Evaluation of deep learning software tool for CT based lung nodule segmentation

Poster No.:	C-3686
Congress:	ECR 2019
Type:	Scientific Exhibit
Authors:	J. Murchison, G. Ritchie, D. Senyszak, E. J. R. Van Beek; Edinburgh/UK
Keywords:	Artificial Intelligence, Lung, Oncology, CAD, CT, Neural networks, Cancer
DOI:	10.26044/ecr2019/C-3686

Aims and objectives

Despite lung cancer preventive strategies, lung cancer remains the third highest cause of cancers worldwide [1] with a rising incidence of the disease [2]. In addition, lung cancer is the commonest cause of cancer related deaths [3] accounting for around 1.7 million annual deaths globally. This high mortality rate is in part due to the fact that lung cancer is often diagnosed at an advanced stage of disease. The results of the National Lung Screening Trial (NLST) showed that early detection of lung cancer is possible using low-dose CT in a high-risk population and that this is associated with a decrease in both lung cancer related and overall mortality. This has led to the approval of lung cancer screening in the USA [4]. The benefits of lung cancer screening and early detection of lung cancer are also supported by the findings of the Benelux NELSON trial.

Lung cancer is ideally diagnosed by histopathological confirmation on a tissue sample. However, the diagnostic process usually begins with detection of pulmonary nodules or masses, usually through medical imaging. Pulmonary nodules are very common and most are benign, however benign and malignant nodules can have identical appearances, so all should be flagged up as potential cancers. The biggest challenges when it comes to pulmonary nodule detection on CT are acceptable sensitivity levels and reading times. The importance of high sensitivity for pulmonary nodule detection is underscored by the fact that many failures in lung cancer diagnosis are due to errors of detection rather than interpretation [5,6]. Over the last two decades a substantial number of studies [7,8,9] have evaluated the performance of (sub-specialist) radiologists for the specific task of detecting pulmonary nodules and have shown that there is room for improvement.

In addition, pulmonary nodule guidelines recommend the use of different cut-off levels for nodule size and/or volume and volume doubling time as metrics to assess nodule size and growth [10-16]. Most recently, there seems to be consensus that semi-automated volume assessment gives the most robust assessment for lung nodules and is most helpful for determining growth during follow up. [14-16] Apart from the measurement of size and/or volume, another important parameter to consider is the makeup of pulmonary nodules (solid vs sub-solid), as their malignant potential is significantly different. [17] In spite of both these size/volume and composition management guidelines, many hospitals currently do not have the the tools to perform these measurements in a timely and accurate manner, both due to a lack of software and due to a lack of suitably trained radiologists. Therefore, software aided detection and classification of lung nodules would be a welcome addition to the radiologist's diagnostic arsenal and could facilitate the roll-out of CT lung cancer screening as has been advocated. [18].

The objective of this study was to evaluate the clinical performance of a Deep Learning computer assisted diagnosis system (CAD) for lung nodule segmentation on CT Chest.

Methods and materials

Patient population: A total of 349 chest CT examinations from 324 unique subjects were retrospectively selected from the NHS Lothian database. Eligibility of CT scans for each of the 5 group was determined using information from the radiology reports with cross referencing to the electronic health records as appropriate. Subjects for the first two groups were selected to mimic a lung cancer screening population. Inclusion criteria were subjects between 50-74 years of age, current smokers or those with a smoking history and/or reported to have radiological evidence of pulmonary emphysema were found eligible for the first two groups. Group 1 consists of 181 CT scans which were clinically reported as being free from pulmonary nodules and group 2 consists of 100 CT scans which were reported to have at least 1 and no more than 10 pulmonary nodule(s). Group 3 consists of 25 CT scans which were followed up for the presence of a pulmonary nodule, group 4 consists of the follow-up CT scans of group 3. Finally, group 5 consists of 18 CT scans with part-solid and/or ground-glass nodule(s) described in the original radiology report. Group 5 was intended to increase the overall number of sub-solid nodules. Specific exclusion criteria were slice thickness >3mm and the presence of diffuse pulmonary disease in the radiology report and/or the CT images, with widespread abnormalities such as interstitial lung disease, which is very likely to lead to significant symptoms and therefore didn't correspond with an asymptomatic screening subject.

CT protocol: This was to be a "real life" group of subjects, so irrespective of type of CT scanner used, presence or absence of intravenous contrast or actual protocol applied. Patients were scanned with Aquilion (n=330), Aquilion-CX (n=2), and Aquilion ONE (n=1) CT scanners from Canon Medical Systems (formerly Toshiba Medical Systems), Otawara, Japan and LightSpeed (n=2), LightSpeed Plus (n=2) CT scanners from General Electric Medical Systems, Waukesha, United States. Intravenous contrast medium was used in 22 CT scans. Image orientation direction were Feet First-Supine (FFS, n=277), Head First-Supine (HFS, n=44), Feet First-Prone (FFP, n=9) and Head First-Prone (HFP, n=7). The mean tube peak potential energies used for scan acquisition was 120 kVp, (median: 120 kVp, range: 120-140 kVp). The average tube current was 243 mAs (median: 232 mAs, range: 80-491 mAs) and the average CTDIvol was 14.0 mGy (median: 14.8 mGy; range: 2.9-29.7). Data were reconstructed at a mean slice thickness of 1.0 mm (median: 1.0mm, range 1.0-2.5mm). The following reconstruction kernels were used for CT scans from Canon Medical Systems FC03 (n=120), FC07 (n=99), FC08 (n=4), FC10 (n=3), FC12 (n=7), FC30 (n=1), FC51 (n=99) and LUNG (n=3), STANDARD (n=1) for CT scans from GE Medical Systems. All CT scans were reconstructed using filtered back-projection.

Nodule definition: The Fleischner Society's definition for pulmonary nodules was broadly used during this study. However the term "pulmonary nodule," was deliberately not firmly defined since the notion of nodule may not represent a single entity capable of verbal

definition [12] and therefore the interpretation of the "noduleness" of a lesion was left at the discretion of the readers, with the proviso that the largest axial diameter was between #3mm and #30mm. Nodules with largest axial diameter between #5mm (or a volume of #80mm3) and #30mm were called "actionable nodules".

CAD software: The CAD software evaluated in this study was Veye Chest version 2.0 (Aidence B.V., Amsterdam, the Netherlands).

Image annotation: A two-phase process was developed for the asynchronous interpretation by a panel of three thoracic radiologists with at least 9 years experience in reading Chest CT scans, JM, GR and EB, expert readers 1, 2 and 3, respectively. Prior to the start of the study each reader received training on the annotation tasks and how to use the annotation tool. A comprehensive set of written instructions was available during the entire annotation process.

In summary, the initial "blinded" phase required readers 1 and 2 to independently perform a free search on all CT scans on a radiology reporting workstation. In half of the CT scans, which were selected at random, the detection results of CAD were made available. The study design ensured that each CT scan was reviewed twice, once by each reader, once by one reader with the results of CAD (AIDED) and once by the other reader without (UNAIDED). Readers were asked to identify all lesions which they considered to be a pulmonary nodule without clear benign morphological characteristics (i.e. calcified nodules). They could mark a pulmonary nodule by adding a manual annotation or classify a CAD prompt as either a true positive or false positive. They were required to register all nodules that were present on CT scans from both groups 3 and 4, where possible. Finally, the readers also classified all false positive prompts in three different groups: micro-nodules (largest axial diameter <3mm), masses (largest axial diameter >30mm), benign nodules (benign calcification pattern or clear benign perifissural appearance), non-nodules (any finding that could not be classified in any of the other sub-groups). Subsequently, non-nodules were further classified as: pleural plaque, scar tissue, atelectasis, fibrosis, fissure thickening, pleural fluid, pleural thickening, intrapulmonary vessels, consolidations, outside of lung tissue, or other (free format). After completing all the readings on the workstations the readers subsequently reviewed their own previously identified nodules on a tablet (iPad Pro). The reader was asked to determine the composition (solid or sub-solid) of the nodule and subsequently segment the nodule on every slice by delineating the border using a stylus (Apple Pencil). After the blinded phase was completed the results from readers 1 and 2 were evaluated for the presence of any discrepancies. Discrepancies were defined as a difference between the results in terms of: location (3D dice coefficient of 0); composition; segmentation (3D dice coefficient < -1 standard deviation of the mean 3D dice coefficient) and nodule registration. The second "unblinded" phase required reader 3 to adjudicate all discrepancies from the blinded phase without the results of CAD, free search was not allowed. The review was performed using the same materials used in the blinded phase.

Reader 3 created a third independent reading for each nodule that had a discrepancy for at least one characteristic.

Reference standard: All segmentations of a nodule from groups 1-3 and 5 were retained.

Data analysis: The segmentation accuracy of readers was calculated as the dice coefficient between each reader's segmentation and the segmentations of the other readers and subsequently averaged (inter-reader dice coefficient). The segmentation accuracy of CAD alone was calculated as the dice coefficient between each CAD segmentation and each individual reader segmentation and subsequently averaged. A dice coefficient score of 1.0 is considered a perfect overlap. In addition, the inter-reader mean diametric and volumetric discrepancy was calculated using the largest axial diameter and volume from each segmentation of each reader's segmentation and compared to those from the other readers, this was also calculated for CAD alone compared to the other readers.

Statistical Analysis: One-tailed Welch's t-test was used to accept the hypothesis that the mean CAD dice score is higher to the mean inter-reader dice score ($p < 0.05$).

Images for this section:

Fig. 1: Figure 1: Software reported lung nodule (axial and MIP), as presented for verification in 50% of cases to thoracic radiologists.

Results

A total of 337 CT scans from 314 subjects (173 women, 164 men) with a total of 470 pulmonary nodules (largest axial diameter between #3mm and #30mm) were included in this study. The mean age of all the subjects was 63 ± 7 years (range 32-88 years). Details regarding the number of CT scans and nodules per group are described in table 1. The mean largest axial diameter of all nodules in groups 1 to 5 was 7.68 ± 3.50 mm (range: 3.42 - 28.45 mm) and the mean volume was 198 ± 333 mm3 (range: 21 - 2797 mm3).

The CAD software was able to successfully segment 95% of the total 428 nodules between #3mm and #30mm in groups 1-3 and 5. The average inter-reader dice coefficient was 0.83 (95% CI: 0.39, 0.96) which was 0.86 (95% CI: 0.51, 0.95) for CAD alone ($p<0.01$). The inter-reader geometric mean diameter discrepancy was 1.15 (95% CI: 1.00, 1.58) which was 1.17 (95% CI: 1.01, 1.69) for CAD alone. The inter-reader geometric mean volumetric discrepancy was 1.39 (95% CI: 1.01, 3.19) which was 1.38 (95% CI: 1.01, 3.38) for CAD alone.

Images for this section:

Groups	Number of unique subjects	Number CT scans	Total number of nodules between	
			≥3 and <5mm	≥5mm / ≥80mm^3 and <30mm
1	178	178	19	71
2	95	95	34	198
3	23	23	0	68
4		23	6	36
5	18	18	2	36
TOTAL	**314**	**337**	**61**	**409**

Table 1. Number of unique subjects, scans and nodules in each of the 5 groups.

Table 1: Table 1: demographics of the groups studied.

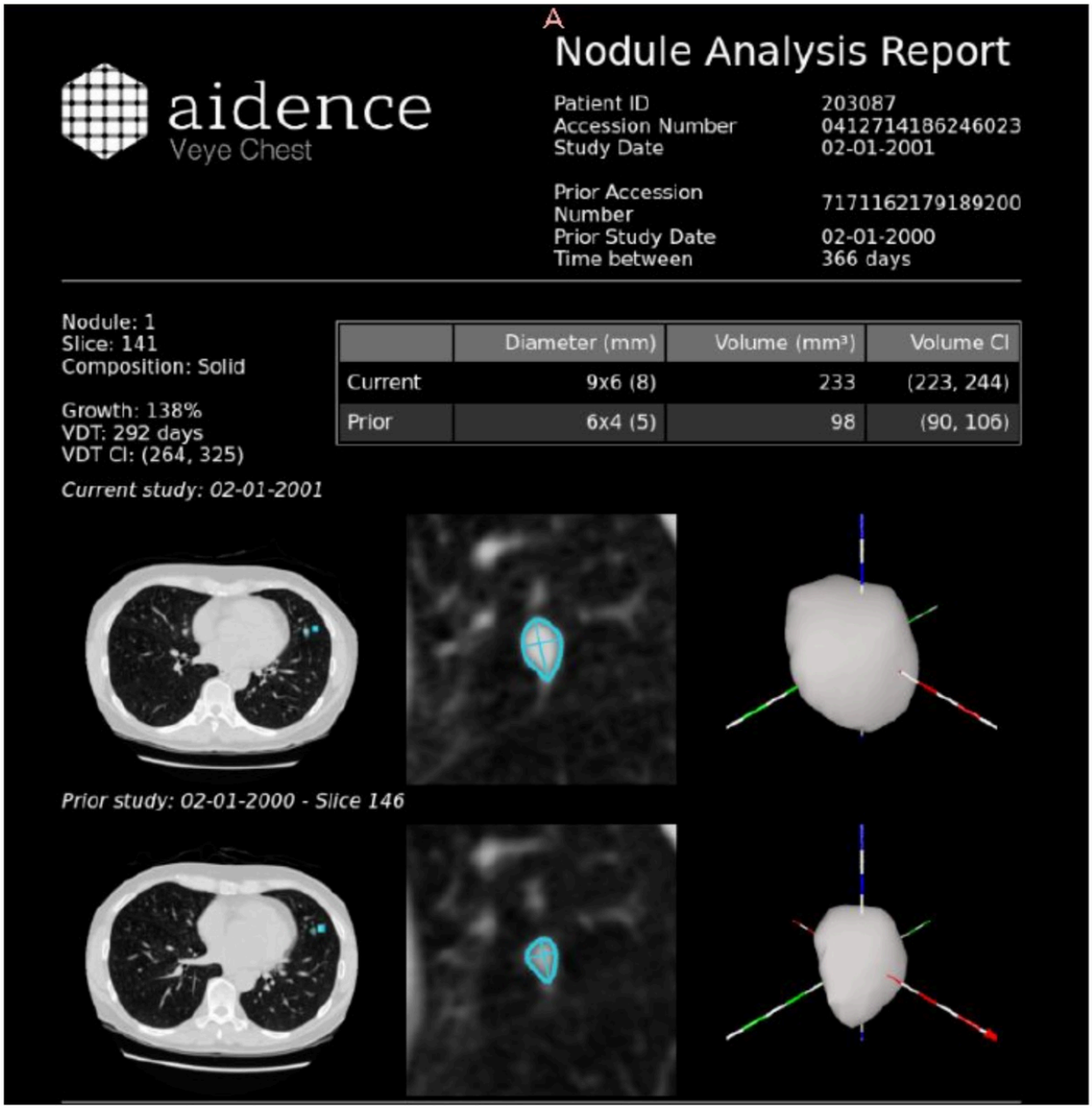

Fig. 2: Figure 2: Nodule detected and segmented, with follow-up demonstrating volume doubling times assessment.

Conclusion

The performance of the CAD software for segmenting pulmonary nodules on Chest CT is comparable to that of experienced thoracic radiologists. These results show that the tool has the potential to assist radiologists at pulmonary nodule management, although visual verification of the segmentation is still advised at present.

Personal information

Address for correspondence:

Professor John T Murchison, Consultant Radiologist, Royal Infirmary of Edinburgh, Little France Crescent, Edinburgh, UK. Email: john.murchison@luht.scot.nhs.uk

Declaration of interest:

This study was funded by NHS England via the SBRI Phase 1 grant for earlier detection and treatment of cancer. Aidence BV received the grant from SBRI and commissioned the University of Edinburgh and Royal Infirmary of Edinburgh to perform the study. Prof. J.T. Murchison, Dr. G. Ritchie and Prof J.R. van Beek received a consulting fee for reading and annotating the studies.

Prof. Van Beek is a member of the Advisory Board of Aidence.

Prof. Murchison, Dr. Ritchie and Mr. Senyszak declare no interest.

References

1. International Agency for Research on Cancer (IARC). GLOBOCAN 2018.

2. Jemal A, Bray F, Center MM, Ferlay J, Ward E and Forman D (2011) Global cancer statistics. CA Cancer J Clin 61:69-90

3. International Agency for Research on Cancer (IARC). GLOBOCAN Cancer Fact Sheets: Lung Cancer. International Agency for Research on Cancer (IARC); 2017, http://gco.iarc.fr/today/data/factsheets/cancers/15-Lung-fact-sheet.pdf. Accessed 2018.

4. Centers for Medicare & Medicaid Services (CMS).(2015) Decision memo for screening for lung cancer with low dose computed tomography (LDCT) (CAG-00439N). CMS website. www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=274. Accessed February 19, 2017

5. Kakinuma R, Ohmatsu H, Kaneko M et al. (1999) Detection failures in spiral CT screening for lung cancer: analysis of CT findings. Radiology 212: 61-6

6. White CS, Romney BM, Mason AC, Austin JH, Miller BH, Protopapas Z (1996) Primary carcinoma of the lung overlooked at CT: analysis of findings in 14 patients. Radiology 199:109-15

7. Kakinuma R, Ashizawa K, Kobayashi T et al (2012) Comparison of sensitivity of lung nodule detection between radiologists and technologists on low-dose CT lung cancer screening images. Br J Radiol 85:e603-8

8 Roos JE, Paik D, Olsen D et al. (2010) Computer-aided detection (CAD) of lung nodules in CT scans: radiologist performance and reading time with incremental CAD assistance. Eur Radiol 20: 549-57

9. Nair A, Gartland N, Barton B et al (2016) Comparing the performance of trained radiographers against experienced radiologists in the UK lung cancer screening (UKLS) trial. Br J Radiol 89:20160301

10. Church TR, Black WC, Aberle D, et al. (2013) National Lung Screening Trial. New Engl J Med 368:1980-91.

11. Callister MEJ, Baldwin DR, Akram AR, et al.(2015) The BTS guideline for the investigation and management of pulmonary nodules. Thorax 2015;70 (suppl 2).

12. MacMahon H, Naidich DP, Goo JM, et al. Guidelines for management of incidental pulmonary nodules detected on CT imaging: From the Fleischner Society 2017. Radiology 284:228-43.

13. McKee BJ, Regis SM, McKee AB, Flacke S, Wald C. (2015) Performance of ACR lung-RADS in a clinical CT lung screening program. J Am Coll Radiol 12:273-6.

14. Bankier AA, MacMahon H, Goo JM, Rubin GD, Schaefer-Prokop CM, Naidich DP. (2017) Recommendations for measuring pulmonary nodules at CT: A statement from the Fleischner Society. Radiology 285:584-600.

15. Deveraj A, van Ginneken B, Nair A, Baldwin D. (2017) Use of volumetry for lung nodule management: theory and practice. Radiology 284:630-44.

16. Heuvelmans MA, Walter JE, Vliegenthart R, et al. (2018) Disagreement of diameter and volume measurements for pulmonary nodule size estimation in CT lung cancer screening. Thorax 73:779-81.

17. Oudkerk M, Deveraj A, Vliegenthart R, et al. (2017). European position statement on lung cancer screening. Lancet Oncol 18:e754-66.

18. Naidich DP, Bankier AA, MacMahon H, et al. (2013) Recommendations for the management of subsolid pulmonary nodules detected at CT: A statement from the Fleischner Society. Radiology 266:304-17.