



Deep learning-based pulmonary nodule detection: Effect of slab thickness in maximum intensity projections at the nodule candidate detection stage

Sunyi Zheng^{a,*}, Xiaonan Cui^{b,c}, Marleen Vonder^d, Raymond N.J. Veldhuis^e, Zhaoxiang Ye^c, Rozemarijn Vliegthart^b, Matthijs Oudkerk^f, Peter M.A. van Ooijen^a

^a Department of Radiation Oncology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands

^b Department of Radiology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands

^c Department of Radiology, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Centre of Cancer, Tianjin, China

^d Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands

^e University of Twente, Enschede, the Netherlands

^f University of Groningen, Groningen, the Netherlands

ARTICLE INFO

Article history:

Received 10 January 2020

Accepted 14 June 2020

Keywords:

Diagnosis

Computer-assisted

Artificial intelligence

Tomography

X-ray computed

Maximum intensity projection

Pulmonary Nodules

ABSTRACT

Background and Objective: To investigate the effect of the slab thickness in maximum intensity projections (MIPs) on the candidate detection performance of a deep learning-based computer-aided detection (DL-CAD) system for pulmonary nodule detection in CT scans.

Methods: The public LUNA16 dataset includes 888 CT scans with 1186 nodules annotated by four radiologists. From those scans, MIP images were reconstructed with slab thicknesses of 5 to 50 mm (at 5 mm intervals) and 3 to 13 mm (at 2 mm intervals). The architecture in the nodule candidate detection part of the DL-CAD system was trained separately using MIP images with various slab thicknesses. Based on ten-fold cross-validation, the sensitivity and the F_2 score were determined to evaluate the performance of using each slab thickness at the nodule candidate detection stage. The free-response receiver operating characteristic (FROC) curve was used to assess the performance of the whole DL-CAD system that took the results combined from 16 MIP slab thickness settings.

Results: At the nodule candidate detection stage, the combination of results from 16 MIP slab thickness settings showed a high sensitivity of 98.0% with 46 false positives (FPs) per scan. Regarding a single MIP slab thickness of 10 mm, the highest sensitivity of 90.0% with 8 FPs/scan was reached before false positive reduction. The sensitivity increased (82.8% to 90.0%) for slab thickness of 1 to 10 mm and decreased (88.7% to 76.6%) for slab thickness of 15–50 mm. The number of FPs was decreasing with increasing slab thickness, but was stable at 5 FPs/scan at a slab thickness of 30 mm or more. After false positive reduction, the DL-CAD system, utilizing 16 MIP slab thickness settings, had the sensitivity of 94.4% with 1 FP/scan.

Conclusions: The utilization of multi-MIP images could improve the performance at the nodule candidate detection stage, even for the whole DL-CAD system. For a single slab thickness of 10 mm, the highest sensitivity for pulmonary nodule detection was reached at the nodule candidate detection stage, similar to the slab thickness usually applied by radiologists.

© 2020 The Author(s). Published by Elsevier B.V.

This is an open access article under the CC BY license. (<http://creativecommons.org/licenses/by/4.0/>)

1. Introduction

Lung cancer is one of the deadliest cancers (18.4% of the total cancer deaths in 2018) worldwide with a low long-term survival rate [1–4]. Accurate lung nodule detection based on low-dose CT is of great importance to diagnose and treat lung cancer at an early

Abbreviations: DL, Deep learning; CAD, Computer-aided detection; FPs, False positives; MIP, Maximum intensity projection; LIDC, Lung image database consortium; IDRI, Image database resource initiative.

* Corresponding author.

E-mail address: s.zheng@umcg.nl (S. Zheng).

stage [5,6]. Clinical research has demonstrated that the maximum intensity projection (MIP) technique is an effective method for radiologists to detect nodules on CT images [7-9]. With the implementation of lung cancer screening all over the world, the use of a computer-aided detection (CAD) system could be essential to reduce the fast-increasing workload of radiologists.

Recently developed CAD systems are mainly based on deep learning algorithms. The hierarchical learning architecture of deep learning algorithms is inspired by artificial intelligence emulating the deep, layered learning process of the primary sensorial areas of the neocortex in the human brain. These algorithms are able to extract features automatically from the underlying data [10]. These features include information (shape, size, intensity) that is also used by human readers. In recent years, a large number of deep learning-based CAD systems (DL-CAD) have been developed in the medical image analysis field [11], especially for the purpose of lung nodule detection [12]. However, DL-CAD systems still have not been widely used in clinical practice for various reasons, including low sensitivity or high false positive rates of the available systems [13]. It is important to improve the performance of current DL-CAD systems to provide a more trustworthy assistance for radiologists.

The MIP technique can boost the performance for nodule detection as shown by the 2-dimensional proprietary DL-CAD system [14]. The slab thickness plays an important role in this technique, since it can significantly influence how clear a nodule can be distinguished from pulmonary bronchi and surrounding vasculature. In other words, detection on the MIP images with different slab thicknesses directly influences the performance at the nodule candidate detection stage. This stage determines the upper limit performance of the whole system [15], which is essential for the development of the nodule detection system. In a recent study, it was found that the optimal slab thickness for radiologists' detection of lung nodules is 10 mm [16]. However, a slab thickness of 10 mm might not be the optimal thickness for a 2-dimensional DL-CAD system, since radiologists differentiate nodules by viewing continuous slices, while the 2-dimensional DL-CAD system detects nodules based on a single slice. Therefore, the aim of this study was to explore the effect of MIP slab thickness on the performance of the DL-CAD system at the nodule candidate detection stage and to find the optimal MIP slab thickness with which the DL-CAD system can detect more nodules among nodule candidates and provide good results for the false positive reduction stage.

2. Materials and methods

2.1. Study population and CT image data

The purpose of the Lung Image Database Consortium and Image Database Resource Initiative (LIDC/IDRI) [17] was to establish a publicly available reference for the medical imaging research community and to stimulate the development of CAD systems. The patient inclusion criteria were specified in a reference work [18]. The dataset contained 1018 helical thoracic CT scans from 1010 patients. With appropriate local IRB approval, the scans were retrospectively collected from seven academic medical centers in the United States. All protected health information was removed by the anonymization software.

The CT scans were acquired by CT systems from different vendors (General Electric, Philips, Siemens, Toshiba) with different reconstruction parameters. The details of the data are shown in Tables 1 and 2. In the selection of study data, scans with a slice thickness of 3 mm or more were excluded because of the deviation from isotropic voxel size [19]. Consequently, 888 CT scans were kept in the study.

Table 1

Distribution of the reconstruction kernels and the number of scans from different vendors.

Vendor	Kernel Type	Number of CT scans	Total number of scans included per vendor
GE	Enhancing	220	662
	Over-enhancing	70	
	Standard	372	
PHILIPS	Standard	21	73
	Enhancing	7	
SIEMENS	Over-enhancing	45	148
	B20s Soft	1	
	B30f Standard	102	
	B31f Standard	1	
	B45f Enhancing	30	
	B50f Enhancing	2	
	B70f	12	
TOSHIBA	Over-enhancing		5
	FC03 Standard	2	
	FC10 Soft	3	
Total	-	888	888

Table 2

Distribution of the reconstructed slice thickness in all scans.

Slice thickness (mm)	Number of scans
0.6	7
0.75	30
0.9	2
1	58
1.25	343
1.5	5
2	123
2.5	320
Total	888

The section thickness of scans in the dataset ranged from 0.6 mm to 2.5 mm. The study from Kim et al. [20] showed that radiologists had a good detection rate based on the scans with 1 mm section slices. Hence, each scan was rescaled to a stack of 1 mm axial section slices by linear interpolation. To explore the general effect of the slab thickness on nodule detection, 1 mm slices were used to generate MIP images with slab thicknesses of 5, 10, 15, 20, 25, 30, 35, 40, 45 and 50 mm with an increment of 1 mm. To further determine the slab thickness that could have a higher sensitivity, MIP images with a slab thickness of 3, 5, 7, 9, 11, 13 mm with an increment of 2 mm were created.

2.2. Image annotation and nodule selection

In this study, we used publicly available data of the LIDC/IDRI dataset of which the radiological evaluation was described in [17]. In short, four experienced radiologists assessed the scans in two phases. First, all radiologists independently annotated all scans, recording information of pulmonary nodule location, diameter, and texture scores in the information sheet. In the second phase, every radiologist reviewed their labeled scans with the anonymized results from other radiologists. The findings included non-nodules, nodules ≤ 3 mm in diameter, and nodules ≥ 3 mm in diameter. The current study only focused on nodules with a diameter ≥ 3 mm. All nodules detected by the majority of radiologists were used as the reference standard. Non-nodules, nodules < 3 mm, and nodules detected by the minority of radiologists were considered as irrelevant findings. After lung nodule selection, 1186 valid nodules were included in this study.

2.3. Deep learning-based CAD system

The DL-CAD system has two stages, namely, nodule candidate detection and false positive reduction. At the first stage, consisting of four streams with 2D convolutional neural networks trained by MIP images with 4 slab thicknesses separately, the system determines locations of potential nodule candidates. At the second stage, false positive reduction, each potential candidate is given a probability of being a nodule by the classifier. The architecture was validated previously for 4 slab thickness settings of 1, 5, 10, 15 mm. The architecture of the system has been described in detail and showed a good performance on a large variation dataset [14].

In the current study, we mainly focus on the nodule candidate detection part of the DL-CAD system. More specifically, the lung parenchyma was segmented out of the whole image to narrow the region of interests for training the DL-CAD system. After segmentation, MIP images with different slab thicknesses were generated. Then the same 2D convolutional neural networks were trained separately using MIP images with 16 slab thicknesses in 16 streams for the detection of lung nodule candidates. After training, the system could mark potential nodule candidates in their appearing MIP images with coordinates. At the false positive reduction stage, two 3D convolutional neural networks described in the previous study [14] with the cube size of 16 and 32 were retrained to remove the false positives. The probability of being nodules for each candidate was averaged by the outputs of these two networks.

2.4. Evaluation

The nodule candidate detection performance of the DL-CAD program using MIP images with varying slab thicknesses was evaluated on the 888 scans by ten-fold cross-validation. Nodules were classified into three groups based on diameter: <5 mm (270 nodules), 5–10 mm (635 nodules), and >10 mm (281 nodules). A good performance of the nodule candidate detection with an optimal MIP slab thickness should have a high sensitivity with a low false positive rate. The high sensitivity determines the ability of lung nodule detection, while the number of false positives is related to extra efforts for the DL-CAD system or radiologists in further diagnosis. Achieving a high sensitivity is more important than reducing the number of false positives at nodule candidate detection. Thus, the performance was assessed by F_2 measure which gives more weight to sensitivity. The F_2 score is equal to 5 times the product of recall and precision, divided by the sum of 4 times precision and recall [21]. To compare with other methods, the F_1 scores were also reported when applying different slab thickness settings [22]. The McNemar's test [23] was applied to determine the difference in sensitivity between two MIP slab thickness settings which have the highest sensitivity or the largest F_2 score, using IBM SPSS Statistics (version 22).

Results from different MIP slab thickness settings were merged to explore whether the sensitivity of the nodule candidate detection could be improved. False negatives were recorded that were still missed in the results from the optimal MIP slab thickness or the combined results from all slab thickness settings. These undetected nodules were analyzed by using the nodule information sheet that was previously filled in by the four radiologists. The nodule density types (solid, part-solid, non-solid) were defined as types for which the majority of the radiologists gave texture scores. Nodule information including component type and diameter was summarized.

After false positive reduction for the results combined from all MIP slab thickness settings, the Competition Performance Metric (CPM) was used to evaluate the performance of the DL-CAD system [24]. This metric calculates the average sensitivity at seven false

positive rates (1/8, 1/4, 1/2, 1, 2, 4, and 8 FPs/scan) in the free-response receiver operating characteristic (FROC) curve [25].

3. Results

The performance of the system at the nodule candidate detection stage was evaluated when it used 1 mm axial slices or MIP images with varying slab thicknesses for nodule detection (Table 3). The sensitivity first went up from 82.8% to 90.0% with increasing slab thickness from 1 to 10 mm and then gradually decreased to 76.6% with slab thicknesses from 15 to 50 mm. At this stage, it had the highest sensitivity for the detection of nodules regardless of size or nodules > 10 mm at a MIP slab thickness of 10 mm. The number of false positives dropped with increasing MIP slab thickness, but is more or less stable with MIP slab thicknesses of 30 mm and higher. Although it showed the highest F_2 score at a MIP slab thickness of 25 mm at the same stage, the sensitivity at a MIP slab thickness of 10 mm is significantly higher than that of 25 mm MIP images (90.0% versus 87.9%, $p = 0.022$).

To further analyze the possible slab thickness with a higher sensitivity, MIP images were reconstructed with slab thicknesses of 3 to 13 mm at 2 mm intervals (Table 4). The program again showed a sensitivity of 90.0% with 9 mm MIP slab thickness images, which is close to the 10 mm MIP images, but more false positives were found with 9 mm MIP images compared to those of 10 mm.

Some examples of MIP images in the same slice with different slab thicknesses are shown in Fig. 1. The nodule in the 1 mm axial section slice is indicated with a blue arrow. With increasing MIP slab thickness, the nodule is easier to distinguish from the vessels

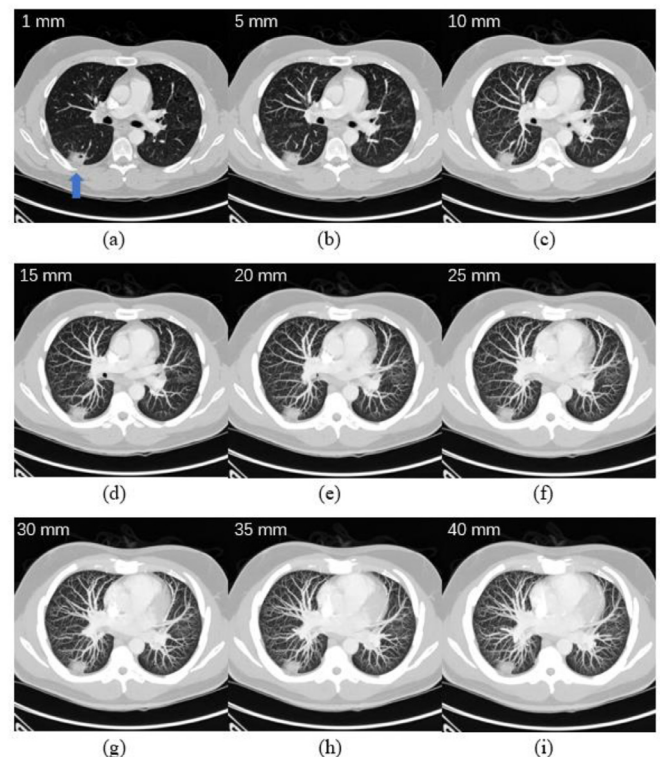


Fig. 1. Examples of the 1 mm axial section slice and MIP images in various slab thicknesses. From (a) to (i), the slab thickness is 1, 5, 10, 15, 20, 25, 30, 35 and 40 mm, respectively. One nodule is indicated with a blue arrow in the right lower lobe lung. With increase of the slab thickness, the nodule stands more out, whereas vessels are more continues. The nodule still can be seen after (f), although more vessels are projected in this slice. This does not add more value for detection but causes a more crowded image.

Table 3

The performance of the nodule candidate detection per MIP slab thickness at 5 mm intervals.

Slab thickness	Number of nodules detected in different sizes(sensitivity)			Sensitivity	Number of FPs	F_1	F_2
	<5 mm $n = 270$	5–10 mm $n = 635$	>10 mm $n = 281$				
1 mm	177 (65.6%)	542 (85.4%)	263 (93.6%)	82.8%	12,940	0.130	0.263
5 mm	199 (73.7%)	575 (90.6%)	268 (95.4%)	87.9%	9792	0.173	0.334
10 mm	222 (82.2%)	579 (91.2%)	266 (94.7%)	90.0%	6895	0.233	0.420
15 mm	212 (78.5%)	575 (90.6%)	265 (94.3%)	88.7%	5602	0.268	0.461
20 mm	215 (79.6%)	571 (89.9%)	262 (93.2%)	88.4%	5101	0.286	0.481
25 mm	204 (75.6%)	576 (90.7%)	263 (93.6%)	87.9%	4503	0.310	0.507
30 mm	196 (72.6%)	556 (87.6%)	261 (92.9%)	85.4%	4272	0.313	0.505
35 mm	182 (67.4%)	550 (86.6%)	255 (90.7%)	83.2%	4186	0.310	0.498
40 mm	179 (66.3%)	524 (82.5%)	258 (91.8%)	81.0%	4224	0.302	0.484
45 mm	169 (62.6%)	516 (81.3%)	254 (90.4%)	79.2%	4204	0.297	0.475
50 mm	154 (57.0%)	502 (79.1%)	252 (89.7%)	76.6%	4297	0.284	0.456

Highest sensitivity and F_2 values are shown in bold.**Table 4**

The performance of the nodule candidate detection per MIP slab thickness at 2 mm intervals.

Slab thickness	Sensitivity	Number of false positives	F_1	F_2
3 mm	86.0%	11,693	0.147	0.292
5 mm	87.9%	9792	0.173	0.334
7 mm	89.8%	8844	0.192	0.363
9 mm	90.0%	8125	0.206	0.383
11 mm	89.2%	6854	0.233	0.418
13 mm	89.0%	6585	0.239	0.426
15 mm	88.7%	5602	0.268	0.461

The highest sensitivity is shown in bold.

Table 5

The number of undetected nodules and sensitivity by nodule characteristics for 10 mm MIP slab thickness.

Characteristics		Total number of nodules	Number of undetected nodules	Sensitivity (%)
Diameter	<5 mm	270	48	82.2
	5–10 mm	635	56	91.2
	>10 mm	281	15	94.7
Density	Non-solid	64	29	54.7
	Part-solid	189	14	92.6
	Solid	933	76	91.9

while showing fewer suspicious lesions on the slice. Although the nodule still can be seen at MIP slab thickness of 25 mm and higher, these thick MIP images are more crowded.

Although it had the highest sensitivity for 10 mm MIP images at the nodule candidate localization, some nodules were still missed. One hundred and nineteen nodules (10.0% of the total) were not detected on 10 mm MIP images. The characteristics of false negatives are shown in Table 5. The size distribution of these undetected nodules was as follows: < 5 mm (48 nodules), 5–10 mm (56 nodules), > 10 mm (15 nodules). When comparing the sensitivities at different densities, only 54.7% of the non-solid nodules were detected, whereas this was 92.6 and 91.9% for the part-solid and solid nodules, respectively.

When the results from the 16 different slab thicknesses were fused, the sensitivity increased to 98.0% and the average false positive rate is 46. Only 24 nodules were undetected on all MIP images. Among these false negatives, there were 6 nodules < 5 mm, 14 nodules in 5–10 mm, and 4 nodules > 10 mm. Among undetected nodules, 37.5% were non-solid, 4.2% were subsolid, 58.3% were solid. It is noteworthy that some non-detected nodules were attached to tissue, which makes detection difficult. Fig. 2 shows some examples of false negatives that were missed at the nodule candidate detection stage in all MIP slab thickness settings.

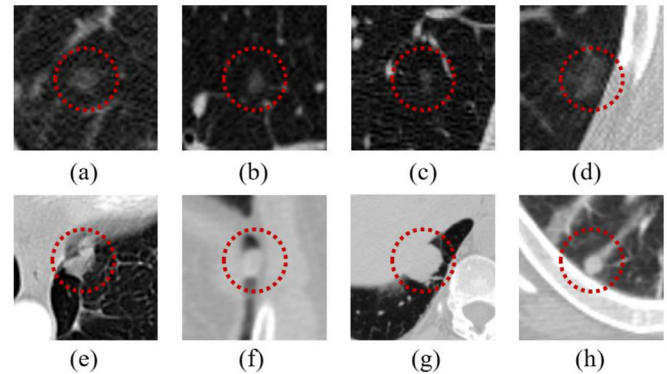


Fig. 2. Examples of false negatives which were missed in all MIP slab thickness settings. The shown slab thickness of the examples is 1 mm. The examples of non-detected nodules were either non-solid nodules (a–d) or attached nodules (e–h).

After false positive reduction, the system (CPM: 0.935) using 16 MIP slab thickness settings outperformed the system (CPM: 0.922) that applied 4 MIP slab thickness settings [14]. The sensitivity of the system with combined results from 16 settings is 0.872, 0.909, 0.925, 0.944, 0.957, 0.966, 0.974 at the false positive rate of 1/8, 1/4, 1/2, 1, 2, 4, 8 FPs/scan, respectively (Fig. 3).

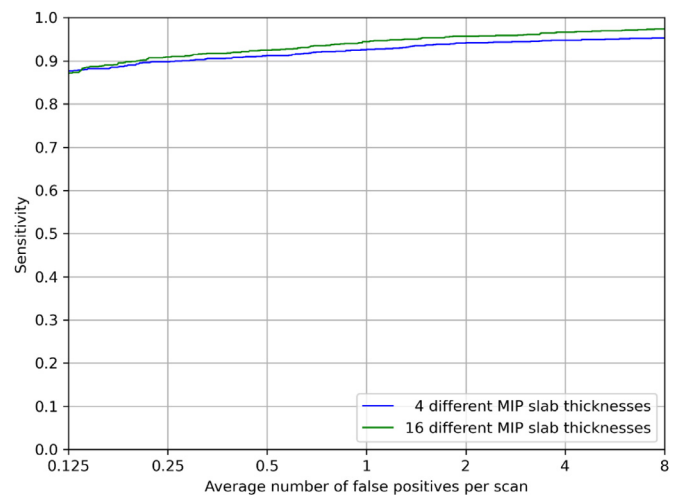


Fig. 3. Free-response receiver operating characteristic (FROC) curves of the system that used MIP images with 4 and 16 slab thicknesses for pulmonary nodule detection.

4. Discussion

The purpose of this study was to explore the effect of slab thickness on lung nodule detection and find the optimal setting at the nodule candidate detection stage. The combination of results from all MIP slab thicknesses improved the sensitivity of the nodule candidate detection to 98%. The results showed that with a slab thickness of 10 mm, the architecture achieved the highest sensitivity for nodule detection.

With different slab thicknesses for MIP images, the same architecture detected different numbers of nodule candidates, comprising true positive and false positive nodules. One reason is that pulmonary nodules stand out differently from vasculatures and lung bronchi depending on the MIP slab thickness (Fig. 1). Moreover, more pulmonary nodules were identified with increasing slab thickness from 1 mm to 10 mm, because vessels are more continuously depicted on thicker MIP images, making it easier to localize isolated nodules. However, beyond a slab thickness of 10 mm, the sensitivity started to decrease. One possible explanation is that more vessels tend to be visible in one single slice, which makes the image more complex for interpretation, resulting in more difficulties for convolutional neural networks to learn complex contextual information of nodules. In addition, thick MIP images may cause overlap of nodules and vessels, leading to false negative results. This finding, of interference of morphological information, has also been reported by Diederich et al. [26] based on the visual assessment by human observers. Nevertheless, although the accuracy reduced with increasing MIP slab thickness, the program found fewer false positives at higher slab thickness (3–50 mm), as shown in Tables 3 and 4. The reason for this is that fewer false positive candidates, such as cross-sectional vessels, appeared in one slice.

To explore the effect of MIP slab thickness on lung nodule detection by human evaluation, prior studies have evaluated different MIP images settings on varied small-scale datasets. Based on visual assessment, Park et al. [7] showed that radiologists found more nodules on 5 mm MIP images than on 1 mm section slices. In another study based on visual assessment, Valencia et al. [27] analyzed the performance of axial 1-mm slices, 5-mm slices, and non-overlapping 10 mm axial/coronal MIP images on the detection of pulmonary nodules. They found 10 mm axial MIP images improved the overall sensitivity because of the higher detection of nodules < 5 mm in diameter. In addition, Li et al. [16] used MIP slab thicknesses of 5 mm, 10 mm, 15 mm, and 20 mm to assess human performance. Their results showed that the nodule detection rate (reader 1: 84.5%; reader 2: 83.6%) on 10 mm MIP images was significantly higher than in other series of MIP images. In the visual evaluation study by Diederich et al. [26], it was found that pulmonary nodules < 5 mm were detected much better on 15 mm MIP images than on 30 mm MIP images. However, this was not seen for nodules > 5 mm, when comparing both these MIP thicknesses. The results of the nodule candidate detection in this study that utilized varied MIP images, are similar to prior studies based on radiologists' findings [14]. Likewise, it also showed a higher sensitivity for 10 mm MIP images than that of 5 mm, 15 mm and 20 mm MIP images. The detection rate of nodules in the 1 mm section slices was lower than that of 5 mm MIP images. Sensitivity in the 15 mm MIP images was higher than in 30 mm MIP images.

The aim of the nodule candidate detection is to find as many true positives as possible, but on the other hand to keep the number of false positive findings as low as possible. If we first only take sensitivity into consideration, the program detected the truest nodules based on the 10 mm MIP images. With the slab thickness of 25 mm, it had the largest F_2 score at the candidate detection stage. Although the number of false positives reduced with 33% when comparing the 25 mm MIP images with the 10 mm MIP images, the program missed 6 more nodules (3 nodules 5–10 mm and

3 nodules > 10 mm). However, these 6 potential undetected nodules would have required follow-up according to the Lung-RADS guidelines [28]. Moreover, the sensitivity of 10 mm was significantly higher than that of 25 mm by the McNemar's test ($p < 0.05$). Therefore, we recommend the use of the 10 mm MIP images setting as optimal setting over the use of the 25 mm MIP images setting, if just a single MIP setting is used. In addition, when the results of 16 different MIP slab thickness settings were merged, the program achieved a high sensitivity of 98.0% at the nodule candidate detection stage. Although more false positives appeared (mean FPs/scan: 46) by combining results from a number of MIP slab thicknesses, it provided good results for the false positive reduction stage. The FROC analysis showed that with the improved sensitivity after combining results from 16 settings at the candidate detection stage, the DL-CAD system can even have good performance (sensitivity: 94.4%, false positive rate: 1.0) for lung nodule detection. In clinical practice, it is not efficient for human observers to review the same scan in multiple slab thicknesses. But the DL-CAD system can process multiple scans at any time and detect more nodules with a low false positive rate by combining results of different MIP slab thickness settings, which shows its potential assistance for radiologists.

Although the program could detect most of the nodules by using 10 mm MIP images or a combination of MIP images with different slab thicknesses at the nodule candidate detection stage, there were still some undetected nodules, most of which were non-solid. These nodules have low attenuation and are easily overlapped by vessels. Extending the training set with scans with more non-solid nodules might improve the detection of these undetected nodules. From the prior study, it is known that the appearance of a solid component, when the non-solid nodule becomes part-solid, is a more suspicious finding [29]. However, for part-solid nodules, the architecture had actually a much better sensitivity with the help of the MIP.

A limitation of this study is that the public dataset was imbalanced for the number of solid, part-solid, and non-solid nodules. Because more solid nodules were present, the architecture thus tended to be better in detecting solid rather than non-solid nodules. This may influence the effect of the optimal MIP slab thickness settings. Also, there might have been a bias due to image quality loss during the process of generating MIP images, resulting in missing some nodules. To create MIP images with a specific slab thickness, the ideal way was using 1 mm axial section slices. But, the public dataset had inconsistent original section thicknesses (0.6–2.5 mm), which makes that slices with a section thickness $\neq 1$ mm had to be interpolated causing a slight change in density values. Another limitation might be that the ground-truth did not have long-term follow-up study or histological verification. It was only determined by the majority vote of the screening radiologists. Some pulmonary nodules can be missed by all four readers, but being detected by the program in some MIP slab thickness settings.

5. Conclusions

We investigated the effect of MIP slab thickness at the nodule candidate detection stage on pulmonary nodule detection. The effect of MIP slab thickness at this stage was comparable to human reader studies. The combination of results from 16 slab thicknesses showed a detection sensitivity of 98% with 46 FPs/scan. For a single MIP setting, with 10 mm MIP images, the scheme had the highest sensitivity for lung nodule detection, similar to the slab thickness usually applied by human observers. The MIP slab thickness of 10 mm and combined results from varying MIP settings can provide better results for false positive reduction in the development of DL-CAD systems.

References

- [1] J.P. de Torres, G. Bastarrrika, J.P. Wisnivesky, A.B. Alcaide, A. Campo, L.M. Seijo, J.C. Pueyo, A. Villanueva, M.D. Lozano, U. Montes, Assessing the relationship between lung cancer risk and emphysema detected on low-dose CT of the chest, *Chest* 132 (2007) 1932–1938.
- [2] F. Bray, J. Ferlay, I. Soerjomataram, R.L. Siegel, L.A. Torre, A. Jemal, Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, *CA Cancer J. Clin.* 68 (2018) 394–424.
- [3] J.E. Walter, M.A. Heuvelmans, K. Ten Haaf, R. Vliegthart, C.M. van der Aalst, U. Yousaf-Khan, P.M. van Ooijen, K. Nackaerts, H.J. Groen, G.H. De Bock, Persisting new nodules in incidence rounds of the NELSON CT lung cancer screening study, *Thorax* 74 (2019) 247–253.
- [4] M. Javid, M. Javid, M.Z. Rehman, S.I. Shah, A novel approach to CAD system for the detection of lung nodules in CT images, *Comput. Methods Programs Biomed.* 135 (2016) 125–139.
- [5] M. Oudkerk, A. Devaraj, R. Vliegthart, T. Henzler, H. Prosch, C.P. Heussel, G. Bastarrrika, N. Sverzellati, M. Mascalchi, S. Delorme, European position statement on lung cancer screening, *Lancet Oncol.* 18 (2017) e754–e766.
- [6] H. Broekhuizen, C.G. Groothuis-Oudshoorn, R. Vliegthart, H.J. Groen, M.J. IJzerman, Assessing lung cancer screening programs under uncertainty in a heterogeneous population, *Value Health* 21 (2018) 1269–1277.
- [7] J.F. Gruden, S. Ouanounou, S. Tigges, S.D. Norris, T.S. Klausner, Incremental benefit of maximum-intensity-projection images on observer detection of small pulmonary nodules revealed by multidetector CT, *Am. J. Roentgenol.* 179 (2002) 149–157.
- [8] A. Jankowski, T. Martinelli, J.-F. Timsit, C. Brambilla, F. Thony, M. Coulomb, G. Ferretti, Pulmonary nodule detection on MDCT images: evaluation of diagnostic performance using thin axial images, maximum intensity projections, and computer-assisted detection, *Eur. Radiol.* 17 (2007) 3148–3156.
- [9] E.-A. Park, J.M. Goo, J.W. Lee, C.H. Kang, H.J. Lee, C.H. Lee, C.M. Park, H.Y. Lee, J.-G. Im, Efficacy of computer-aided detection system and thin-slab maximum intensity projection technique in the detection of pulmonary nodules in patients with resected metastases, *Invest. Radiol.* 44 (2009) 105–113.
- [10] M.M. Najafabadi, F. Villanustre, T.M. Khoshgoftaar, N. Seliya, R. Wald, E. Muharemagic, Deep learning applications and challenges in big data analytics, *J. Big Data* 2 (2015) 1.
- [11] R. Miotto, F. Wang, S. Wang, X. Jiang, J.T. Dudley, Deep learning for healthcare: review, opportunities and challenges, *Brief. Bioinform.* 19 (2017) 1236–1246.
- [12] A.A.A. Setio, A. Traverso, T. De Bel, C. van den Bogaard, P. Cerello, H. Chen, Q. Dou, M.E. Fantacci, B. Geurts, Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge, *Med. Image Anal.* 42 (2017) 1–13.
- [13] C. Jacobs, E.M. van Rikxoort, K. Murphy, M. Prokop, C.M. Schaefer-Prokop, B. van Ginneken, Computer-aided detection of pulmonary nodules: a comparative study using the public LIDC/IDRI database, *Eur. Radiol.* 26 (2016) 2139–2147.
- [14] S. Zheng, J. Guo, X. Cui, R.N. Veldhuis, M. Oudkerk, P.M. van Ooijen, Automatic pulmonary nodule detection in CT scans using convolutional neural networks based on maximum intensity projection, *IEEE Trans. Med. Imaging* 39 (2020) 797–805.
- [15] Z. Hu, A. Muhammad, M. Zhu, Pulmonary nodule detection in CT images via deep neural network: nodule candidate detection, in: *Proceedings of the 2nd International Conference on Graphics and Signal Processing*, 2018, pp. 79–83.
- [16] W.-J. Li, Z.-G. Chu, Y. Zhang, Q. Li, Y.-N. Zheng, F.-J. Lv, Effect of slab thickness on the detection of pulmonary nodules by use of CT maximum and minimum intensity projection, *Am. J. Roentgenol.* (2019) 1–6.
- [17] S.G. Armato III, G. McLennan, L. Bidaut, M.F. McNitt-Gray, C.R. Meyer, A.P. Reeves, B. Zhao, D.R. Aberle, C.I. Henschke, E.A. Hoffman, The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans, *Med. Phys.* 38 (2011) 915–931.
- [18] S.G. Armato III, G. McLennan, M.F. McNitt-Gray, C.R. Meyer, D. Yankelevitz, D.R. Aberle, C.I. Henschke, E.A. Hoffman, E.A. Kazerooni, H. MacMahon, Lung image database consortium: developing a resource for the medical imaging research community, *Radiology* 232 (2004) 739–748.
- [19] D.P. Naidich, A.A. Bankier, H. MacMahon, C.M. Schaefer-Prokop, M. Pistolesi, J.M. Goo, P. Macchiarini, J.D. Crapo, C.J. Herold, J.H. Austin, Recommendations for the management of subsolid pulmonary nodules detected at CT: a statement from the Fleischner Society, *Radiology* 266 (2013) 304–317.
- [20] J.-S. Kim, J.-H. Kim, G. Cho, K.T. Bae, Automated detection of pulmonary nodules on CT images: effect of section thickness and reconstruction interval—Initial results, *Radiology* 236 (2005) 295–299.
- [21] Y.J. Huang, R. Powers, G.T. Montelione, Protein NMR recall, precision, and F-measure scores (RPF scores): structure quality assessment measures based on information retrieval statistics, *J. Am. Chem. Soc.* 127 (2005) 1665–1674.
- [22] D.M. Powers, Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation, 2011.
- [23] N. Hawass, Comparing the sensitivities and specificities of two diagnostic procedures performed on the same group of patients, *Br. J. Radiol.* 70 (1997) 360–366.
- [24] M. Niemeijer, M. Loog, M.D. Abramoff, M.A. Viergever, M. Prokop, B. van Ginneken, On combining computer-aided detection systems, *IEEE Transactions on Medical Imaging* 30 (2010) 215–223.
- [25] A.I. Bandos, H.E. Rockette, T. Song, D. Gur, Area under the free-response ROC curve (FROC) and a related summary index, *Biometrics* 65 (2009) 247–256.
- [26] S. Diederich, M. Lentschig, T. Overbeck, D. Wormanns, W. Heindel, Detection of pulmonary nodules at spiral CT: comparison of maximum intensity projection sliding slabs and single-image reporting, *Eur. Radiol.* 11 (2001) 1345–1350.
- [27] R. Valencia, T. Denecke, L. Lehmkuhl, F. Fischbach, R. Felix, F. Knollmann, Value of axial and coronal maximum intensity projection (MIP) images in the detection of pulmonary nodules by multislice spiral CT: comparison with axial 1-mm and 5-mm slices, *Eur. Radiol.* 16 (2006) 325–332.
- [28] S.J. van Riel, C. Jacobs, E.T. Scholten, R. Wittenberg, M.M.W. Wille, B. de Hoop, R. Sprengers, O.M. Mets, B. Geurts, M. Prokop, Observer variability for Lung-RADS categorisation of lung cancer screening CTs: impact on patient management, *Eur. Radiol.* 29 (2019) 924–931.
- [29] H. MacMahon, D.P. Naidich, J.M. Goo, K.S. Lee, A.N. Leung, J.R. Mayo, A.C. Mehta, Y. Ohno, C.A. Powell, M. Prokop, Guidelines for management of incidental pulmonary nodules detected on CT images: from the Fleischner Society 2017, *Radiology* 284 (2017) 228–243.