Impact of Concurrent Use of Artificial Intelligence Tools on Radiologists Reading Time: A Prospective Feasibility Study

Felix C. Müller, FM, Henriette Raaschou, HR, Naurien Akhtar, NA, Mathias Brejnebol, MBr, Lene Collatz, LC, Michael Brun Andersen, MBA

Rational and Objectives: This study investigated how an AI tool impacted radiologists reading time for non-contrast chest CT exams.

Materials and Methods: An AI tool was implemented into the PACS reading workflow of non-contrast chest CT exams between April and May 2020. The reading time was recorded for one CONSULTANT RADIOLOGIST and one RADIOLOGY RESIDENT by an external observer. After each case radiologists answered questions regarding additional findings and perceived case overview. Reading times were recorded for 25 cases without and 20 cases with AI tool assistance for each reader. Differences in reading time with and without the AI tool were assessed using Welch’s t-test for non-inferiority with non-inferiority limits defined as 100 seconds for the consultant and 200 seconds for the resident.

Results: The mean reading time for the radiology resident was not significantly affected by the AI tool (without AI 370s vs with AI 437s; +67s 95% CI - -28s to +163s, p = 0.70). The AI tool led to additional actionable findings in 5/40 (12.5%) studies and better overview in 18/20 (90%) of studies for the resident.

Conclusion: A PACS based implementation of an AI tool for concurrent reading of chest CT exams did not increase reading time with additional actionable findings made as well as a perceived better case overview for the radiology resident.

Key Words: Work Performance; Image Interpretation; Computer-Assisted; Multidetector Computed Tomography; Radiology.

© 2021 The Association of University Radiologists. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

INTRODUCTION

The number of commercially available Artificial Intelligence (AI) tools in radiology is rapidly rising. AI tools can assist the radiologist in solving interpretation tasks such as lesion detection, automatic measurement, and decision support. And AI tools have been shown to potentially reduce the reading time for radiologists, especially in the tasks of detecting breast-, lung- and colon-cancer (1–5). But, the AI tools are often retrospectively investigated in research environments, with no interaction between the AI tool, and the clinical reading environment. In addition, the increasing number of AI tools will lead to multiple AI tools simultaneously affecting radiologists in clinical practice. We therefore wanted to study the influence of multiple AI functionalities on radiologist reading time in a clinical setting.

In this manuscript the term AI tool refers to programs or algorithms, which automatically and independently assess imaging studies and inform the radiologists about findings. One of these AI tools, recently introduced at our institution, allows simultaneous reporting of multiple findings in non-contrast low-dose CT scans of the chest. At our institution, this type of scan is primarily used for follow-up of pulmonary nodules identified on previous CT scans or radiographs of the chest.

AI tools can reduce reading times for CT scans of the chest — but this depends on the implementation of the algorithm. Matsumoto et al found that when AI tools for the detection of CT nodules were used concomitantly to normal reading then radiologist reading times for both residents and senior radiologists were not substantially affected (6). But when using the AI tool for second reading (i.e. the radiologist first assesses the scan without an AI tool), the reading time increased. (5–7) A similar effect has been shown for AI tools for pulmonary embolism detection, where second reading increased the reading time for CT angiography studies by 22 seconds, while concurrent reading reduced the reading time by 12%–17%. (8,9) Recently two smaller studies of concomitant use of AI tools for the detection of corona virus 19 disease also found reduction in average
reading times for chest CT studies (10,11). The intended use of the algorithm as described by the vendor might restrict concurrent use of some AI tools.

The clinical environment also affects radiologists’ reading time. Some factors are hardly surprising, such as the time of day, the experience of the radiologist and the number of interruptions — others more puzzling, such as the size and luminescence of the screen, the use of key images in reporting, and the image order in the hanging protocol. (12-18)

Given this complexity, it does not seem unreasonable to assume that the reading times when using AI can be differently affected in a clinical environment and a research environment. This seems supported by studies which find that CAD findings affect the reading times of individual radiologists differently. (3,9)

The primary aim of this feasibility study was to investigate if radiologists’ reading times, when using multiple AI tools concurrently, was non-inferior to reading times without AI tool, in non-contrast low-dose chest CT exams. Secondary aims were to assess if the AI tool lead to additional findings or affected the readers’ diagnostic confidence.

MATERIALS AND METHODS

This was a prospective interventional feasibility study with commercially available and regulatory approved software. Ethical approval and need for patient consent were waved due to the nature of the study. The manuscript was prepared using the SQUIRE-2 statement. (19)

AI Tool

The AI tool (AI Chest Companion, Siemens Healthineers, Erlangen, Germany) for the automatic analysis of CT studies of the thorax without contrast was installed at our institution. Non-contrast CT studies of the thorax were automatically and always forwarded to a local DICOM node, which de-identified scans and forwarded them to the cloud-based AI tool. Cases were then assessed by the AI tool and results re-identified and stored in our institutional picture archiving and communications system (PACS) alongside original scan series (Fig 1). Results of the AI tool were not reviewed by investigators or radiographers prior to display to the radiologists in the PACS viewer. The AI tool performed 9 tasks and provided the radiologists with the results and measurements outlined in Table 1.

The AI tool generated an overview image which displayed the assessed organs (lung, aorta, heart, and spine) color coded based on presence of findings. Five additional series were always sent to PACS per case: 1) single overview image 2) a single coronal curved planar image of the straightened vertebral column with height and bone marrow density overlay 3) a transverse lung series with areas below -950 HU, lung lobes, and nodules marked 4) cross sectional images of the aorta at nine measurement locations 5) a transverse series of the heart with coronary calcification. An additional series was only sent when nodules were found in the lung and consisted of a 3D radial range at eight angles with the detected nodule marked. (Fig 2)

Pre-Study Interview and Baseline Assessment

Prior to the initiation of the study NA performed structured interviews with two specialists in thoracic radiology and one radiology resident. The purpose of this interview was to understand radiologists’ expectations and assumptions of the AI tool and its presumed influence on their workflow. Results of the pre-study interviews resulted in selection of quantitative and qualitative outcome measures for this study.

Radiological Case Observations

We observed readings of 90 consecutive low dose CT scans of the thorax without intravenous contrast media at the department of radiology at Herlev and Gentofte Hospital in Denmark. The cases were chosen between April and May 2020 and were based on the availability on days on which data collection for the study was possible. Cases were not pre-screened and no cases were excluded.

Radiologists

We observed two radiologists (a thoracic radiology consultant with 13 years’ experience and a second-year radiology resident [PGY-2 doing rotation in thoracic radiology) during their routine clinical case reading. Radiologists read cases, including prior exams, in their normal working environment at their own discretion with use of any software they liked. Radiologists were free to view the results of the AI tool before, concurrent with or after viewing conventional CT reconstructions.

Baseline

For a baseline assessment we recorded the reading time (time from opening the case to submitting the report) of radiologists without deployment of the AI tool in 25 unique cases each.
Post Intervention Measurements

After the AI tool had been installed, radiologists had time to familiarize themselves with the AI tool. At their own discretion they indicated that they felt proficient using the additional series. Then reading time of the radiologist was measured in 20 unique cases each.

Data Collection

The reading time for each radiologist was measured at baseline and post intervention using a stopwatch by a single observer (NA). The timer was stopped when the radiologist was interrupted by other tasks such as answering questions from colleagues or telephone calls and resumed when reading continued.

Immediately after each case the observer posed five questions to the radiologist. (Original questions in Danish) (Do you feel more confident when using the AI tool [Yes / No / Equivocal]? Has the AI tool provided additional information in this case [Yes / No / Equivocal]? Has the AI tool changed the diagnosis in this case [Yes / No / Equivocal]? Do you have a better overview of the case using the AI tool [Yes / No / Equivocal]? Based on your subjective experience, how has the AI tool affected reading time for this case [Increased / Decreased / Unchanged / Equivocal]?). Radiologists were asked to elaborate on the categorical answers.

Reader performance was not objectively assessed as part of this study.

Statistical Analysis and Power Calculation

We previously measured reading times for non-contrast chest CT at 7.03 min (SD 1.8 min) for radiology consultants and 10.68 min (SD 3.56 min) for radiology residents at our

| TABLE 1. AI Tool Features and Results in PACS. |
|-----------------------------------------------|
| Feature | Image Series Sent to PACS |
| Overview | Single image with color coded organs according to findings from the eight other features |
| Segmentation of pulmonary lobes | Axial series with markup on findings and lobes + optional 3d radial range of lungs if nodule detected |
| Quantification of pulmonary areas below -950 HU - a surrogate marker for emphysema | Axial series with marking of coronary calcification |
| Detection and measurement of solid pulmonary nodules | Cross sectional reconstructions at each measurement location |
| Quantification of degree of coronary calcification | Midsagittal curved planar reconstruction of spine with anterior, middle, and posterior height of vertebral corpus overlaid |
| Quantification of cardiac volume | None |
| Cross sectional diameter measurement of aorta at 9 predefined locations | |
| Quantification of vertebral corpus height at anterior, middle and posterior aspects and comparison to vertebra above and below - A surrogate marker for spinal compression fractures | |
| Quantification of mean HU values in the vertebral corpus - a surrogate marker for osteoporosis | |

Figure 2. Result series from the AI tool. Note – Series A-E were available for all cases, while series F was only displayed if nodules were detected. The results of series B-E are summarized in the overview image A.
inclusion. Sample size was calculated with the intent of testing for non-inferiority at a power of 0.9. The clinically relevant limit of non-inferiority for the consultant radiologist was decided to be 100 seconds and 200 seconds for the resident in this feasibility study. This resulted in a required sample size of at least 20 cases in each subgroup for each reader.

Reading times with and without the AI tool were compared using Welch t-test per reader, with \( p \) values < 0.05 indicating a significant difference. Statistical calculations were done using R (version 4.0.1)

**RESULTS**

**Reading Time**

The mean reading time for the radiology resident was 370s (SD 169s; Range: 90-704s) at baseline and 437s (SD 149s; Range: 199-684s) with the AI tool (Fig 3). The difference was not significant at +67s (95% CI -28s to +163s, \( p = 0.16 \)).

The reading time for the radiology consultant was 366s (SD 110s; Range: 171-526s) at baseline and 380s (SD 121s; Range: 245-739s) with the AI tool. The difference was also not significant at +13s (95% CI -57s to 84s, \( p = 0.70 \)).

**Quantitative Outcomes**

The consultant radiologists stated in 5/20 cases that additional findings were made by the AI tool, while the resident noted no additional findings. The AI tool provided a better case overview in 18/20 cases for the resident but only 2/20 cases for the consultant. The perceived reading time was not affected by use of the AI tool in 19/20 cases for the consultant but only unaffected in 10/20 cases for the resident (Table 2 and Fig 4).

**Qualitative Outcomes**

The comments indicated that the AI tool found two cases of aortic ectasia, one case of coronary calcification and two vertebral compression fractures, which were added to the conclusion, and considered a new actionable finding by the consultant radiologist.

There were comments in 21 cases (resident 17 / consultant 4) where the radiologists indicated that the AI tool increased the conceived diagnostic confidence or gave a better overview. In 15 comments (resident 14 / consultant 1) it was indicated that the tool was used after the reading to confirm that there were no additional findings (i.e. the AI tool was used for second reading), in five cases (resident 3 / consultant 2) the AI tool was used to give an overview before reading the case, nine comments (resident 6 / consultant 3) indicated that the color-coded overview image was helpful, six comments (resident 4 / consultant 2) indicated that the 3D pulmonary model was helpful, and 4 other comments.

The perceived increase in reading time was attributed to problems with false positive findings in two cases (resident),

### Table 2. Impact of AI Tool for Low Dose Chest CT Scans on Radiologist Workflow.

| Question                          | Answer | Radiology Resident | Radiology Consultant |
|-----------------------------------|--------|--------------------|----------------------|
| Additional Findings               | Yes    | 0/20 (0%)          | 5/20 (25%)           |
|                                  | No     | 19/20 (95%)        | 15/20 (75%)          |
|                                  | Equivocal | 1/20 (5%)        | 0/20 (0%)            |
| Impression Changed                | Yes    | 0/20 (0%)          | 5/20 (25%)           |
|                                  | No     | 20/20 (100%)       | 15/20 (75%)          |
|                                  | Equivocal | 0/20 (0%)        | 0/20 (0%)            |
| Better Overview                   | Yes    | 18/20 (90%)        | 4/20 (20%)           |
|                                  | No     | 2/20 (10%)         | 15/20 (75%)          |
|                                  | Equivocal | 0/20 (0%)        | 1/20 (5%)            |
| Increased Diagnostic Confidence   | Yes    | 6/20 (30%)         | 2/20 (10%)           |
|                                  | No     | 14/20 (70%)        | 18/20 (90%)          |
|                                  | Equivocal | 0/20 (0%)        | 0/20 (0%)            |
| Perceived effect on reading time  | Decreased | 0/20 (0%)         | 0/20 (0%)            |
|                                  | Unchanged | 10/20 (50%)       | 19/20 (95%)          |
|                                  | Increased | 5/20 (25%)        | 1/20 (5%)            |
|                                  | Equivocal | 5/20 (25%)        | 0/20 (0%)            |
too many series in one case (consultant), PACS problems not attributable to the AI tool in one case (resident), and other reasons in two cases (resident).

**DISCUSSION**

In this prospective feasibility study we assessed the influence of a PACS integrated AI tool on the workflow of two radiologists. Reading times for each radiologist were not significantly affected by the AI tool. The 95% confidence interval for the difference in mean reading time ranged between +163s to -28s for the radiology resident and +84s to -57s for the radiology consultant. These ranges were within the non-inferiority limit defined a priori for this feasibility study. There were additional findings due to the use of the AI tool in 12.5% (5/40) of cases. The AI tool provided a better overview in only 20% (4/20) of cases for the radiology consultant. In contrast, the impact on the radiology resident was more striking: a better overview or perceived increased confidence was noted in 90% (18/20) of cases. Comments suggested that this was primarily due to the presentation of an overview image.

We compared results of our study against studies which we identified using the search string “radiology reading time” on MedLine on the 26.11.2020. We identified a study by Brown et al, which also investigated multiple AI tools on CT scans of the chest, and found a reduction in reading time (20). We found no reduction in reading time and even a noticeable, but non-significant, increase in reading time for the radiology resident. Results from the study by Brown et al. might be fundamentally different from our results. They used a paired design, wherein each case was read twice by the same radiologists, a task which involved the systematic measurement of the aortic diameter in all cases. Our radiologists only measure this diameter when they suspect aortic ectasia.

Radiologists commented in multiple cases on a perceived increase in reading time due to false positive findings. This is in line with findings by Silva et., who concluded that CAD systems for nodule detection complemented radiologist reading but require visual confirmation for the reduction of false positives (21). Reading times for both the radiology resident and the consultant did not differ at baseline 366s vs 370s. This was substantially faster than median 731s reading time for CT scans of the chest without contrast reported by Forsberg et al. (22) This difference is probably due to our inclusion of low-dose protocols only, while Forsberg et al. might have included reading times for HRCT scans.

Overall, results from this feasibility study put forward the theory that PACS based integration of AI tools for the assisted reading of chest CT images are feasible without high reading time impact, while providing diagnostic benefits to the radiologist.

A strength of the study was the assessment of real-world performance, with no cases excluded. The time measurement by an independent observer allowed the radiologist to maintain their focus on their work and allowed a direct interview of the radiologist after each case. This allowed us to compare reading time, impact of the AI tool on the radiologists reading tasks, and qualitative comments by the radiologists for each case. This feasibility study had several limitations. First, the workflow impact of the AI tool was only measured on two radiologists with varying experience. We cannot rule out that the workflow impact might be fundamentally different for other radiologists or affect radiologists differently. This reader dependent effect on reading time as modified by the AI is currently not understood, and might reduce inferences which can be drawn from these study types. However, reading times found in this study corresponded to previously estimated reading times at our institution. Future studies should try to include additional radiologists. It is also likely, that other factors like PACS system familiarity, seniority, and experience with Computer Assisted Detection systems affects reading time and should be individually addressed in future, larger studies. Second, the number of cases assessed was low, with
only 45 cases per reader. We previously assessed reading times at our institution for radiology residents and radiology consultants and discerned that a reading time increase of more than 3.5 minutes and 2 minutes respectively were absolute unacceptable performance decreases preventing any further investigation of the AI tool. Mean and variation in reading times for both radiologists fell within this predefined range, but the confidence intervals of this feasibility study are too wide to rule out a negative impact of the algorithm on reading time. In order to investigate the impact of this AI tool with higher precision a larger sample of cases would be necessary. Third, the consultant radiologist reading the cases was also interviewed prior to the study. His personal preference regarding use of AI tools might have affected the chosen outcome measures. Finally, we only evaluated a single AI tool. The tool was chosen based on availability at our institution, its ability to send results directly to the PACS without confirmation, and the ability of this AI tool to assess multiple disease entities simultaneously. While results are not necessarily generalizable to other AI tools, they do emphasize that a PACS based integration of AI tools is feasible potentially without negative impact on reading time, while still maintaining positive results such as additional findings, and increased diagnostic confidence. The investigated AI tool can also be implemented such that all findings must be preapproved before sending findings to the PACS system, but this option was not investigated.

In conclusion, in this prospective interventional feasibility study we investigated the workflow impact of an AI tool, which assisted radiologists in the reading of low-dose non-contrast CT scans of the chest. The PACS based integration of the AI tool did not significantly increase reading time for the radiology resident or the thoracic radiology consultant with additional actionable findings made as well as a perceived better overview for the radiology resident. We recommend outcomes to be verified in a full-scale study.

ACKNOWLEDGEMENTS

The study was financially supported by the Capital Region of Denmark and the department of radiology at Herlev and Gentofte Hospital, Denmark, who had no influence on the design of the study, data collection, data analysis or manuscript preparation.

REFERENCES

1. De Boo DW, van Hoom F, van Schuppen J, et al. Observer training for computer-aided detection of pulmonary nodules in chest radiography. Eur Radiol 2012; 22(8):1659–1664.

2. Rodríguez-Ruíz A, Krupinski E, Merdang JJ, et al. Detection of breast cancer with mammography: effect of an artificial intelligence support system. Radiology 2019; 290(2):305–314.

3. Jung NY, Kang BJ, Kim HS, et al. Who could benefit the most from using a computer-aided detection system in full-field digital mammography? World J Surg Oncol 2014; 12:168.

4. Miyake M, Iinuma G, Taylor SA, et al. Comparative performance of a primary-reader and second-reader paradigm of computer-aided detection for CT colonography in a low-prevalence screening population. Jpn J Radiol 2013; 31(5):310–319.

5. Mozaffary A, Trabzonolu TA, Lombardi P, et al. Integration of fully automated computer-aided pulmonary nodule detection into CT pulmonary angiography studies in the emergency department: effect on workflow and diagnostic accuracy. Emerg Radiol 2019; 26(6):609–614.

6. Matsumoto S, Ohno Y, Aoki T, et al. Computer-aided detection of lung nodules on multidetector CT in concurrent-reader and second-reader modes: a comparative study. Eur J Radiol 2013; 82(8):1332–1337.

7. Vassallo L, Traverso A, Agnello M, 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):144–152.

8. Wittenberg R, Berger FH, Peters JF, et al. Acute pulmonary embolism: effect of a computer-assisted detection prototype on diagnosis—an observer study. Radiology 2012; 262(1):305–313.

9. Wittenberg R, Peters JF, van den Berk IA, et al. Computed tomography pulmonary angiography in acute pulmonary embolism: the effect of a computer-aided detection prototype used as a concurrent reader. J Thorac Imaging 2013; 28(6):315–321.

10. Chen J, Wu L, Zhang J, et al. Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography. Sci Rep 2020; 10(1):19196.

11. Ni Q, Sun ZY, Qi L, et al. A deep learning approach to characterize 2019 coronavirus disease (COVID-19) pneumonia in chest CT images. Eur Radiol 2020; 30(12):6517–6527.

12. Zhan H, Schartz K, Zygmont ME, et al. The impact of fatigue on complex CT case interpretation by radiology residents. Acad Radiol 2021; 28(3):424–432. doi:10.1016/j.acra.2020.06.005.

13. van Geel K, Kok EM, Krol JP, et al. Reversal of the hanging protocol of contrast enhanced mammography leads to similar diagnostic performance yet decreased reading times. Eur J Radiol 2019; 117:62–68.

14. Wynn RM, Howe JL, Kelahan LC, et al. The impact of interruptions on chest radiograph interpretation: effects on reading time and accuracy. Acad Radiol 2018; 25(12):1515–1520.

15. Krupinski EA. Increasing display luminance as a means to enhance interpretation accuracy and efficiency when reducing full-field digital mammography dose. J Med Imaging 2018; 5(3):035001.

16. Fan SF, Xu Z, He HQ, Ding JR, Teng GJ. The role of key image notes in CT imaging study interpretation. J Digit Imaging 2011; 24(2):366–372.

17. Faggionti L, Neri E, Bargellini I, et al. iPad-based primary 2D reading of CT angiography examinations of patients with suspected acute gastrointestinal bleeding: preliminary experience. Br J Radiol 2015; 88(1047):20140477.

18. Kelly BS, Rainford LA, Darcy SP, et al. The Development of expertise in radiology: in chest radiograph interpretation, "expert" search pattern may predate "expert" levels of diagnostic accuracy for pneumothorax identification. Radiology 2016; 280(1):252–260.

19. Ogrinc G, Davies L, Goodman D, et al. SQUIRE 2.0 (Standards for Quality Improvement reporting excellence): revised publication guidelines from a detailed consensus process. BMJ Quality & Safety 2016; 25(12):986–992.

20. Brown M, Browning P, Wahi-Anwar MW, et al. Integration of chest CT CAD into the clinical workflow and impact on radiologist efficiency. Acad Radiol 2019; 26(5):626–631.

21. Silva M, Schaef er-Prokop CM, Jacobs C, et al. Detection of subsolid nodules in lung cancer screening: complementary sensitivity of visual reading and computer-based diagnosis. Invest Radiol 2018; 53(8):441–449.

22. Forberg D, Rospik B, Sunshine JL. Radiologists’ variation of time to read across different procedure types. J Digit Imaging 2017; 30(1):86–94.