# 博 士 学 位 論 文

1mm厚コンピュータ断層画像での肺結節検出における、深層学習を備えたコンピュータ支援診断（CAD）システムの有効性

近 畿 大 学 大 学 院 医 学 研 究 科

医 学 系 放 射 線 診 断 ・ 画 像 応 用 治 療 学

小 　 塚 　 健 　 倫

Doctoral Dissertation

Efficiency of a computer-aided diagnosis (CAD) system with deep learning in detection of pulmonary nodules on 1-mm-thick images of computed tomography

November 2020

Department of Radiology, Major in Medical Sciences
Kindai University Graduate School of Medical Sciences

**Takenori Kozuka**

同　意　書

2020 年　11 月　11 日

近畿大学大学院
　医学研究科長　殿

共著者　松久保 祐子　㊞　　　共著者　任 誠雲　㊞

共著者　門場 智也　㊞　　　共著者　甲斐田 勇人　㊞

共著者　小田 晃義　㊞　　　共著者　柳生 行伸　㊞

共著者　鈴木 約子　㊞　　　共著者　鶴崎 正勝　㊞

共著者　兵頭 朋子　㊞　　　共著者　松木 充　㊞

論文題目

Efficiency of a computer-aided diagnosis (CAD) system with deep
learning in detection of pulmonary nodules on 1-mm-thick images of
computed tomography

下記の学位論文提出者が、標記論文を貴学医学博士の学位論文（主論文）

として使用することに同意いたします。

また、標記論文を再び学位論文として使用しないことを誓約いたします。

記

1．学位論文提出者氏名　　　　　　　　小塚　健倫

2．専攻分野　　医学系　　　　　　　近畿大学医学部放射線医学講座
　　　　　　　　　　　　　　　　　　放射線診断学部門

近畿大学大学院医学研究科

論博

# 同　意　書

2020 年　11 月　11 日

近畿大学大学院
　医学研究科長　殿

共著者　石井一成　㊞　　　　共著者 ＿＿＿＿＿＿＿ ㊞

共著者 ＿＿＿＿＿＿＿ ㊞　　　共著者 ＿＿＿＿＿＿＿ ㊞

共著者 ＿＿＿＿＿＿＿ ㊞　　　共著者 ＿＿＿＿＿＿＿ ㊞

共著者 ＿＿＿＿＿＿＿ ㊞　　　共著者 ＿＿＿＿＿＿＿ ㊞

共著者 ＿＿＿＿＿＿＿ ㊞　　　共著者 ＿＿＿＿＿＿＿ ㊞

論文題目

Efficiency of a computer-aided diagnosis (CAD) system with deep
learning in detection of pulmonary nodules on 1-mm-thick images of
computed tomography

　下記の学位論文提出者が、標記論文を貴学医学博士の学位論文（主論文）

として使用することに同意いたします。

また、標記論文を再び学位論文として使用しないことを誓約いたします。

記

1．学位論文提出者氏名　　　　　　　　小塚　健倫

2．専攻分野　　医学系　　　　　　近畿大学医学部放射線医学講座
　　　　　　　　　　　　　　　　　　放射線診断学部門

近畿大学大学院医学研究科

**Title:**

Efficiency of a computer-aided diagnosis (CAD) system with deep learning in
detection of pulmonary nodules on 1-mm-thick images of computed tomography

Authors:

Takenori Kozuka,      takkozuka@gmail.com

Yuko Matsukubo,      moucocco_sheepdog@yahoo.co.jp

Tomoya Kadoba,      t_kadoba@yahoo.co.jp

Teruyoshi Oda,      teru.oda0321@gmail.com

Ayako Suzuki,      i_am_sue_19861211_pinky@yahoo.co.jp

Tomoko Hyodo,      neneth@m.ehime-u.ac.jp

SungWoon Im,      imrad5222@gmail.com

Hayato Kaida,      kaida@med.kindai.ac.jp

Yukinobu Yagyu,      y-yagyu@med.kindai.ac.jp

Masakatsu Tsurusaki,           mtsuru@dk2.so-net.ne.jp

Mitsuru Matsuki,      rad053@osaka-med.ac.jp

Kazunari Ishii      ishii@med.kindai.ac.jp

Corresponding author: Takenori Kozuka

Department of Radiology, Kindai University Faculty of Medicine,

377-2 Ohno-Higashi, Osaka-Sayama, Osaka, 589-8511, Japan.

TEL: +81-72-366-0221（3133）

FAX: +81-72-367-1685

the type of manuscript: Original article

Efficiency of a computer-aided diagnosis (CAD) system with deep learning in detection of pulmonary nodules on 1-mm-thick images of computed tomography

**Abstract**

Purpose: To evaluate the performance of a deep learning-based Computer-aided Diagnosis (CAD) system at detecting pulmonary nodules on CT by comparing radiologists' readings with and without CAD.

Materials and methods: A total of 120 chest CT images were randomly selected from patients with suspected lung cancer. The gold standard of nodules ≥3mm was established by a panel of three expert radiologists. Two less-experienced radiologists read the images without and afterward with CAD system. Their reading times were recorded.

Results: The radiologists' sensitivity increased from 20.9% to 38.0% with the introduction of CAD. The positive predictive value (PPV) decreased from 70.5% to 61.8%, and the F1-score increased from 32.2% to 47.0%. The sensitivity significantly increased from 13.7% to 32.4% for small nodules (3–6 mm) and from 33.3% to 47.6% for medium nodules (6–10 mm). CAD alone showed a sensitivity of 70.3%, a PPV of 57.9%, and an F1-score of 63.5%. Reading time decreased by 11.3% with the use of CAD.

Conclusion: CAD improved the less-experienced radiologists' sensitivity in detecting pulmonary nodules of all sizes, especially including a significant improvement in the detection of clinically important-sized medium nodules (6–10 mm) as well as small nodules (3–6 mm) and reduced their reading time.

**Introduction**

Lung cancer is one of the most widespread diseases worldwide and the leading cause of cancer-related death. As reported in Global Cancer Statistics 2018, lung cancer remains the leading cause of cancer incidence and mortality worldwide, with 2.1 million new lung cancer cases and 1.8 million deaths predicted in 2018, representing close to 1 in 5 (18.4%) cancer deaths [1].

Early diagnosis is important in lung cancer practice to improve the effectiveness of treatment and increase patients' chances of survival [2]. Advances in computed tomography (CT) technology have enabled early diagnosis [3]. Low-dose computed tomography (LDCT) is a recommended modality for lung cancer screening in the United States [4]. The largest National Lung Screening Trial (NLST) showed a 20% reduction in lung cancer mortality rate in participants screened with low-dose helical CT compared with radiography [5]. The NLST also showed that more than 60% of identified lung cancers were stage I or II [5, 6]. Many such early-stage lung cancers are small and may be overlooked, and such failure to detect small nodules may cause delays in diagnosis and treatment [7]. Because CT screening work is a huge workload on radiologists, they may overlook small nodules and/or misinterpret images [8, 9]. A study reported that 20%–35% of small lung nodules were missed in screening diagnosis by a single radiologist [10]. Less-experienced radiologists generally have lower lung cancer detection rates than experienced radiologists have. Several studies have reported that computer-aided diagnosis (CAD) systems can improve less-experienced radiologists' detection rates [8, 11, 12].

The field of artificial intelligence (AI) has progressed greatly with advances in deep learning technology in the 2010s. In general, deep learning consists of massive multilayer networks of artificial neurons that can automatically discover useful features (i.e., representations of input data needed for tasks such as detection) given large amounts of training data [13-15]. While the performance of conventional machine learning algorithms was highly dependent on extracted feature quantities designed by humans [13], deep learning has improved performance because feature quantities are automatically extracted. Deep learning has the advantage that it can analyze huge amounts of medical images without the acquisition of systematic knowledge of diagnostic imaging anatomy without fatigue. Several studies have shown that deep learning-based CAD systems have excellent detection rates and lower false positive (FP) rates [16-18].

In this study, we used a CAD system to detect pulmonary nodules on chest CT images so as to assist radiologists, and the final diagnosis must be made by doctors. The CAD system was based on a two-stage object detection system deriving from the well-known Faster R-CNN framework, which could achieve a high localization and identification accuracy compared with the one-stage object detection system [19].

The CAD system used in this study was based on deep learning with a neural network structure that detects pulmonary nodules on chest CT images. The CAD system is intended to assist radiologists, and the final diagnosis must be made by doctors.

The primary objective of the study was to confirm that the assistance of deep learning-based

CAD improves less-experienced radiologists' detection rates of pulmonary nodules on CT images by comparing the sensitivity between reading with and without CAD. The secondary objective was to confirm that CAD reduces radiologists' reading time.

**Materials and Methods**

This was a retrospective study that used CT images of cases with suspected lung cancer. The study was approved by the ethics committee of Kindai University, and a waiver of informed consent was obtained.

**CT studies**

The study dataset consisted of 120 chest CT images randomly selected from cases of suspected lung cancer in patients aged 20 years or older who underwent CT examination at Kindai University Hospital between November and December 2018. Cases of inappropriate image quality, contrast medium, pneumonia, diffuse lung disease, massive pleural effusion/atelectasis, and severe postoperative complications were excluded. CT scans were performed on an 80-detector row CT scanner (Aquilion™ Prime, Canon Medical Systems Corporation, Japan). Unenhanced CT was performed with the following settings: tube voltage, 120kV; effective tube current, 128/194 mAs with Volume Exposure Control (SD: 7.27 for 5-mm image); pitch, 1.5; filter and function, high-frequency algorithm (FC86); detector configuration, 40×1.0 mm; window level, 1200HU, window width, -700HU. All images were reconstructed at 1 mm/1 mm and stored on the hospital PACS (SYNAPSE).

The gold standard regarding the presence of nodules was established by a panel of three expert radiologists with 26, 6, and 12 years of diagnostic experience. Initially, two experienced radiologists read the CT images to determine the locations of nodule marks and types of nodules. Only when the two experienced radiologists' results were different did the third experienced radiologist make the final judgment. This result served as the gold standard.

Two less-experienced radiologists (reader A: 5 years, reader B: 1 year of diagnostic experience) performed comparative reading tests, first without CAD and then with CAD. In the first part, each reader was asked to read 120 thoracic CT images without CAD. Workstations with general interpretation functions (e.g., change of window level and window width, zoom, pan, magnifying glass) were provided to the readers. Each reader was asked to mark each nodule and annotate the type of nodule. The second part was conducted after at least 14 days' interval. In the second part, the same two radiologists read the same CT images again using CAD that displayed the images with marks. Marks were considered true-positive (TP) findings if the marks annotated by the readers were within the gold standard marks, otherwise, they were regarded as FP findings. If the readers did not mark the location of gold standard nodules, it was regarded as a false negative (FN).

**Nodules**

The nodules to be detected in this study were those with a major axis of ≥3mm [20]. All sizes were included in the detection during the gold standard preparation and comparative reading test, and nodules <3mm were excluded at the aggregation step. The types of nodules were classified into solid nodules, part-solid nodules, calcified nodules, and ground-glass nodules (GGNs).

**Computer-Aided Diagnosis System**

The CAD system used in this study was "InferRead CT Lung," developed by Infervision Co., Ltd. It is a pulmonary nodule detection AI software package based on deep learning. This CAD system has functions to display marks, density, major axis, and the volume of detected nodules (Fig. 1). The backbone of this CAD system was the well-known Faster R-CNN framework, which was a two-stage object detection system achieving a high localization and identification accuracy. The input of the CAD system was CT images after data enhancement including randomly resizing, cropping, padding, and flipping the original CT images, while the output of this system was the results of object detection, i.e., object bounding boxes and scores. The CAD system was mainly composed of four parts, including Region Proposal Network (RPN), Region of Interest (ROI) Pooling Layer, classification, and regression. Firstly, the RPN was utilized to generate candidate object bounding boxes, and then the corresponding features of each ROI were extracted by ROI Pooling Layer for the subsequent classification and regression parts [19]. The location of the object bounding-boxes was specified by using the regression part, and the candidate object bounding-boxes were distinguished true-positives from false-positives through the classification part. Fig. 6 shows the basic structure of the current CAD system. Other detailed information on this CAD system including the source, the size, and the quality of the training data as well as the hyper-parameters of the neural network have not been published. Furthermore, no research has ever evaluated the functionality of this CAD system.

**Reading time**

Each reader's reading time for all cases was recorded and compared between reading without and with CAD.

**Statistical Analysis**

Statistical analyses were performed from two perspectives: per-nodule and per-patient. In the per-nodule analysis, TP, FP, and FN were judged for each nodule on the results of the comparative reading tests without and with CAD and using CAD alone. Then, each sensitivity, positive predictive value (PPV), and F1-score was calculated. The F1-score is defined as the weighted harmonic mean of the test's PPV and sensitivity. The F1-score can provide a more realistic measure of a test's performance by using both PPV and sensitivity. Machine Learning is a good application of the F1-score.

In the per-patient analysis, sensitivity and specificity were calculated for each case [21]. Per-patient sensitivity was defined as the ratio between the number of patients with at least one TP nodule

6

detected by the reading test and the total number of patients with at least one nodule according to the gold standard. Per-patient specificity was defined as the ratio between the number of patients with no nodules found during the reading test and the total number of patients without nodules according to the gold standard. Per-nodule and per-patient sensitivity in readings without and with CAD were only compared using McNemar's test. The sensitivity in readings without CAD vs. CAD alone and in readings with CAD vs. CAD alone were not compared using McNemar's test because CAD is just assistance for clinical diagnosis and CAD alone does not diagnose in the real clinical scene. A p-value of less than 0.05 was considered statistically significant.

## Results

Fig. 2 shows the flowchart of the study's selection process. The number of patients randomly selected was 120, of whom 3 were excluded because of diffuse nodules. Therefore, the number of effective cases was 117 (62 males, 55 females: age 21-89, mean age 65.0). According to the gold standard, 111 patients had at least one ≥3mm nodule, and 6 patients had no nodules. 743 nodules ≥3mm were detected; 12 nodules were diagnosed to lung cancer, 44 nodules were followed up as nodules suspected lung cancer, and 687 nodules were followed up as nodules suspected benign. The major axis of the nodules averaged 5.7 mm and had a median of 4.7 mm, and the number of nodules in cases containing at least one nodule averaged 6.9 (median: 5). The most common category of major axis size was 3–6 mm. which accounted for 72% (532/743) of detected nodules. The most common nodule type was solid nodule, which accounted for 70% (518/743) of nodules found (Table 1).

## Per-nodule analysis

Fig. 3 shows examples of typical detection by CAD. The result of the per-nodule analysis is shown in Table 2 and 3. The total sensitivity of two readers without CAD was 20.9% (310/1486, 95% CI: 0.188–0.230), and the value with CAD was 38.0% (564/1486, 95% CI: 0.355–0.405, p<0.01). The number of TP increased to 564 from 310 when CAD was used. The sensitivity of reader A and reader B increased from 22.7% to 35.3% and from 19.0% to 40.6%, respectively. The total PPV decreased from 70.5% (310/440, 95% CI: 0.660–0.747) to 61.8% (564/912, 95% CI: 0.586–0.650), and the number of FP increased to 348 from 130. The PPV without and with CAD were almost unchanged (69.0% and 69.5%, respectively) in reader A, whereas the PPV of reader B decreased from 72.3% to 56.4%. The numbers of FP increased from 76 to 115 in reader A and from 54 to 233 in reader B. The total F1-score of both readers increased from 32.2% to 47.0% with CAD.

The performance of CAD alone had a sensitivity of 70.3% (95% CI: 0.668–0.735), a PPV of 57.9% (95% CI: 0.546–0.611), and an F1-score of 63.5%. CAD alone had higher sensitivity and F1-score values than those of radiologists regardless of whether they used CAD. Instead, CAD alone showed the highest number of FP (380). Fig. 4 shows examples of FP findings by CAD.

7

The readers' sensitivity increased with CAD regardless of the nodules' major axis size. The sensitivity increased for nodules 3–6 mm, 6–10 mm, 10–15 mm, 15–20 mm, and >20 mm from 13.7% (95% CI: 0.117–0.159) to 32.4% (95% CI: 0.296–0.353) (p<0.01), from 33.3% (95% CI: 0.279–0.391) to 47.6% (95% CI: 0.417–0.535) (p<0.01), from 51.1% to 58.7% (p=0.07), from 46.4% to 60.7% (p=0.2), and from 57.1% to 78.6% (p=0.2), respectively. Thus, the detection of 3–6-mm nodules was improved the most (p<0.01). The sensitivity of CAD alone for nodules 3–6 mm, 6–10 mm, 10–15 mm, 15–20 mm, and >20 mm were 73.9% (393/532), 60.4% (87/144), 58.7% (27/46), 64.3% (9/14), and 85.7% (6/7), respectively.

In analyses by nodule type, CAD improved readers' total sensitivity for solid nodules, part-solid nodules, calcified nodules, and GGNs from 18.6% to 32.6% (p<0.01), from 31.5% to 58.5% (p<0.01), from 30.4% to 54.7% (p<0.01), and from 18.0% to 40.1% (p<0.01), respectively. The detection of GGNs was improved the most. The sensitivity of CAD alone in detecting solid nodules, part-solid nodules, calcified nodules, and GGNs, were 68.1% (353/518), 70.8% (46/65), 82.4% (61/74), and 72.1% (62/86), respectively, with higher sensitivity for calcified nodules. Fig. 5 shows examples of CAD detection of GGNs that were overlooked by readers without CAD.

**Per-patient analysis**

The results of the per-patient analysis are shown in Table 4. The readers' total sensitivity was significantly improved, from 68.0% (151/222, 95% CI: 0.614–0.741) to 85.1% (189/222, 95% CI: 0.798–0.895), with the use of CAD (p<0.01). The readers' total specificity decreased from 91.7% (11/12) to 83.3% (10/12) with the use of CAD. The performance of CAD alone was a sensitivity of 95.5% (106/111) and a specificity of 83.3% (5/6).

**Reading time**

As shown in Table 5, the total reading time decreased by 11.3% with CAD, from 373 minutes to 331 minutes. The reading time decreased by 10.4% in Reader A, by 11.9% in Reader B. The mean reading time for one case was 3.1 minutes without CAD and 2.8 minutes with CAD.

**Discussion**

In this study, we evaluated improvement in the detection rate of pulmonary nodules among less-experienced radiologists generated by using CAD with AI. Similar to the results of previous studies, our study showed that less-experienced radiologists assisted with CAD with AI also could detect more nodules than they could without the CAD [7, 21-27]. The readers' sensitivity was significantly improved (from 20.9% to 38.0%) by using CAD (p<0.01). Both readers showed similar improvement tendencies. These results suggest that assistance by the CAD system could help to prevent less-experienced radiologists from overlooking pulmonary nodules.

Although CAD alone detected more nodules than readers' performance with CAD, the CAD

alone introduced a 5.8-fold (380/65) increase in FP compared with the readers' performance without CAD. We found that the CAD system sometimes detected non-nodule objects such as blood vessels, protruding pleura, and ground-glass opacities caused by lack of air intake or inflammation (Fig. 4). Because CAD systems generally detect many candidate nodules, including FP findings, the number of FP should also be taken into consideration for the evaluation of CAD systems. The mean number of FP found using CAD alone was 3.2 per case, a comparable value to those found in previous studies [21, 28-31]. The number of FP in the radiologists' readings increased 2.7-fold (348/130) with the introduction of this CAD because of FP findings by the CAD. The reduction in PPV with CAD use was observed only in reader B (from 72.3% to 56.4%). This suggests that radiologists' judgment abilities are important after the detection of candidate nodules.

Regarding the analysis according to the major axis, readings without CAD had a sensitivity of 13.7% for small nodules (3–6 mm), which was significantly lower than that for nodules ≥6mm ($p < 0.01$). The CAD significantly improved readers' sensitivity in detecting small nodules (3–6 mm) to 32.4% ($p < 0.01$). This tendency was found in both readers. We found that the CAD alone had a great advantage in detecting small nodules (3–6 mm), as it showed a higher sensitivity (73.9%) than the readers' results ($p < 0.01$). Furthermore, readers' sensitivity for medium nodules (6–10 mm) significantly improved from 33.3% to 47.6% with the use of CAD ($p < 0.01$). Several studies have shown that the prevalence of malignancy varied by size: 0%–1% for nodules <5mm and 6%–28% for nodules 5–10 mm [32]. Therefore, the result of significant improvement for medium nodules (6–10 mm) suggests that the CAD could be effective for early detection of malignant nodules.

In the Dutch Belgian Lung Cancer Screening Trial, 22 (36%) lung cancers diagnosed on post-screening CTs were overlooked, and 20 of 22 missed lung cancers (91%) were due to errors of detection [33]. A report based on the International Early Lung Cancer Action Program study showed that in 75% of patients with confirmed cancer, the corresponding small nodules could be detected in previous CT scans [34]. We expect that CAD with AI could complement the detection of small lung nodules that doctors may overlook and contribute to the early detection of lung cancers.

Regarding the analysis according to nodule type, CAD significantly improved the radiologists' sensitivity for all nodule types. The sensitivity increase was the greatest for GGNs. It has been reported that part-solid nodules and GGNs are more likely to be malignant than solid nodules are [35]; however, it is generally difficult for CAD systems to detect part-solid nodules and GGNs [29]. This study showed that the sensitivity values of the CAD alone for part-solid nodules and GGNs were not inferior to those for solid nodules (Fig. 5).

In clinical screening for pulmonary nodules, it is problematic if radiologists diagnose patients with at least one nodule as having no nodules because such patients may miss opportunities for treatment. It is therefore a great advantage that CAD with AI significantly improved per-patient sensitivity from 68.0% to 85.1% ($p < 0.01$). CAD alone showed a specificity of 83.3% (value comparable to that of radiologists)

despite many FP; however, statistical evaluation was difficult because only 6 patients had no nodules of diameter ≥3mm.

A general concern with CAD systems is increased reading time. In previous studies, both cases of increase and decrease have been reported [7, 21]. In this study, CAD with AI technology reduced radiologists' reading time by an average of 11.3%. This result suggests that CAD with AI could reduce the burden on radiologists and improve the workflow of chest CT with a large number of examinations.

Deep learning requires a massive number of training images for learning before practical use. The "InferRead CT Lung" system has already been used in clinical practice in more than 200 medical institutions in China. In addition, significant amounts of training data have already been obtained. Because the diagnosis based on the original image depends on the accuracy of radiologists' reading, the diagnosis result is not always true. Therefore, the training data need to be determined based on certain decisions. In many cases, the correct answer is determined by a consensus of several doctors. In the case of this CAD system, the ground truth annotations of the training data were formed based on evaluations by multiple doctors. However, no research has ever evaluated the functionality of this AI software.

Our study has several limitations. Without histological proof, lung nodules are radiological evidence and thus tend to be interpreted variably even among experienced radiologists [36]. However, we believe that variation could have been minimized by the use of a panel of three expert radiologists to establish the gold standard.

Because this CAD system was designed for CT images with 1-mm-thick sections, we reconstructed 1-mm-thick images for this study, although the radiologists usually read 5-mm-thick images. This method may have caused the more-experienced radiologists to find more nodules in the preparation process of our gold standard than they would find under usual medical care. Furthermore, the less-experienced radiologists might have been unfamiliar with reading 1-mm-thick images.

The sensitivity of this examination was relatively low compared with former literatures. This may be due to the fact that the readers had only one or five years of experience. However, pulmonary nodules may not be always detected by experienced clinicians, but rather by inexperienced radiologists or non-specialist clinicians in daily clinical scenes. This suggests that this CAD may be useful for clinical practice. The definition of lung nodule is ambiguous. Especially ground glass nodule and ground-glass attenuation or solid nodule and consolidation are difficult to distinguish. Furthermore, lung nodules adjacent to pleura and pleural nodules are difficult to distinguish. The distinction may be difficult for inexperienced radiologists.

In conclusion, CAD with AI significantly improved the assisting (i.e., less-experienced) radiologists' sensitivity for the detection of pulmonary nodules and reduced their reading time. The assistance of the CAD system significantly improved the readers' sensitivity in detecting clinically important-sized medium nodules (6–10 mm) as well as small nodules (3–6 mm). The results suggest that the use of the CAD with AI can help to prevent physicians from overlooking pulmonary nodules and

contribute to efficient reading for early detection of lung cancer.

**References**

1.  Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global Cancer Statistics 2018. CA: Cancer J Clin. 2018; 68:394-424.

2.  A. El-Baz, J. Suri. Lung Imaging and Computer Aided Diagnosis, 1st edition. 2011; Taylor & Francis, Abingdon.

3.  Li Q. Recent progress in computer-aided diagnosis of lung nodules on thin-section CT. Computed Imaging Graph. 2007; 31(4-5):248-57.

4.  Centers for Medicare & Medicaid Services (CMS). Decision memo for screening for lung cancer with low dose computed tomography (LDCT) (CAG-00439N). 2015; Available via https://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=274.

5.  National Lung Screening Trial Research Team; Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med. 2011; 365(5):395-409.

6.  Aberle DR, DeMello S, Berg CD, Black WC, Brewer B, Church TR, et al. Results of the two incidence screenings in the National Lung Screening Trial. N Engl J Med. 2013; 369(10):920-31.

7.  Lo SCB, Freedman MT, Gillis LB, White CS, Mun SK. Computer-aided detection of lung nodules on CT with a computerized pulmonary vessel suppressed function. Am J Roentgenol. 2018; 210:480-8.

8.  B Al Mohammad, PC Brennan, C Mello-Thoms. A review of lung cancer screening and the role of computer-aided detection. Clin Radiol. 2017; 72(6):433-42.

9.  Li F, Sone S, Abe H, MacMahon H, Armato SG III, Doi K. Lung cancers missed at low-dose helical CT screening in a general population: Comparison of clinical, histopathologic, and imaging findings. Radiology. 2002; 225:673-83.

10. Torres EL, Fiorina E, Pennazio F, Peroni C, Saletta M, Camarlinghi N, et al. Large scale validation of the M5L lung CAD on heterogeneous CT datasets. Med Phys. 2015; 42:1477-89.

11. Goo JM, Kim HY, Lee JW, Lee HJ, Lee CH, Lee KW, et al. Is the computer-aided detection scheme for lung nodule also useful in detecting lung cancer? J Comput Assist Tomogr. 2008; 32(4):570-5.

12. Marten K, Seyfarth T, Auer F, Wiener E, Grillhösl A, Obenauer S, et al. Computer-assisted detection of pulmonary nodules: performance evaluation of an expert knowledge-based detection system in consensus reading with experienced and inexperienced chest radiologists. European Radiology. 2004; 14:1930-8 .

13. Sahiner B, Pezeshk A, Hadjiiski LM, Wang X, Drukker K, Cha KH, et al. Deep learning in medical imaging and radiation therapy. Med. Phys. 2019; 46 (1):e1-36.

14. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015; 521:436-44.

15. Schmidhuber J. Deep learning in neural networks: an overview. Neural Netw. 2015; 61:85-117.

16. da Silva GLF, Valente TLA, Silva AC, de Paiva AC, Gattass M. Convolutional neural network-based PSO for lung nodule false positive reduction on CT images. Comput Methods Programs Biomed. 2018; 162:109-18.

17. Dou Q, Chen H, Yu L, Qin J, Heng PA. Multilevel contextual 3-D CNNs for false positive reduction in pulmonary nodule detection. IEEE Trans Biomed Eng. 2017; 64:1558-67.

18. Li W, Cao P, Zhao D, Wang J. Pulmonary nodule classification with deep convolutional neural networks on computed tomography images. Comput Math Methods Med. 2016; 6215085.

19. Ren, S., He, K., Girshick, R., & Sun, J. Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence. Faster R-CNN. 2017; 39(6), 1137-1149.

20. MacMahon H, Bankier AA, Naidich DP. Lung cancer screening: what is the effect of using a larger nodule threshold size to determine who is assigned to short-term CT follow-up? Radiology. 2014; 273:326–7.

21. Vassallo L, Traverso A, Agnello M, Bracco C, Campanella D, Chiara G, et al. A cloud-based computer-aided detection system improves identification of lung nodules on computed tomography scans of patients with extra-thoracic malignancies. Eur Radiol. 2019; 29(1):144-52.

22. Li L, Liu Z, Huang H, Lin M, Luo D. Evaluating the performance of a deep learning - based computer - aided diagnosis (DL - CAD) system for detecting and characterizing lung nodules: Comparison with the performance of double reading by radiologists. Thoracic Cancer. 2019; 10:183-92.

23. Armato SG III, Li F, Giger ML, MacMahon H, Sone S, Doi K. Lung cancer: Performance of automated lung nodule detection applied to cancers missed in a CT screening program. Radiology. 2002; 225:685-92.

24. Yuan R, Vos PM, Cooperberg PL. Computer-aided detection in screening CT for pulmonary nodules. AJR Am J Roentgenol. 2006; 186:1280-7.

25. Lee IJ, Gamsu G, Czum J, Wu N, Johnson R, Chakrapani S. Lung nodule detection on chest CT: Evaluation of a computer-aided detection (CAD) system. Korean J Radiol. 2005; 6:89-93.

26. Jacobs C, van Rikxoort EM, Murphy K, Prokop M, Schaefer-Prokop CM, van Ginneken B. Computer-aided detection of pulmonary nodules: A comparative study using the public LIDC/IDRI database. Eur Radiol. 2016; 26:2139-47.

27. Zhao Y, de Bock GH, Vliegenthart R, van Klaveren RJ, Wang Y, Bogoni L, et al. Performance of computer-aided detection of pulmonary nodules in low-dose CT: Comparison with double reading by nodule volume. Eur Radiol. 2012; 22:2076-84.

28. Rubin GD, Lyo JK, Paik DS, Sherbondy AJ, Chow LC, Leung AN, et al. Pulmonary nodules on multi-

detector row CT scans: performance comparison of radiologists and computer-aided detection. Radiology. 2005; 234(1):274-83.

29. Retico A. Computer-aided detection for pulmonary nodule identification: improving the radiologist's performance? Imaging Med. 2013; 5:249-63.

30. Beigelman-Aubry C, Raffy P, Yang W, Castellino RA, Grenier PA. Computed-aided detection of solid lung nodules on followup MDCTscreening: evaluation of detection, tracking, and reading time. AJR Am J Roentgenol. 2007; 189:948-55.

31. Godoy MCB, Kim TJ, White CS, Bogoni L, de Groot P, Florin C, et al. Benefit of Computer-Aided Detection Analysis for the Detection of Subsolid and Solid Lung Nodules on Thin- and Thick-Section CT. AJR Am J Roentgenol. 2013; 200:74-83.

32. Wahidi MM, Govert JA, Goudar RK, Gould MK, McCrory DC. Evidence for the treatment of patients with pulmonary nodules: when is it lung cancer? ACCP evidence-based clinical practice guidelines (2nd edition). Chest (3 Suppl). 2007; 94S-107S.

33. Scholten ET, Horeweg N, de Koning HJ, Vliegenthart R, Oudkerk M, Mali WP, et al. Computed tomographic characteristics of interval and post screen carcinomas in lung cancer screening. Eur Radiol. 2015; 25:81-8.

34. Henschke CI, Yankelevitz DF, Yip R, Reeves AP, Farooqi A, Xu D, et al. Lung cancers diagnosed at annual CT screening: Volume doubling times. Radiology. 2012; 263:578-83.

35. Henschke CI, Yankelevitz DF, Mirtcheva R, McGuinness G, McCauley D, Miettinen OS. CT screening for lung cancer: frequency and significance of part-solid and nonsolid nodules. AJR Am. J. Roentgenol. 2002; 178(5):1053-7.

36. Armato SG 3rd, McNitt-Gray MF, Reeves AP, Meyer CR, McLennan G, Aberle DR, et al. The Lung Image Database Consortium (LIDC): an evaluation of radiologist variability in the identification of lung nodules on CT scans. Acad Radiol. 2007; 14:1409-21.

**Table and Figure legends**

Fig. 1: Computer-Aided Diagnosis System

Fig. 2: Flowchart of the study population

Fig. 3: Examples of typical CAD findings
(A) Solid nodule, (B) Part-solid nodule, (C) Calcified nodule, (D) GGN

Fig. 4: Examples of false-positive findings by CAD
(A) Incorrect detection of a blood vessel as a calcified nodule. (B) Detection of ground-glass opacity.
(C)(D) Detection of protruding pleura.

Fig. 5: Examples of CAD detection of GGNs overlooked by readers

(A) GGN with a major axis of 10 mm in right lung S3. (B) GGN with a major axis of 12 mm in left lung S10.

In both cases, differential diagnosis among primary lung cancer in situ, atypical adenomatous hyperplasia, and adenomatous. hyperplasia is required.

Table 1: The number of nodules according to the major axis and nodule types, the number of cases according to the number of nodules

Table 2: Per-nodule sensitivity of CAD alone and radiologists' reading without and with CAD, according to the major axis and nodule type. P-values are for the comparisons between total radiologists' reading without and with CAD.

Table 3: Per-nodule sensitivity, PPV, F1-score of CAD alone, and radiologists' reading with/without CAD. P-values are for the comparisons between total radiologists' reading without and with CAD.

Table 4: Per-patient sensitivity and specificity. P-values are for the comparisons between total radiologists' reading without and with CAD.
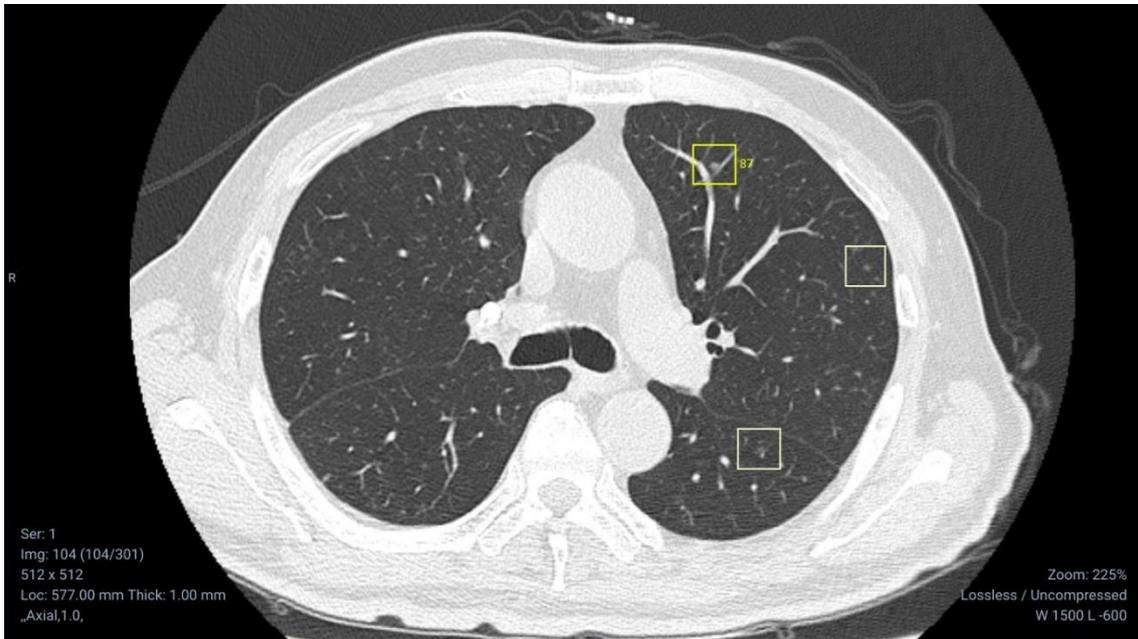
Table 5: Reading time of Reader A and Reader B.
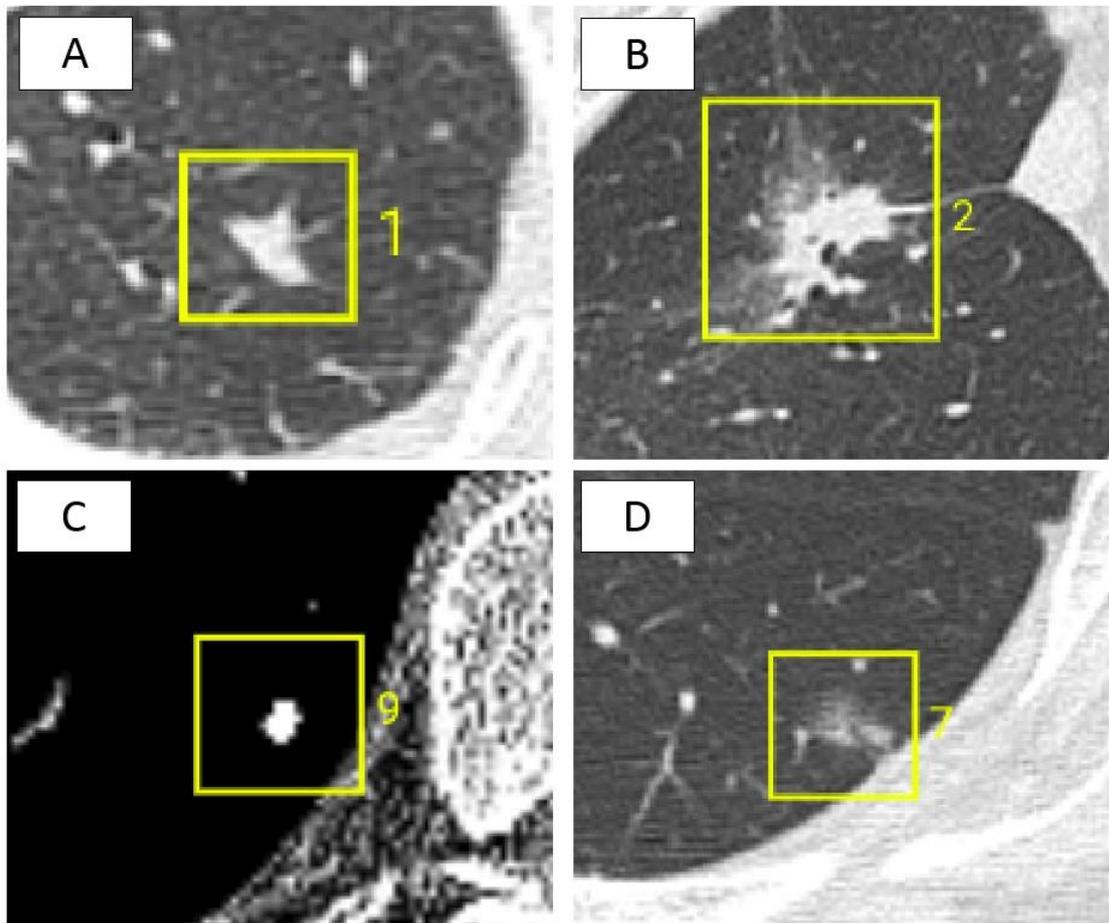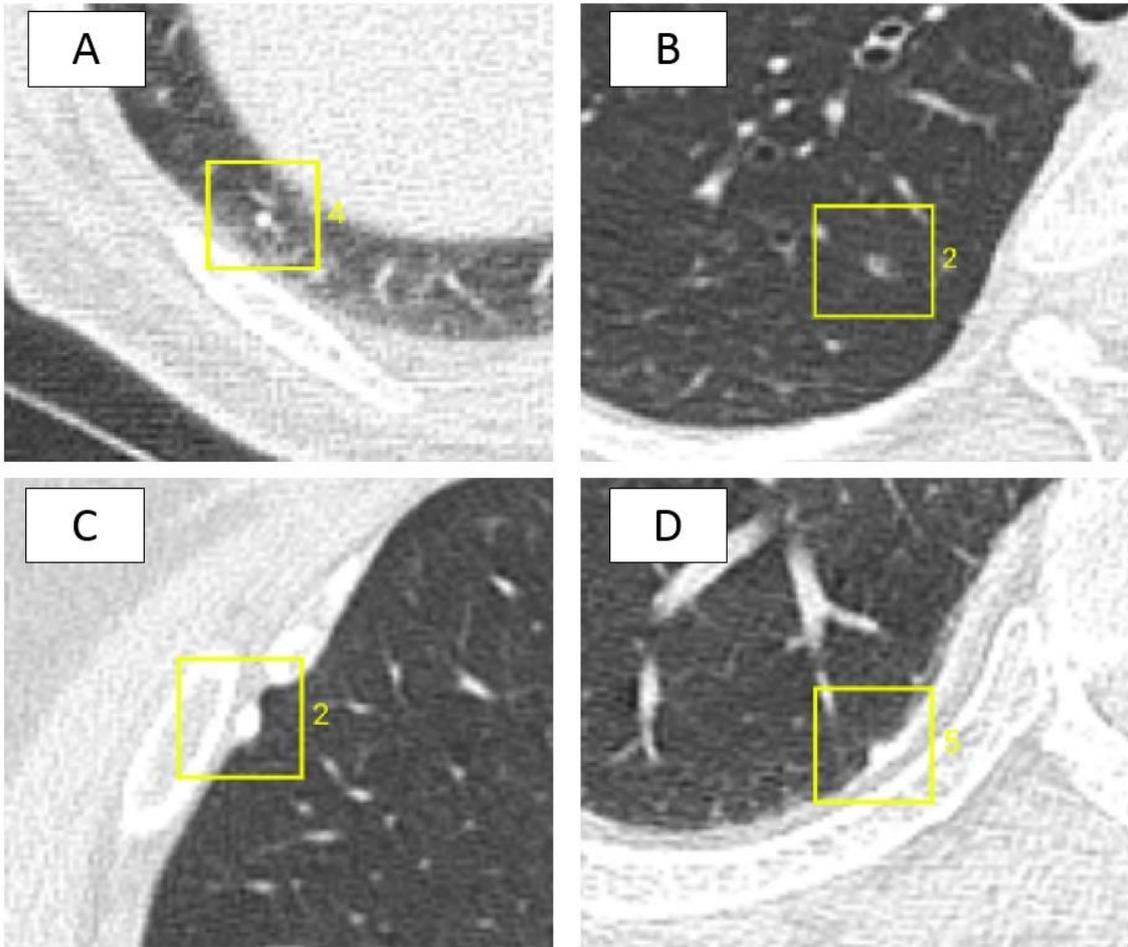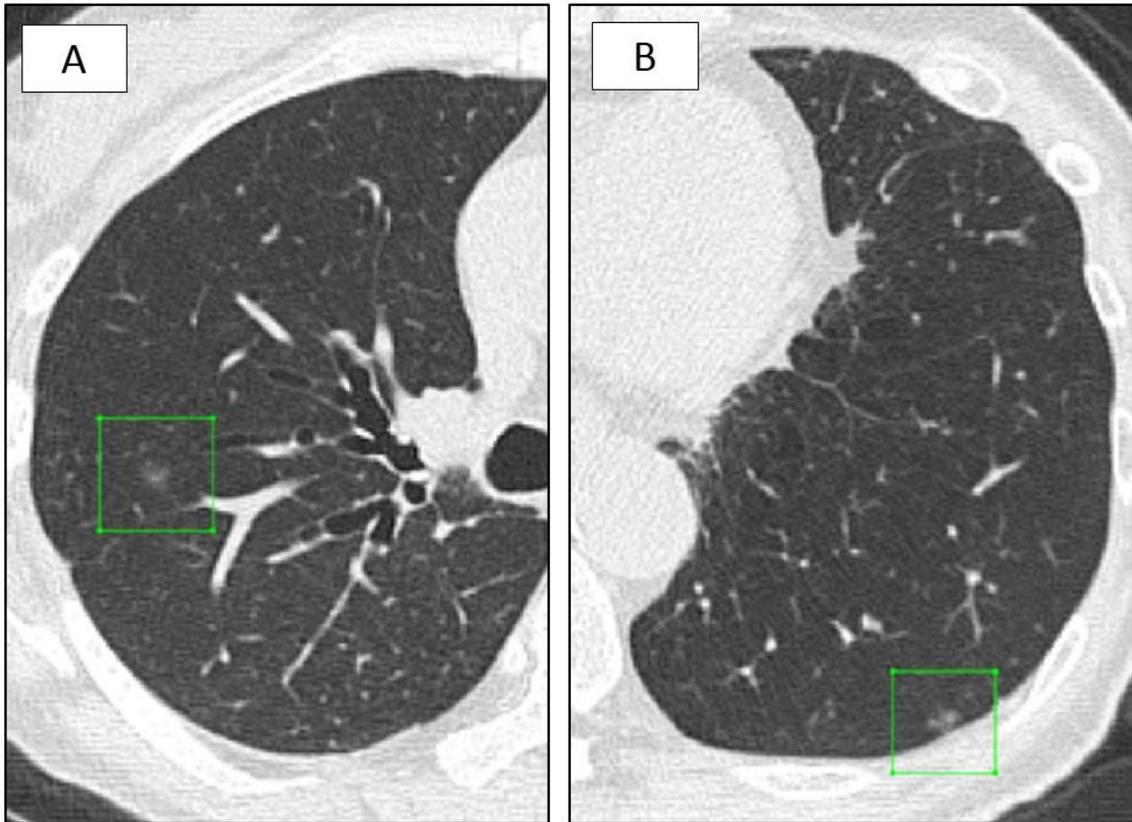
Fig. 1:



Fig. 2:

Fig. 3:

Fig. 4:

Fig. 5:

Table 1:

| Major axis | Number of nodules | Type of nodule | Number of nodules | Number of nodules | Number of cases |
|---|---|---|---|---|---|
| 3mm - 6mm | 532 | Solid nodule | 518 | 0 | 6 |
| 6mm - 10mm | 144 | Part-solid nodule | 65 | 1 | 11 |
| 10mm - 15mm | 46 | Calcified nodule | 74 | 2 - 3 | 27 |
| 15mm - 20mm | 14 | GGN | 86 | 4 - 9 | 46 |
| 20mm - | 7 | | | 10 - | 27 |
| Total | 743 | Total | 743 | Total | 117 |

Table 2:

| | CAD-alone | | Reading without CAD (Reader A + Reader B) | | Reading with CAD (Reader A + Reader B) | | |
|---|---|---|---|---|---|---|---|
| Major axis | Sensitivity | 95%CI | Sensitivity | 95%CI | Sensitivity | 95%CI | $p$-value |
| 3 – 6 mm | 73.9% (393/532) | 0.699-0.776 | 13.7% (146/1064) | 0.117-0.159 | 32.4% (345/1064) | 0.296-0.353 | <0.01 |
| 6 – 10 mm | 60.4% (87/144) | 0.519-0.685 | 33.3% (96/288) | 0.279-0.391 | 47.6% (137/288) | 0.417-0.535 | <0.01 |
| 10 – 15 mm | 58.7% (27/46) | 0.432-0.730 | 51.1% (47/92) | 0.404-0.617 | 58.7% (54/92) | 0.479-0.689 | 0.07 |
| 15 – 20 mm | 64.3% (9/14) | 0.351-0.872 | 46.4% (13/28) | 0.275-0.661 | 60.7% (17/28) | 0.406-0.785 | 0.2 |
| 20 mm – | 85.7% (6/7) | 0.421-0.996 | 57.1% (8/14) | 0.289-0.823 | 78.6% (11/14) | 0.492-0.953 | 0.2 |
| Nodule type | | | | | | | |
| Solid nodule | 68.1% (353/518) | 0.639-0.721 | 18.6% (193/1036) | 0.163-0.211 | 32.6% (338/1036) | 0.298-0.356 | <0.01 |
| Part-solid | 70.8% (46/65) | 0.582-0.814 | 31.5% (41/130) | 0.237-0.403 | 58.5% (76/130) | 0.495-0.670 | <0.01 |
| Calcified nodule | 82.4% (61/74) | 0.718-0.903 | 30.4% (45/148) | 0.231-0.385 | 54.7% (81/148) | 0.463-0.629 | <0.01 |
| GGN | 72.1% (62/86) | 0.614-0.812 | 18.0% (31/172) | 0.126-0.246 | 40.1% (69/172) | 0.327-0.479 | <0.01 |
| Total | 70.3% (522/743) | 0.668-0.735 | 20.9% (310/1486) | 0.188-0.230 | 38.0% (564/1486) | 0.355-0.405 | <0.01 |

Table 3:

| | CAD alone | | Reader A | | | | Reader B | | | | Total (Reader A + Reader B) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 95%CI | Without CAD | 95%CI | With CAD | 95%CI | Without CAD | 95%CI | With CAD | 95%CI | Without CAD | 95%CI | With CAD | 95%CI | p-value |
| Sensitivity | 70.3% (522/743) | 0.668-0.735 | 22.7% (169/743) | 0.198-0.259 | 35.3% (262/743) | 0.318-0.388 | 19% (141/743) | 0.162-0.220 | 40.6% (302/743) | 0.371-0.443 | 20.9% (310/1486) | 0.188-0.230 | 38% (564/1486) | 0.355-0.405 | <0.01 |
| PPV | 57.9% (522/902) | 0.546-0.611 | 69% (169/245) | 0.628-0.747 | 69.5% (262/377) | 0.646-0.741 | 72.3% (141/195) | 0.655-0.785 | 56.4% (302/535) | 0.521-0.607 | 70.5% (310/440) | 0.660-0.747 | 61.8% (564/912) | 0.586-0.650 | - |
| F1-score | 0.635 | - | 0.342 | - | 0.468 | - | 0.301 | - | 0.473 | - | 0.322 | - | 0.470 | - | - |

Table 4:

| | CAD alone | | Reader A | | | | Reader B | | | | Total (Reader A + Reader B) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 95%CI | Without CAD | 95%CI | With CAD | 95%CI | Without CAD | 95%CI | With CAD | 95%CI | Without CAD | 95%CI | With CAD | 95%CI | p-value |
| Sensitivity | 95.5% (106/111) | 0.898-0.985 | 68.5% (76/111) | 0.590-0.77 | 84.7% (94/111) | 0.766-0.908 | 67.6% (75/111) | 0.580-0.761 | 85.1% (95/111) | 0.776-0.915 | 68% (151/222) | 0.614-0.741 | 85.1% (189/222) | 0.798-0.895 | <0.01 |
| Specificity | 83.3% (5/6) | 0.359-0.996 | 83.3% (5/6) | 0.359-0.996 | 83.3% (5/6) | 0.359-0.996 | 100% (6/6) | 0.421-1.000 | 83.3% (5/6) | 0.359-0.996 | 91.7% (11/12) | 0.615-0.998 | 83.3% (10/12) | 0.516-0.979 | - |

Table 5:

| | Without CAD (min) | With CAD (min) | Decrease ratio |
|---|---|---|---|
| Reader A | 326 | 292 | 10.4% |
| Reader B | 420 | 370 | 11.9% |
| Mean | 373 | 331 | 11.3% |