screening.

While our study hopefully contributes to the CADe community and helps to step towards an efficient system accepted by physicians, a large ground for improvement still exists. Our results showed that decomposing the detection task by lesion types is an effective principle. By generalizing our concepts in the future, separating small and early stage tumors into subtypes can be a ground for future investigations, and may help to improve the detection of very small and subtle nodules.

Another generalization of the current scheme may be to involve enhancers sensitive to the same type of nodules. Using a more sophisticated merging method, the confidence of the result set may be strengthened. Of course when involving more algorithms, care should be taken to keep running times low enough for clinical applicability.

Comparison with other solutions showed that our system is sometimes inferior because omitting results in opaque lung areas. With proper modification of the filters or a robust background compensation, enabling detection in opaque areas can lead to a huge improvement.

Our future goal is the development of a complex medical decision support system where besides X-ray images all other available and relevant knowledge about a patient (e.g. anamnesis) is to be taken into consideration. This extra information can help to make the CADe response more accurate.

Regarding medical applicability, a possible working point may be 60% sensitivity with an average of two false positives measured on the JSRT database. It means 80% sensitivity for radiographs like in the PCSU database. This output may help to improve the sensitivity of lung screening without exposing the examiner to much extra work. For confirmation an observer study is needed in the future. The utility of the OAF in real-world should be also confirmed. We experienced that the subjective judgment of the system got better after we integrated the OAF, but an exact measurement should be carried out.

Acknowledgments

The authors express their thanks to all physicians of Pulmonological Clinic, Semmelweis University, who helped us with validated images and to the development staff of ImSoft ltd. This work was partly supported by the National Development Agency under contract KMOP-1.1.1-07/1-2008-0035.

References

- Acharyya, M., Chakravarty, S. and Stoeckel, J. (2008). Use of random process-based fractal measure for characterization nodules and suspicious regions in lung. *Proc. of SPIE*, **6915**: 69151F.
- Campadelli, P., Casiraghi, E. and Artioli D. (2006). A fully automated method for lung nodule detection from postero-anterior chest radiographs. *IEEE T. Med. Imaging*, **25** (12): 1588–1603.
- Carreira, M.J., *et al.* (1998). Computer-aided diagnoses: Automatic detection of lung nodules. *Med. Phys.*, **25**.
- Chakraborty, D.P. and Berbaum, K.S. (2004). Observer studies involving detection and localization: modeling, analysis, and validation. *Med. Phys.*, **31**: 2313.
- Chen, S., Suzuki, K. and MacMahon, H. (2011). Development and evaluation of a computer-aided diagnostic scheme for lung nodule detection in chest radiographs by means of two-stage nodule enhancement with support vector classification. *Med. Phys.*, **38**: 1844.
- Coppini, G., *et al.* (2003). Neural networks for computer-aided diagnosis: detection of lung nodules in chest radiograms. *IEEE T. Inf. Technol. B.*, **7** (4): 344–357.
- Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Mach. learn.*, **20** (3): 273–297.
- De Boo, D.W., *et al.* (2009). Computer-aided detection (CAD) of lung nodules and small tumours on chest radiographs. *Eur. J. Radiol.*, **72** (2): 218–225.
- Doi, K. (2007). Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Comput. Med. Imag. Grap.*, **31** (4-5): 198–211.
- Giger, M.L., *et al.* (1990). Pulmonary nodules: computer-aided detection in digital chest images. *Radiographics*, **10** (1): 41.
- van Ginneken, B., *et al.* (2002). Automatic detection of abnormalities in chest radiographs using local texture analysis. *IEEE T. Med. Imaging*, **21** (2): 139–149.
- van Ginneken, B., Hogeweg, L. and Prokop, M. (2009). Computer-aided diagnosis in chest radiography: Beyond nodules. *Eur. J. Radiol.*, **72** (2): 226–230.
- Haralick, R.M., Shanmugam, K. and Dinstein, I.H. (1973). Textural features for image classification. *IEEE T. Syst. Man. Cyb.*, **3** (6): 610–621.
- Hardie R. C., *et al.* (2008). Performance analysis of a new computer aided detection system for identifying lung nodules on chest radiographs. *Med. Image Anal.*, **12** (3): 240–258.
- Haykin, S. (1994). *Neural networks: a comprehensive foundation*. Prentice Hall, Upper Saddle River, NJ.
- Heath, M.D. and Bowyer, K.W. (2000). Mass detection by relative image intensity. *Proc. of the 5th International Workshop on Digital Mammography (IWDM-2000)*, 219–225.
- de Hoop, B., *et al.* (2010). Computer-aided detection of lung cancer on chest radiographs: effect on observer performance. *Radiology*, **257** (2): 532.
- Juhász, S., *et al.* (2010). Segmentation of Anatomical Structures on Chest Radiographs. *IFMBE Proc. of MEDICON 2010*, **29**: 359–362.
- Katsuragawa, S. and Doi, K. (2007). Computer-aided diagnosis in chest radiography. *Comput. Med. Imag. Grap.*, **31** (4-5): 212–223.

- Keserci, B. and Yoshida, H. (2002). Computerized detection of pulmonary nodules in chest radiographs based on morphological features and wavelet snake model. *Med. Image Anal.*, **6** (4): 431–447.
- Kobayashi, T., *et al.* (1996). Effect of a computer-aided diagnosis scheme on radiologists' performance in detection of lung nodules on radiographs. *Radiology*, **199**: 843–848.
- Le, K. (2011). Chest X-Ray Analysis for Computer-Aided Diagnostic. *Advanced Computing CCSIT 2011*, 300–309.
- Lee, J.S. (1980). Digital image enhancement and noise filtering by use of local statistics. *IEEE T. Pattern. Anal.*, **2**: 165–168.
- Li, Q., Katsuragawa, S. and Doi, K. (2001). Computer-aided diagnostic scheme for lung nodule detection in digital chest radiographs by use of a multiple-template matching technique. *Med. Phys.*, **28**: 2070.
- Li, F., *et al.* (2010). True Detection Versus "Accidental" Detection of Small Lung Cancer by a Computer-Aided Detection (CAD) Program on Chest Radiographs. *J. Digit. Imaging*, **23** (1): 66–72.
- Lo, S.C.B., *et al.* (1995). Artificial convolution neural network techniques and applications for lung nodule detection. *IEEE T. Med. Imaging*, **14** (4): 711–718.
- MacMahon, H., *et al.* (1990). Computer-aided diagnosis in chest radiology. *J. Thorac. Imag.*, **5** (1): 67.
- MacMahon, H., *et al.* (1999). Computer-aided diagnosis of pulmonary nodules: Results of a large-scale observer test. *Radiology*, **213**: 723–726.
- Mao, F., *et al.* (1998). Fragmentary window filtering for multiscale lung nodule detection: Preliminary study. *Acad. Radiol.*, **5** (4): 306–311.
- Matsumoto, T., *et al.* (1992). Potential usefulness of computerized nodule detection in screening programs for lung cancer. *Invest. Radiol.*, **27** (6): 471.
- Monnier-Cholley, L., *et al.* (1998). Computer-aided diagnosis for detection of interstitial opacities on chest radiographs. *Am. J. Roentgenol.*, **171** (6): 1651.
- Mountain, C.F. (1997). Revisions in the international system for staging lung cancer. *Chest*, **111** (6): 1710.
- Muhm, J.R., *et al.* (1983). Lung cancer detected during a screening program using four-month chest radiographs. *Radiology*, **148** (3): 609.
- Penedo, M.G., *et al.* (1998). Computer-aided diagnosis: a neural-network-based approach to lung nodule detection. *IEEE T. Med. Imaging*, **17** (6): 872–880.
- Pereira, C.S., *et al.* (2007a). Detection of lung nodule candidates in chest radiographs. *Pattern Recognition and Image Analysis*, 170–177.
- Pereira, C.S., *et al.* (2007b). Evaluation of contrast enhancement filters for lung nodule detection. *ICIAR 2007 Proc.*, 878–888.
- Quekel, L.G.B.A., *et al.* (1999). Miss Rate of Lung Cancer on the Chest Radiograph in Clinical Practice*. *Chest*, **115** (3): 720.
- Schilham, A.M.R., Van Ginneken, B. and Loog, M. (2006). A computer-aided diagnosis system for detection of lung nodules in chest radiographs with an evaluation on a public database. *Med. Image Anal.*, **10** (2): 247–258.
- Shah, P.K., *et al.* (2003). Missed Non-Small Cell Lung Cancer: Radiographic Findings of Potentially Resectable Lesions Evident Only in Retrospect1. *Radiology*, **226** (1): 235.
- Shiraishi, J., *et al.* (2000). Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules. *Am. J. Roentgenol.*, **174** (1): 71.
- Shiraishi, J., *et al.* (2006). Computer-aided diagnostic scheme for the detection of lung nodules on chest radiographs: localized search method based on anatomical classification. *Med. Phys.*, **33**: 2642.

- Shiraishi, J., Li, F. and Doi, K. (2007). Computer-aided diagnosis for improved detection of lung nodules by use of PA and lateral chest radiographs. *Acad. Radiol.*, **14** (1): 28.
- Snoeren, P.R., *et al.* (2010). Training a Computer Aided Detection System with Simulated Lung Nodules in Chest Radiographs. *Proc. of The Third International Workshop on Pulmonary Image Analysis*, 139.
- Sundararajan, R., *et al.* (2010). A multiresolution support vector machine based algorithm for pneumoconiosis detection from chest radiographs. *Biomedical Imaging: From Nano to Macro, 2010 IEEE International Symposium on*, 1317–1320.
- Suzuki, K., *et al.* (2005). False-positive reduction in computer-aided diagnostic scheme for detecting nodules in chest radiographs by means of massive training artificial neural network. *Acad. Radiol.*, **12** (2): 191–201.
- Wei, J., *et al.* (2002). Optimal image feature set for detecting lung nodules on chest X-ray images. *Proc. of CARS2002*, 706–711.
- Weng, Q., *et al.* (2009). Computer-aided Diagnosis: A Support-Vector-Machine-Based Approach of Automatic Pulmonary Nodule Detection in Chest Radiographs. *Proc. of the 2009 International Symposium on Bioelectronics and Bioinformatics*, 60.
- Xu, X.W., *et al.* (1997). Development of an improved CAD scheme for automated detection of lung nodules in digital chest images. *Med. Phys.*, **24**: 1395.
- Yoshida, H. (2004). Local contralateral subtraction based on bilateral symmetry of lung for reduction of false positives in computerized detection of pulmonary nodules. *IEEE T. Bio-med. Eng.*, **51** (5): 778–789.
- Yoshida, H., *et al.* (1995). Computer-aided diagnosis scheme for detecting pulmonary nodules using wavelet transform. *Proc. of SPIE*, **2434**: 621.

Biography

Gergely ORBÁN, Email: orbanger@mit.bme.hu

He took his MSc degree in 2009 in technical informatics at the Budapest University of Technology and Economics (BUTE). Started his PhD studies in 2009 at BUTE, PhD school of computer sciences. His research topic is application of complex methods for image based diagnostics. He takes part in a project aiming at the development of a computer aided detection system for signs of lung cancer on chest radiographs. His main interests are image processing and machine learning.

Figure 20 here

Gábor HORVÁTH, Email: horvath@mit.bme.hu (corresponding author)

Gábor Horváth received Dipl. Eng. Degree (EE) from the Budapest University of Technology and Economics in 1970 and the candidate of science (PhD) degree from the Hungarian National Academy of Sciences in digital signal processing in 1988.

Since 1970 he has been with the Department of Measurement and Information Systems Budapest University of Technology and Economics where he has held various teaching positions. In the last few years till the latest days he has served as the Head of the Department. He is a senior member of IEEE. He has published more than 150 papers and he is the author or coauthor of 8 books. His research and educational activity is related to the theory and application of neural networks machine learning, and medical image processing.

Figure 21 here

Figures

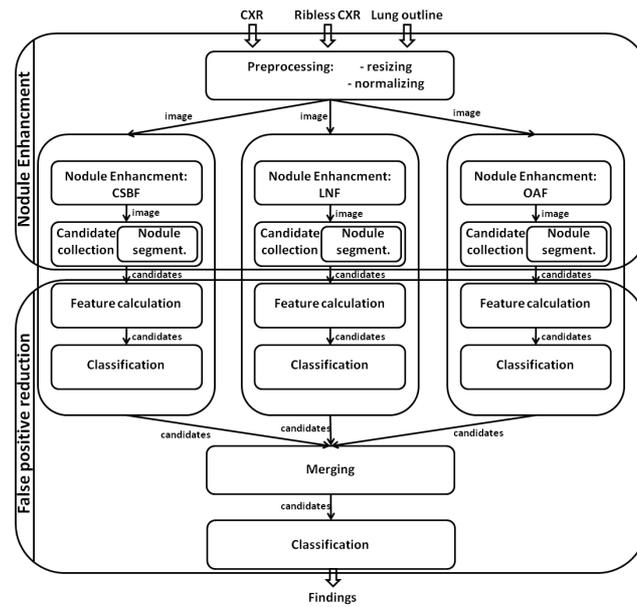


Fig. 1. Overview of the lesion finder subsystem of the CADe scheme (second and third steps).

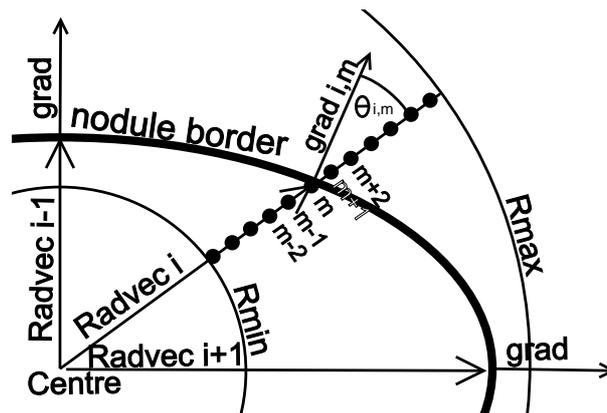


Fig. 2. An illustration of the CSBF filter.

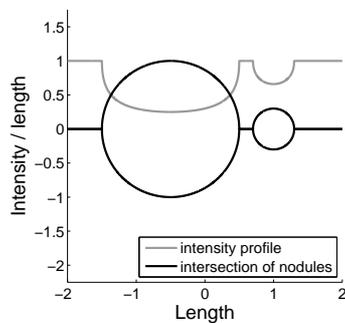


Fig. 3. An estimated intensity profile of a large, circular nodule (of unit radius) and a smaller one ($r = 0.3$ unit). Note that the profile of the larger nodule is more rectangular.

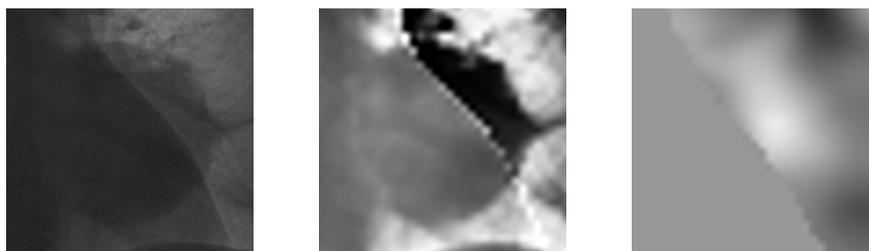


Fig. 4. A large nodule partly overlapped by the heart (left), the LCE output (middle) and the final LNF output (right).

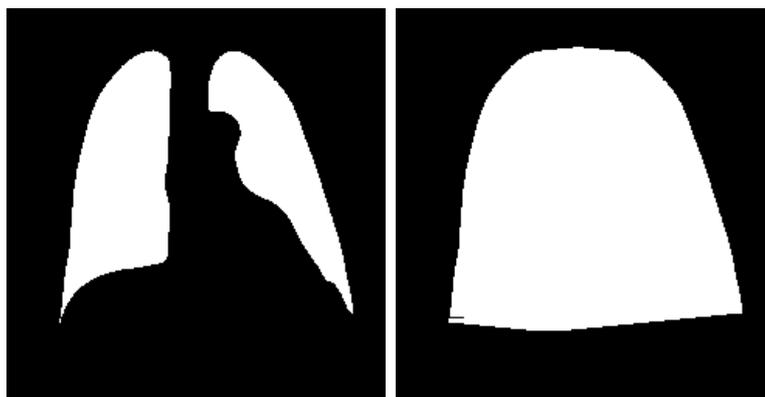


Fig. 5. The masks of the viewable lung area (left) and the estimated area of the whole lung (right).

32 *Gergely Orbán, Gábor Horváth*



Fig. 6. An example for a typical large infiltrated area. Hard-to-define border and non-circular shape is common for these structures.

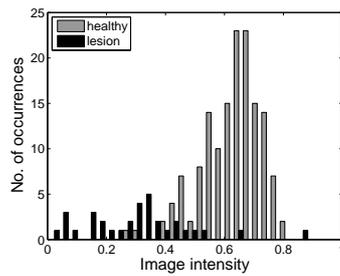


Fig. 7. Histogram of intensities for a fixed location (coordinates are $p = 0.8, \varphi = 0.6\pi$ in the LPCS system described later in section 3.4). Samples come from the PCSU database and are separated according to whether the location is inside a marked area of a lesion or in a completely healthy area.

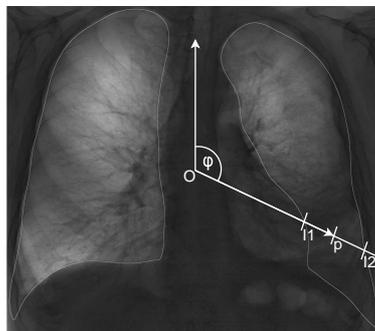


Fig. 8. Illustration of LPCS calculation for a given point p .



Fig. 9. An input image (left), the outlying pixels in LPCS space (center) and output markers of the OAF (right).

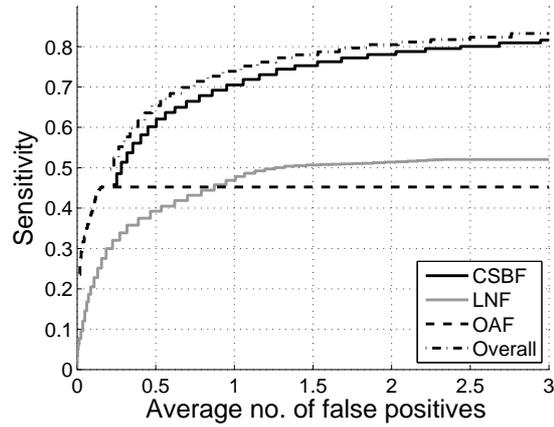


Fig. 10. FROC curves on the PCSU database. Three curves indicate the performance of systems using a single enhancer only and one curve shows the result of algorithm fusion.

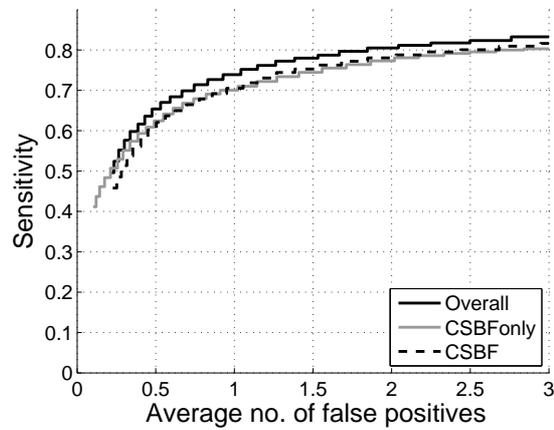


Fig. 11. An FROC curve for comparing algorithm fusion versus a CSBF sensitive to small and large nodules.

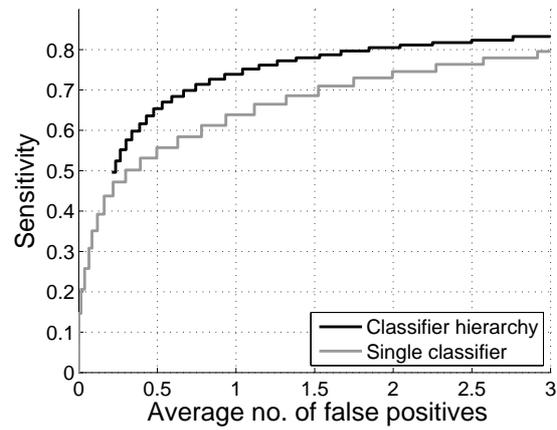


Fig. 12. Comparison of the classifier hierarchy to a system using only one SVM.

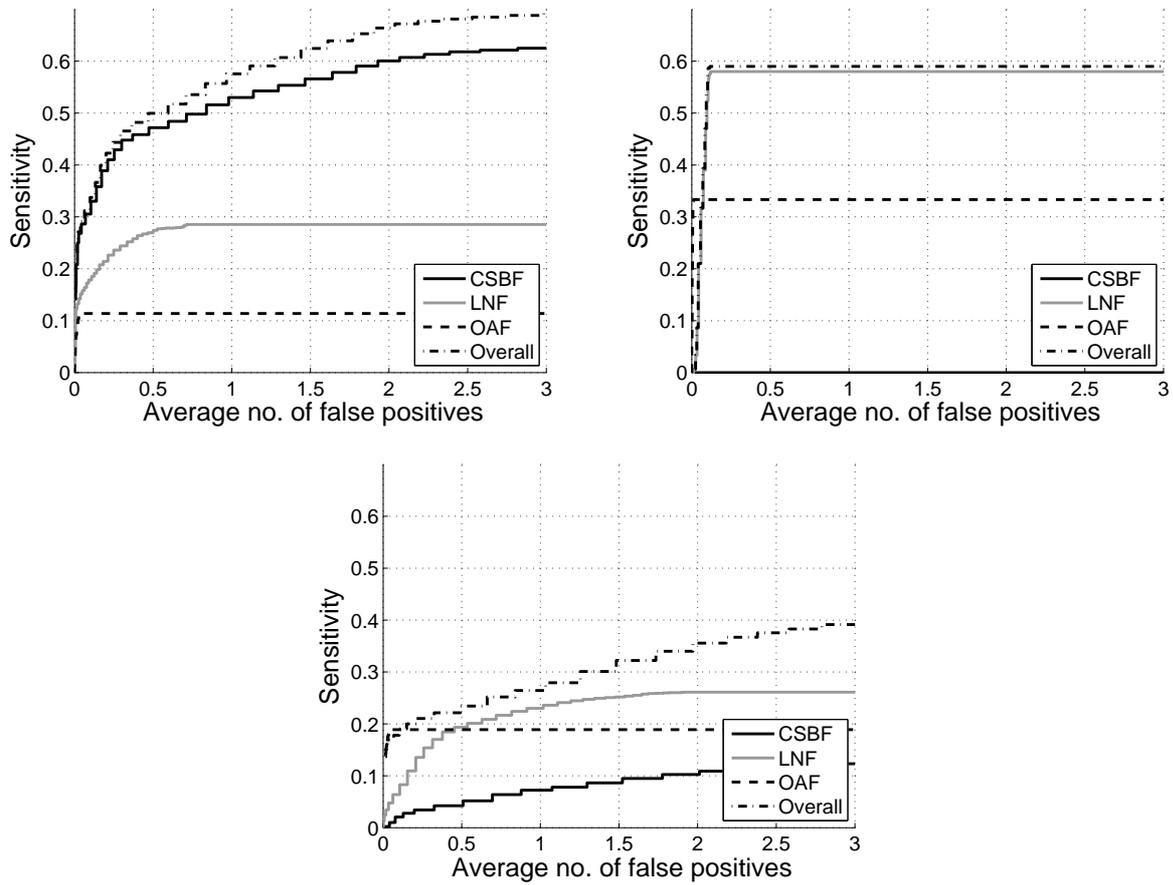


Fig. 13. Characteristics of the algorithms for different kinds of lesions. FROC curve for small (≤ 30 mm) nodules (top left), large (> 30 mm) nodules (top right) and infiltrated areas (bottom).

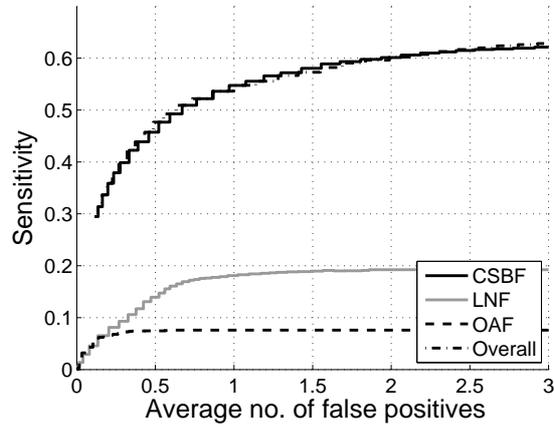


Fig. 14. FROC curve on the JSRT database. Three curves indicate the performance of systems using a single enhancer only and one curve shows the result of algorithm fusion.

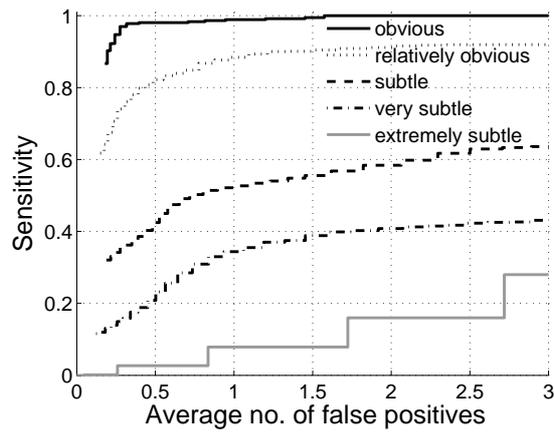


Fig. 15. Performance of the overall system for nodules of different subtlety on the JSRT database.

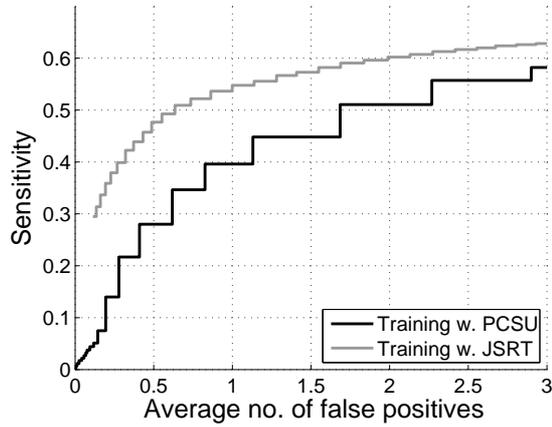


Fig. 16. FROC curve comparing performance on the JSRT database when training on the PCSU database versus training on the JSRT database using cross validation.

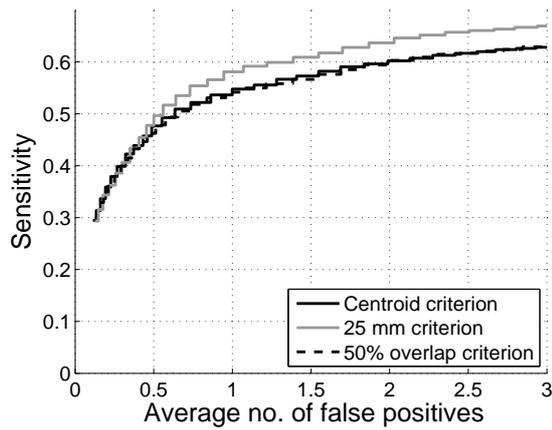


Fig. 17. FROC curve on the JSRT database with same settings except using different criteria.

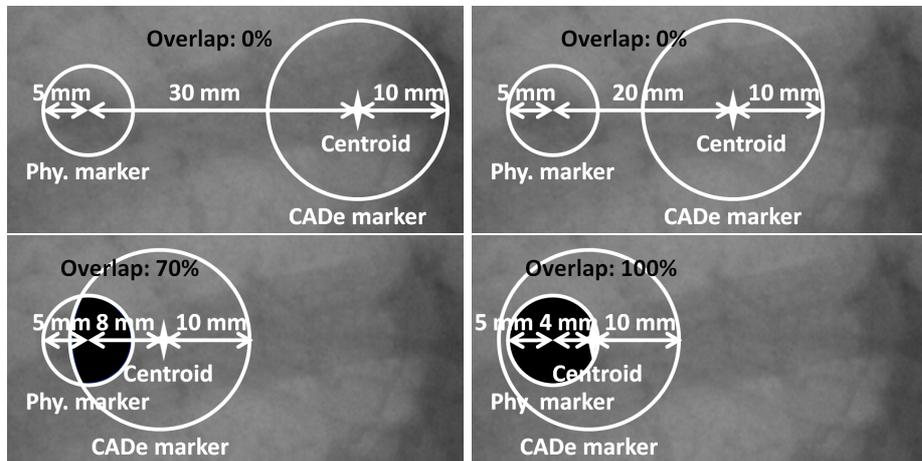


Fig. 18. An example scenario when the three criteria give inconsistent results. When the distance between the reference marker and the CADe marker is 30 mm, all the criteria give a false positive label. If the distance is 20 mm, the '25 mm criterion' gives true, the others false. At 8 mm distance the '>50% overlap criterion' also gives true, while the 'centroid criterion' holds the false label. At 4 mm all the criteria gives a true positive label.

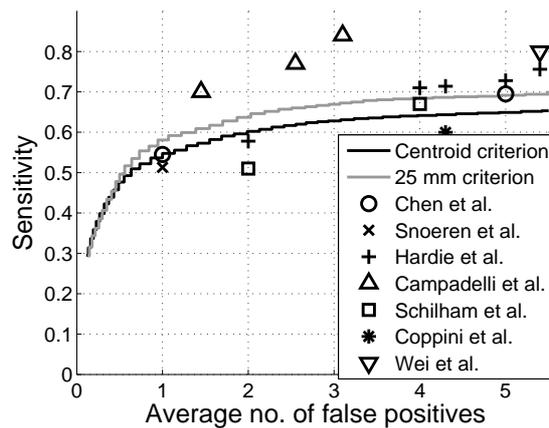


Fig. 19. FROC curve on the JSRT database (154 images) using the 'centroid' and '25 mm criterion' compared to other published results.

40 *Gergely Orbán, Gábor Horváth*



Fig. 20. Gergely ORBÁN



Fig. 21. Gábor HORVÁTH