Study protocol

1 Title

Diagnostic Performanc**e** of Deep learning image reconstruc**T**ion in low Dos**e** **CT** for the Detection of **Acute** Abdominal Conditions

Working title: DETECT Acute

2 Participants

On-site Oslo University Hospital, Ullevål, Norway (Oslo):

Principal investigator Oslo: Anselm Schulz, MD, PhD, Department of Radiology and Nuclear Medicine, Oslo University Hospital, Norway

Johann Baptist Dormagen, MD, PhD, Department of Radiology and Nuclear Medicine, Oslo University Hospital, Norway

Kristin Jensen, MSc, PhD, Department of Diagnostic Physics, Oslo University Hospital, Norway

Bjørn Helge Østerås, MSc, PhD, Department of Diagnostic Physics, Oslo University Hospital, Norway

Tom Mala, MD, Prof. II, Department for surgery, inflammatory medicine and transplantation, Oslo University Hospital, Norway

Tormund Njølstad, MD, PhD candidate, Department of Radiology and Nuclear Medicine, Oslo University Hospital, Norway

Knut Waagan, PhD, Senior Statistician, University of Oslo, Norway

On-site Odense University Hospital, Denmark (Odense):

Principal investigator Odense: Bo Mussmann, Research Radiographer, PhD, Associate Professor at University of Southern Denmark and Oslo Metropolitan University

Ole Graumann, Associate Professor, MD, PhD, Radiology Department, Odense University Hospital (OUH), Denmark

Stefan Posth, Associate Professor, MD, PhD, Department of Clinical Research, OUH, Denmark

Mark Bremholm Ellebæk, Associate Professor, MD, PhD, Department for abdominal surgery, OUH.

Mohammad Talal Elhakim, MD, PhD student, Radiology Department, Odense University Hospital (OUH), Denmark

3 Background

Computed Tomography (CT) has become an essential tool in modern clinical medicine (1, 2). With widespread availability, a rapid increase in the use of CT imaging has been observed over the last decades (3). With the associated increase in radiation exposure, the potential increased risk for radiation-induced malignancy has become a public health concern (4). This is especially true for CT scans of the abdomen and pelvis which currently account for 50% of the collective CT dose (5). As the benefit of dose reduction in general is offset by deterioration of image quality, technological advances to reduce radiation dose without compromising image quality are aspired in clinical practice.

In CT-image reconstruction, filtered back projection (FBP) has been the dominant image reconstruction technique algorithm since the early 1970s, complemented by the first commercial iterative reconstruction (IR) algorithms in 2009 (6, 7).

A novel deep learning image reconstruction (DLIR) algorithm received clinical approval in 2019 (TrueFidelity, GE Healthcare, Milwaukee, WI). Other vendor-specific algorithms for deep learning image reconstruction are also emerging (AiCE, Canon Medical Systems, Otawara, Japan). As explained by a technical white paper (8), having been trained with high-dose and low-dose FBP datasets across phantom and patient cases, the DLIR algorithm strives to suppress image noise without compromising image quality. The use of deep learning image reconstruction has demonstrated potential for improved image quality (9-11) and dose reduction without shifting noise texture (12-14).

For patients with acute abdominal conditions, CT of the abdomen and pelvis is considered the best first- or second-line diagnostic approach (15-18). For these patients a fast and accurate diagnosis is of great importance to avoid treatment delay and subsequent complications such as gastrointestinal perforation in case of appendicitis or diverticulitis (19). On the other hand, it is also important to avoid unnecessary surgical intervention and the related complications. A possible low-dose CT protocol must therefore provide a non-inferior diagnostic performance to facilitate fast diagnosis and avoid overtreatment and inconclusive examinations.

Promising results have been reported regarding low-dose CT examinations with model-based IR and dose reduction of up to 75-80% (20, 21). However, with the introduction of DLIR even further dose reduction seems feasible. Our own results from an image quality perception study with DLIR indicate that a dose reduction of up to 92.5% compared to standard CT might preserve acceptable diagnostic image quality (yet unpublished work).

On this basis, the purpose of this study is to assess the diagnostic performance of low-dose CT with DLIR for the diagnosis of acute abdominal conditions in a non-inferiority setting with a large sample size provided by two major trauma centers in northern Europe.

4 Aims

Primary:

To evaluate the diagnostic performance for acute abdominal conditions of contrast enhanced low-dose CT with DLIR “TrueFidelity” (TF) compared to standard full-dose CT.

Secondary:

To evaluate technical and perceived image quality (qualitatively and quantitatively).

5 Ethics

Approval will be obtained from the regional ethics committee and the institutions data protection officer.

Written informed consent will be obtained from all participants. Please see the attached patient consent form.

This project will be in accordance with the Helsinki Declaration.

5.1 Risks

Minimal risks exist due to a slight increase in radiation exposure. The additional radiation exposure of 27.5% is within the national variation of radiation exposure from CT exams performed for corresponding clinical tasks (22). We estimated the mean additional effective dose to 1.5 mSv which corresponds to about 4 months with natural background radiation exposure in Norway (4.1 mSv/year) (23). The additional radiation exposure translates into a theoretical excess lifetime risk of deadly radiation induced cancer between 0.004 – 0.03%.[1] The clinical risks from this exposure are considered to be minimal/not significant.

6 Material and Methods

The study will be registered at ClinicalTrials.gov prior to initiation. Study methods and results will be reported in agreement with the Standards for Reporting of Diagnostic Accuracy Studies (STARD) statement of 2015 (24). It should be noted that the STARD-AI Steering Group is preparing an AI-specific extension (25). If these STARD-AI guidelines are published before end of study, the findings will also be reported in accordance herewith. To compensate for AI specific elements not addressed in STARD, we will, when relevant, rely on the Checklist for Artificial Intelligence in Medical Imaging (CLAIM) (26) which is modelled after the STARD guideline.

6.1 Inclusion

- Patients under evaluation for an acute abdominal condition who are referred to CT of the abdomen and pelvis.
- Age >18 years

[1] Calculated for a 70 year old male and an 18 year old female, respectively, using https://www.xrayrisk.com/calculator/calculator-normal-studies.php

- The patients must be able to give their oral and written consent to study participation.

6.2 Exclusion:

- Contraindications regarding contrast enhanced CT examinations like known iodinated contrast media adverse reactions or claustrophobia.
- Pregnancy.

6.3 Pilot

A study pilot including 10 patients divided equally between Oslo and Odense will be performed to allow for testing of study logistics and adjustments of the radiation dose level of the low-dose CT.

6.4 Examination protocol / imaging

Examinations will be carried out according to local routine procedures and established CT protocols (CT scanner: GE Revolution). Please find detailed imaging protocols for Oslo and Odense in the attachment.

In addition to the CT with standard examination protocol a low-dose CT scan will be performed, not exceeding 30% radiation dose of the standard CT. Low-dose CT images will be reconstructed with TF high. The low dose CT will be performed directly after the standard CT to avoid bias from differences in the timing of the contrast phase.

6.5 Location and local study population

The study will be carried out as a multicenter study involving Oslo and Odense with prospective data collection.

The estimated total study population will be divided equally between the two Hospitals.

6.6 Image evaluation

The low-dose CT will not be used for diagnostic purposes or patient treatment. Image evaluation and comparison will be conducted separated from clinical routine workflow.

All low-dose CT exams will be evaluated independently by two resident radiologists and by two experienced radiologists specialized in abdominal radiology with more than 10 years of experience in abdominal CT. The readers will be blinded for all information from previous exams, the primary CT report, any finding by the other readers, all treatment related information and for the final diagnosis.

They will have access to clinical referrals and laboratory tests performed prior to the original CT examination. Image evaluation will be performed in the radiologists' clinical environment using diagnostic monitors.

In the outcome analysis, the diagnosis for each patient from low dose CT will be compared to the original radiological diagnosis based on full dose CT.

For intra reader agreement a random selection from 10% of the cases will be presented twice to each reader.

Technical image quality is assessed by positioning regions of interest (ROI) in a homogeneous segment of the portal vein, adjacent normal liver parenchyma aorta, erector spinae muscles and in the subcutaneous fat. Contrast-to-Noise Ratio (CNR) will be calculated using the formula (27).

$$CNR = \frac{|HU_{Kidney} - HU_{Water}|}{\sqrt{\frac{SD_{Kidney}^2 + SD_{Water}^2}{2}}}$$

Perceived image quality will be assessed by at least two radiologists on a Likert-type scale along image quality criteria based on the European guidelines for image quality in abdominal CT (28).

6.7 Statistics

Dedicated statistical software like Stata and SPSS will be used for analysis of study data. The alpha significance level will be set to 5% and 95% confidence intervals will be used. Kappa statistics will be used for inter and intra reader agreement. Logistic regression will be used for image quality assessment. Appropriate parametric or non-parametric tests will be used for evaluation of numeric variables. The diagnostic performance will be defined by area under the curve, sensitivity, specificity, positive and negative predictive value. Significant differences in sensitivity and specificity will be determined by McNemar's test.

6.7.1 Power calculation and sample size

A non-inferiority study design will be used to show noninferiority regarding the diagnostic performance of the low-dose CT compared to standard CT (29). We estimated the sensitivity of the standard CT to 90% (30). The prevalence of acute abdominal conditions with a visible correlate on standard CT is estimated to 70% among all referrals meeting inclusion criteria. A non-inferiority margin of 10% for sensitivity was considered as clinical acceptable i.e. the probability for positive findings on low-dose CT in case of positive standard CT was assumed to be 90%.

S_L = Probability (positive low-dose CT | positive standard CT)

The 0-hypothesis was defined as:

S_L < 90%

The alternative hypothesis was then defined as:

S_L > 90%

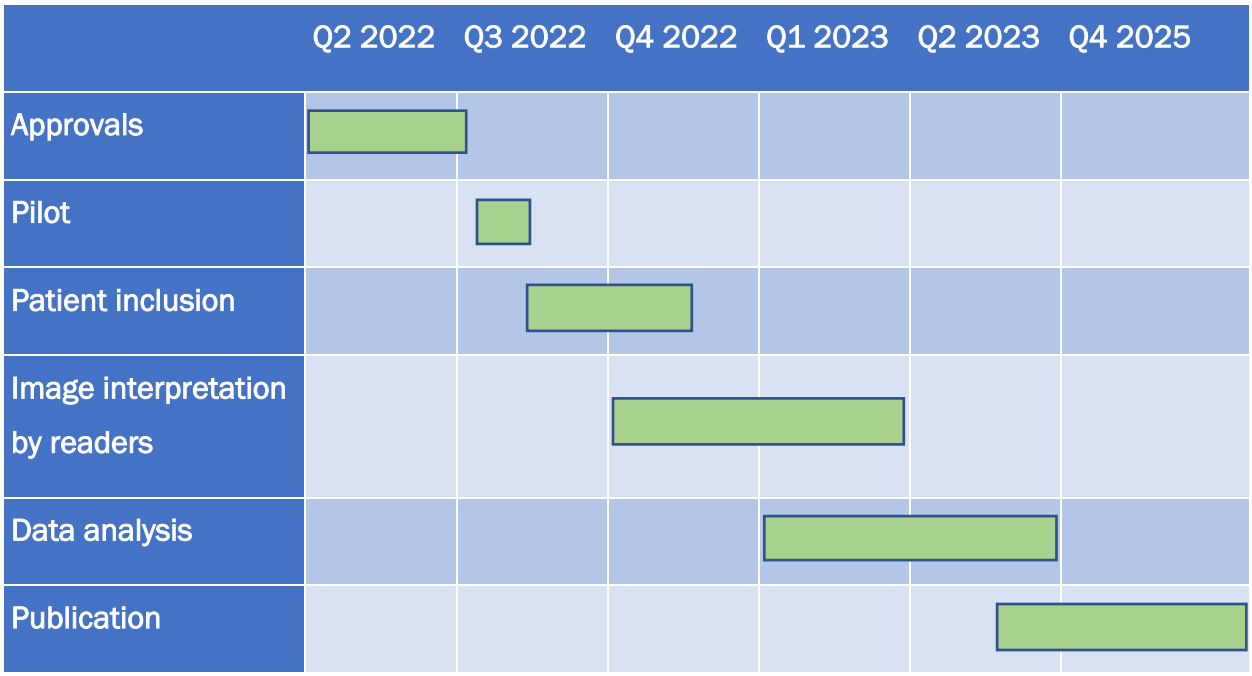
To identify a one-sided 6% difference (increase) from the non-inferiority margin with a power of 80% and an alpha significance level of 5%, we estimated the required patients with positive CT findings to n=116 (binominal distribution). The total number of required patients was then calculated to 116/0.7=166.

6.8 Time schedule

The necessary approvals will be obtained during q2-3 of 2022.

Inclusion and data collection q3-4 2022.

Data analysis and publication q4 2022 - 2025.



7 Variables

Variables for data collection.

7.1 Demographic

Sex, age and bmi

7.2 Primary

Clinical diagnosis (ICR code); clinical diagnosis cat (categorical, derived from “clinical diagnosis”); CT diagnosis (categorical); radiation dose as DLP/$CTDI_{vol}$ (numerical); CT protocol (categorical); vendor (categorical)

7.3 Secondary

Reader experience (ordinal); patient time in hospital (numerical); treatment (categorical); time to read one CT exam (numerical); image quality (ordinal, 5-point Likert scale); image noise (numerical, several points of measurement); contrast-to-noise ratio (numerical, several points of measurement).

8 References

1. Brenner DJ, Hall EJ. Computed Tomography — An Increasing Source of Radiation Exposure. New Engl J Med 2007;357(22):2277-2284. doi: 10.1056/NEJMra072149
2. Novelline RA, Rhea JT, Rao PM, Stuk JL. Helical CT in Emergency Radiology. Radiology 1999;213(2):321-339. doi: 10.1148/radiology.213.2.r99nv01321
3. OECD. Computed tomography (CT) exams. 2018.
4. Berrington de Gonzalez A, Mahesh M, Kim KP, Bhargavan M, Lewis R, Mettler F, Land C. Projected cancer risks from computed tomographic scans performed in the United States in 2007. Arch Intern Med 2009;169(22):2071-2077. doi: 10.1001/archinternmed.2009.440
5. Mettler FA, Jr., Thomadsen BR, Bhargavan M, Gilley DB, Gray JE, Lipoti JA, McCrohan J, Yoshizumi TT, Mahesh M. Medical radiation exposure in the U.S. in 2006: preliminary results. Health Phys 2008;95(5):502-507. doi: 10.1097/01.HP.0000326333.42287.a2
6. Pan X, Sidky EY, Vannier M. Why do commercial CT scanners still employ traditional, filtered back-projection for image reconstruction? Inverse Probl 2009;25(12):1230009. doi: 10.1088/0266-5611/25/12/123009
7. Beister M, Kolditz D, Kalender WA. Iterative reconstruction methods in X-ray CT. Phys Med 2012;28(2):94-108. doi: 10.1016/j.ejmp.2012.01.003
8. Hsieh JL, E.; Nett, B.; Tang, J.; Thibault JB.; Sahney, S. A new era of image reconstruction: TrueFidelity. Technical white paper on deep learning image reconstruction. 2019.
9. Akagi M, Nakamura Y, Higaki T, Narita K, Honda Y, Zhou J, Yu Z, Akino N, Awai K. Deep learning reconstruction improves image quality of abdominal ultra-high-resolution CT. European radiology 2019;29(11):6163-6171. doi: 10.1007/s00330-019-06170-3
10. Jensen CT, Liu X, Tamm EP, Chandler AG, Sun J, Morani AC, Javadi S, Wagner-Bartak NA. Image Quality Assessment of Abdominal CT by Use of New Deep Learning Image Reconstruction: Initial Experience. AJR American journal of roentgenology 2020;215(1):50-57. doi: 10.2214/AJR.19.22332

11. Njolstad T, Schulz A, Godt JC, Brogger HM, Johansen CK, Andersen HK, Martinsen ACT. Improved image quality in abdominal computed tomography reconstructed with a novel Deep Learning Image Reconstruction technique - initial clinical experience. Acta Radiol Open 2021;10(4):20584601211008391. doi: 10.1177/20584601211008391
12. Solomon J, Lyu P, Marin D, Samei E. Noise and spatial resolution properties of a commercially available deep learning-based CT reconstruction algorithm. Medical physics 2020;47(9):3961-3971. doi: 10.1002/mp.14319
13. Greffier J, Hamard A, Pereira F, Barrau C, Pasquier H, Beregi JP, Frandon J. Image quality and dose reduction opportunity of deep learning image reconstruction algorithm for CT: a phantom study. European radiology 2020;30(7):3951-3959. doi: 10.1007/s00330-020-06724-w
14. Brady SL, Trout AT, Somasundaram E, Anton CG, Li Y, Dillman JR. Improving Image Quality and Reducing Radiation Dose for Pediatric CT by Using Deep Learning Reconstruction. Radiology 2021;298(1):180-188. doi: 10.1148/radiol.2020202317
15. Larson DB, Johnson LW, Schnell BM, Salisbury SR, Forman HP. National trends in CT use in the emergency department: 1995-2007. Radiology 2011;258(1):164-173. doi: 10.1148/radiol.10100640
16. Expert Panel on Gastrointestinal I, Garcia EM, Camacho MA, Karolyi DR, Kim DH, Cash BD, Chang KJ, Feig BW, Fowler KJ, Kambadakone AR, Lambert DL, Levy AD, Marin D, Moreno C, Peterson CM, Scheirey CD, Siegel A, Smith MP, Weinstein S, Carucci LR. ACR Appropriateness Criteria((R)) Right Lower Quadrant Pain-Suspected Appendicitis. Journal of the American College of Radiology : JACR 2018;15(11S):S373-S387. doi: 10.1016/j.jacr.2018.09.033
17. Expert Panel on Gastrointestinal I, Peterson CM, McNamara MM, Kamel IR, Al-Refaie WB, Arif-Tiwari H, Cash BD, Chernyak V, Goldstein A, Grajo JR, Hindman NM, Horowitz JM, Noto RB, Porter KK, Srivastava PK, Zaheer A, Carucci LR. ACR Appropriateness Criteria((R)) Right Upper Quadrant Pain. Journal of the American College of Radiology : JACR 2019;16(5S):S235-S243. doi: 10.1016/j.jacr.2019.02.013
18. Rud B, Vejborg TS, Rappeport ED, Reitsma JB, Wille-Jorgensen P. Computed tomography for diagnosis of acute appendicitis in adults. Cochrane Database Syst Rev 2019;2019(11). doi: 10.1002/14651858.CD009977.pub2
19. Kabir SA, Kabir SI, Sun R, Jafferbhoy S, Karim A. How to diagnose an acutely inflamed appendix; a systematic review of the latest evidence. International journal of surgery (London, England) 2017;40:155-162. doi: 10.1016/j.ijsu.2017.03.013
20. Moloney F, James K, Twomey M, Ryan D, Grey TM, Downes A, Kavanagh RG, Moore N, Murphy MJ, Bye J, Carey BW, McSweeney SE, Deasy C, Andrews E, Shanahan F, Maher MM, O'Connor OJ. Low-dose CT imaging of the acute abdomen using model-based iterative reconstruction: a prospective study. Emergency radiology 2019;26(2):169-177. doi: 10.1007/s10140-018-1658-z
21. Poletti PA, Becker M, Becker CD, Halfon Poletti A, Rutschmann OT, Zaidi H, Perneger T, Platon A. Emergency assessment of patients with acute abdominal pain using low-dose CT with iterative reconstruction: a comparative study. European radiology 2017;27(8):3300-3309. doi: 10.1007/s00330-016-4712-9
22. Widmark A. Diagnostic reference level (DRL) in Norway 2017. Results, revision:and establishment of new DRL.NRPA Report 2018:3. Norwegian Radiation Protection Authority, Østerås 2018.

23. Komperød M, Rudjord AL, Skuterud L, Dyve JE. Radiation Doses from the Environment. Calculations of the Public's Exposure to Radiation from the Environment in Norway. Strålevern Rapport 2015:11 Østerås: Norwegian Radiation Protection Authority 2015.
24. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, Lijmer JG, Moher D, Rennie D, de Vet HC, Kressel HY, Rifai N, Golub RM, Altman DG, Hooft L, Korevaar DA, Cohen JF, Group S. STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies. Radiology 2015;277(3):826-832. doi: 10.1148/radiol.2015151516
25. Sounderajah V, Ashrafian H, Golub RM, Shetty S, De Fauw J, Hooft L, Moons K, Collins G, Moher D, Bossuyt PM, Darzi A, Karthikesalingam A, Denniston AK, Mateen BA, Ting D, Treanor D, King D, Greaves F, Godwin J, Pearson-Stuttard J, Harling L, McInnes M, Rifai N, Tomasev N, Normahani P, Whiting P, Aggarwal R, Vollmer S, Markar SR, Panch T, Liu X, Committee S-AS. Developing a reporting guideline for artificial intelligence-centred diagnostic test accuracy studies: the STARD-AI protocol. BMJ Open 2021;11(6):e047709. doi: 10.1136/bmjopen-2020-047709
26. Mongan J, Moy L, Kahn CE, Jr. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A Guide for Authors and Reviewers. Radiol Artif Intell 2020;2(2):e200029. doi: 10.1148/ryai.2020200029
27. Abdi AJ, Mussmann B. Evaluation of patient dose saving in grid-less x-ray mammography acquisition compared with full field digital mammography (FFDMG) acquisition. Biomedical Physics & Engineering Express 2017;3(1):015026. doi: 10.1088/2057-1976/aa5d1d
28. Report EUR 16262 EN. European guidelines on quality criteria for computed tomography. 2000.
29. Ahn S, Park SH, Lee KH. How to demonstrate similarity by using noninferiority and equivalence statistical testing in radiology research. Radiology 2013;267(2):328-338. doi: 10.1148/radiol.12120725
30. Eng KA, Abadeh A, Ligocki C, Lee YK, Moineddin R, Adams-Webber T, Schuh S, Doria AS. Acute Appendicitis: A Meta-Analysis of the Diagnostic Accuracy of US, CT, and MRI as Second-Line Imaging Tests after an Initial US. Radiology 2018;288(3):717-727. doi: 10.1148/radiol.2018180318