

# BMJ Open CONFIDENT-trial protocol: a pragmatic template for clinical implementation of artificial intelligence assistance in pathology

Rachel N Flach , Nikolas Stathonikos, Tri Q Nguyen, Natalie D Ter Hoeve, Paul J van Diest, Carmen van Doijeweert

**To cite:** Flach RN, Stathonikos N, Nguyen TQ, *et al.* CONFIDENT-trial protocol: a pragmatic template for clinical implementation of artificial intelligence assistance in pathology. *BMJ Open* 2023;**13**:e067437. doi:10.1136/bmjopen-2022-067437

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2022-067437>).

Received 17 August 2022  
Accepted 21 May 2023



© Author(s) (or their employer(s)) 2023. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

Department of Pathology, University Medical Centre Utrecht, Utrecht, The Netherlands

**Correspondence to**  
Dr Carmen van Doijeweert;  
[c.vandooijeweert-3@umcutrecht.nl](mailto:c.vandooijeweert-3@umcutrecht.nl)

## ABSTRACT

**Introduction** Artificial intelligence (AI) has been on the rise in the field of pathology. Despite promising results in retrospective studies, and several CE-IVD certified algorithms on the market, prospective clinical implementation studies of AI have yet to be performed, to the best of our knowledge. In this trial, we will explore the benefits of an AI-assisted pathology workflow, while maintaining diagnostic safety standards.

**Methods and analysis** This is a Standard Protocol Items: Recommendations for Interventional Trials–Artificial Intelligence compliant single-centre, controlled clinical trial, in a fully digital academic pathology laboratory. We will prospectively include prostate cancer patients who undergo prostate needle biopsies (CONFIDENT-P) and breast cancer patients who undergo a sentinel node procedure (CONFIDENT-B) in the University Medical Centre Utrecht. For both the CONFIDENT-B and CONFIDENT-P trials, the specific pathology specimens will be pseudo-randomised to be assessed by a pathologist with or without AI assistance in a pragmatic (bi-)weekly sequential design. In the intervention group, pathologists will assess whole slide images (WSI) of the standard hematoxylin and eosin (H&E)-stained sections assisted by the output of the algorithm. In the control group, pathologists will assess H&E WSI according to the current clinical workflow. If no tumour cells are identified or when the pathologist is in doubt, immunohistochemistry (IHC) staining will be performed. At least 80 patients in the CONFIDENT-P and 180 patients in the CONFIDENT-B trial will need to be enrolled to detect superiority, allocated as 1:1. Primary endpoint for both trials is the number of saved resources of IHC staining procedures for detecting tumour cells, since this will clarify tangible cost savings that will support the business case for AI.

**Ethics and dissemination** The ethics committee (MREC NedMec) waived the need of official ethical approval, since participants are not subjected to procedures nor are they required to follow rules. Results of both trials (CONFIDENT-B and CONFIDENT-P) will be published in scientific peer-reviewed journals.

## INTRODUCTION

### Background and rationale

Since the introduction of digital pathology, the number of studies on artificial

### STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ This is the first clinical trial to examine the added value of artificial intelligence in the daily pathology workflow.
- ⇒ By maintaining the current diagnostic safety standards, patients are not at risk of an inferior diagnosis during the trial.
- ⇒ This is a pragmatic template for prospective AI trials for object-identifying algorithms in pathology.
- ⇒ A limitation is that this is a single-centre study, which may hamper generalisability.
- ⇒ Due to the existing clinical workflow, randomisation of patients and (double-)blinding of the participating pathologists and researchers is not possible.

intelligence (AI) within the field of pathology has increased exponentially.<sup>1 2</sup> Algorithms have been created for tumour detection, tumour grading, tumour subtyping, evaluating biomarkers and more.<sup>1 3</sup> Due to demographic trends, the needs for healthcare are increasing globally, which combined with a lack of specialists, increases the current workload.<sup>2 4</sup> Therefore, AI has great potential to alleviate pathologists' workload<sup>2</sup> and improve diagnostics by improving accuracy, reproducibility and speed.<sup>2</sup> In fact, several algorithms have shown to be comparable, or even superior to pathologists (under time constraint).<sup>2 5–10</sup>

AI and human intelligence are not mutually exclusive, they complement each other, a concept which is known as 'augmented intelligence', where AI can enhance, rather than replace human intelligence.<sup>11</sup> In the (very) early AI-adoption phase, and presumably also in later phases, pathologist supervision remains of key importance. This is particularly relevant as, despite the promising results of retrospective studies and the availability of CE-IVD approved algorithms, prospective validation and

clinical implementation of AI are currently lacking. For example, 6 years after the successful CAMELYON-16 Grand Challenge,<sup>6</sup> the top algorithms have yet to be implemented in daily clinical practice, showing that the time between development of an AI model and clinical implementation is considerable. Likewise, numerous promising prostate cancer (PCa) grading algorithms have been developed, yet implementation studies have yet to be performed,<sup>7 8 12</sup> whereas 9 AI pathology devices received CE-IVD approval in 2021.<sup>13</sup>

### Trial rationale

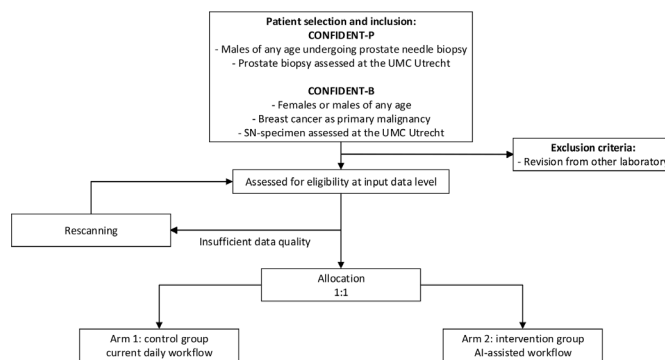
As a pathology laboratory with a fully digital workflow for over 7 years, we are eager to explore the full potential of working digitally by adding the benefit of AI in daily pathology practice. We decided to start with the object localisation task of tumour detection, where an objective reference standard is in place in the routine clinical workflow (ie, pathologist supervision and/or immunohistochemistry (IHC) staining).

We developed the CONFIDENT-trial template, in which AI tumour detection algorithms can be safely implemented in prospective clinical trials, while ensuring that patients are not at risk of receiving an inferior diagnosis, since IHC is always performed when no tumour cells are visible, but also when pathologists need more confirmation about the diagnosis.

### CONFIDENT-B and CONFIDENT-P

Our trials aim to prospectively investigate the added value of an AI-assisted pathology workflow in the identification of PCa in prostate needle biopsies (CONFIDENT-P) and the identification of sentinel node (SN) metastases in patients with breast cancer (BCa) (CONFIDENT-B). As both PCa and BCa are the most common (non-skin) malignancies in men and women, respectively, implementation of AI assistance may have a great impact on diagnostic processes.<sup>14</sup> However, it is important to emphasise that this trial serves as a template for other pragmatic AI-intervention trials for object-localisation tasks as well.

We obtained CE-IVD-approved algorithms for detection and grading of PCa in prostate needle biopsies and an algorithm for detecting lymph node metastases in patients with BCa. In both cases, the task of the pathologist is both labour intensive and expensive, due to the performed IHC stains in case no tumour cells are morphologically observed. However, IHC is expensive and these costs sometimes even exceed reimbursement for the entire specimen (eg, in case of multiple blocks of multiple SNs). This raises the question whether AI may be of added value to morphologically detect cancer cells without the need for IHC use. Thereby, the number of performed IHC stains may be reduced, which may lead to tangible costs savings that will help to build the business case for AI, while potentially decreasing the workload of pathologists as well.<sup>2</sup>



**Figure 1** Flowchart with patient selection. AI, artificial intelligence; SN, sentinel node.

### Study objective

The primary objective is to explore whether an AI-assisted workflow reduces the number of spent resources on IHC, while maintaining diagnostic safety standards in both patients with PCa who underwent prostate needle biopsies (CONFIDENT-P) and patients with BCa who underwent an SN procedure.

Secondary objectives are to investigate whether time management improves in an AI-assisted workflow and to analyse how much IHC staining may have been safely omitted after AI implementation.

## METHODS AND ANALYSIS

### Trial design

The study protocol is structured following the Standard Protocol Items: recommendations for Interventional Trials–Artificial Intelligence (SPIRIT-AI) statement 2020.<sup>15</sup> This study is a single-centre, parallel-group controlled trial, assessing superiority. The allocation ratio is 1:1. Eligible patients will be assigned to arm 1 (control group) or arm 2 (AI-assisted workflow), based on a bi-weekly time schedule. Eligibility criteria are summarised in figure 1. The CONFIDENT trials will be carried out in 2022–2023.

### Study setting

The trial will take place in the daily practice of a single academic hospital (University Medical Centre (UMC) Utrecht, the Netherlands), with a fully digital pathology enabled clinical set-up, where all slides are digitised using ultrafast whole slide image (WSI) using Hamamatsu S360 scanners and reviewed using the Sectra pathology Picture Archiving and Communication System (PACS). Although the UMC Utrecht is an academic hospital, primary routine pathology diagnostics is performed for non-academic hospitals as well (ie, the Alexander Monro Breast Cancer Hospital, Bilthoven, the Netherlands).

### Study population

For PCa, WSI of all males who undergo a prostate needle biopsy in the UMC Utrecht will be included. For BCa, WSI of all females or males with BCa as primary malignancy (ie, invasive BCa) who undergo an SN procedure

in the Alexander Monro Breast Cancer Hospital or the UMC Utrecht will be included. Patients will be excluded, if they were redirected to the UMC Utrecht for a second opinion.

### Assessment of specimen

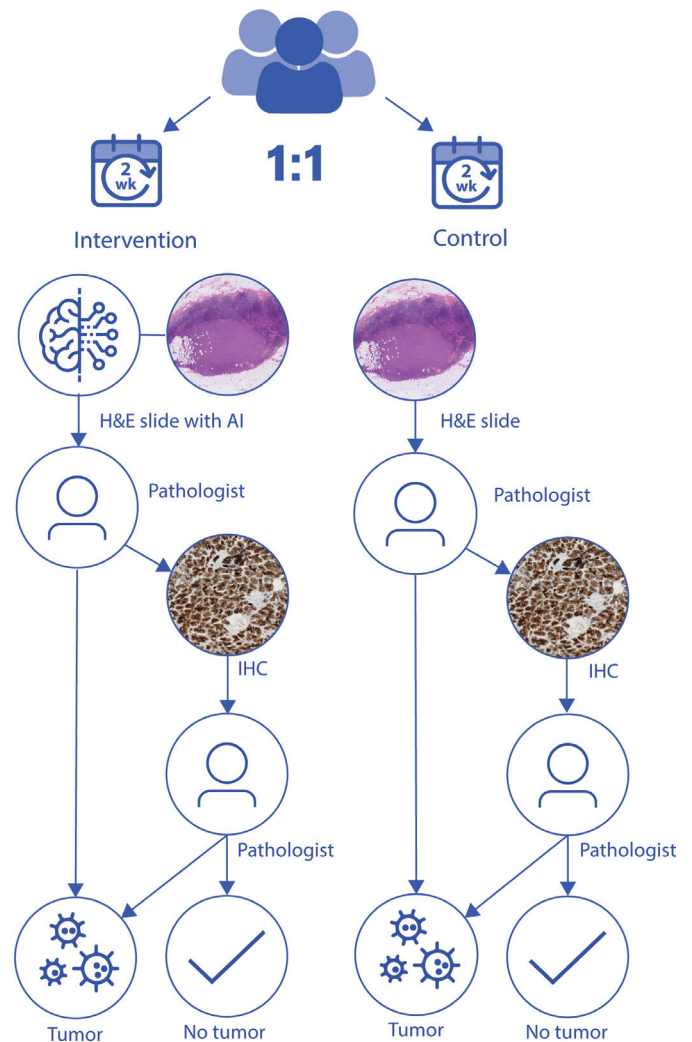
During the study period, all WSI will be assessed by the same group of pathologists; that is, two expert urological pathologists for the PCa biopsies, and three expert breast pathologists for the lymph node assessment from patients with BCa.

For both the CONFIDENT-B and CONFIDENT-P trials, the specific pathology specimens will be assigned to be assessed by a pathologist with or without AI assistance in a pragmatic (bi-)weekly sequential design. This is considered feasible as the change in the case mix and time trends are unlikely to occur within the inclusion period of about 6–9 months. Furthermore, both specialised breast and urological pathologists within the UMC Utrecht work according to weekly schedules. Therefore, using AI every other week or every other 2 weeks, as opposed to switching by day, ensures that all pathologists are equally distributed between groups. Lastly, it would be impractical to switch from AI assistance in the intervention group to no AI assistance in the control group on a case-to-case basis. For obvious reasons, allocation concealment and blinding of pathologists and researchers are not applicable.

### Control and intervention

All eligible specimens will be assigned to either the control group or the intervention group. In the control group, pathologists will assess H&E stained WSI of patients digitally, according to the current clinical workflow. For PCa biopsies, IHC is routinely performed on all cases. For BCa lymph nodes, if no metastases or tumour are present, IHC staining will be performed. Additional IHC staining will also be performed by additional request of the pathologist in case of doubt.

In the intervention group, pathologists will assess the H&E specimens digitally with the outcome of the algorithm provided in their first assessment of the specimen. For PCa, they will use the CE-IVD certified Paige Prostate Suite algorithms for tumour detection and tumour volume percentage calculations, which reaches sensitivity and specificity of 99% and 93%, respectively, and which are based on a weakly supervised deep learning algorithm as described by Campanella *et al.*<sup>16 17</sup> For BCa, pathologists will use the CE-IVD certified Metastasis Detection App by Visiopharm, a deep-learning algorithm for lymph node metastases of BCa and colon carcinoma with a combined sensitivity and specificity of 98.7% and 99.6%, respectively.<sup>18</sup> These algorithms will be integrated within the Sectra PACS where the output of the algorithms will be graphically displayed. AI analysis of the WSI will be performed right after scanning to avoid delays in the clinical workflow. If the AI-assisted pathologist does not detect metastases or tumours on the H&E slide, routine additional IHC staining will be performed by P503S/



**Figure 2** Study flow chart. H&E, hematoxylin and eosin; AI, artificial intelligence; IHC, immunohistochemistry.

p63/CK HMW for PCa and CAM5.2 for BCa, to ensure no metastases or tumours are missed. Pathologists can also request an additional IHC if they feel they need this to make an adequate diagnosis (figure 2).

### Outcome measures

Primary outcome for the CONFIDENT-P trial is the added value of AI assistance in the detection of PCa and in the detection of tumour volume in prostate needle biopsies in daily pathology practice. The primary outcome for the CONFIDENT-B trial is the added value of AI assistance in the detection of BCa SN metastases. The outcome measures for both trials will be the number of spent resources, that is, the number of IHC stains performed in both groups.

Secondary outcome measures will be sensitivity and specificity of the AI-assisted pathologist, time spent on WSI analysis, the number of IHC stains that may have been omitted after AI-implementation and a pathologists' evaluation by a questionnaire on the AI-assisted work process. Sensitivity and specificity analyses of the algorithm itself have already been well documented, and are, therefore,



outside the scope of the paper, as we focus on the combination of pathologist and AI to explore cost savings.

### Input data

Input data for the algorithm will be WSI of H&E stained slides scanned at 40× of either prostate needle biopsies, and WSI of H&E stained slides of BCa lymph nodes. As per routine in our daily clinical practice, WSI will be quality controlled after scanning for colour, focus quality and completeness of the scan. When necessary, the specimens are rescanned.

### Sample size

#### CONFIDENT-P

We performed power calculations using a two-sample proportion superiority test, using expected percentages of IHC staining in both study arms. We assume that the pathologists in the control arm can detect 50% of the tumours without using IHC. We expect AI-assisted pathologists to detect 80% of the tumours, without using IHC. These percentages were conservatively derived from the validity study by Raciti (74% for pathologists without AI and 90% for pathologists with AI, respectively),<sup>9</sup> by expert pathologist opinion, and taking into account that pathologists under time constraint of daily practice do not detect tumours as well as pathologists without time constraint during retrospective studies.<sup>19</sup> We assume that this effect will be larger for the biopsies assessed without AI than with AI, as AI is assumed to make tumour detection easier.

A sample size of 60 (30 per arm) would give a power of approximately 80%, using a one-sided 5% significance level. However, uncertainties remain regarding the sample size parameters. We, therefore, inflated our sample size to 80 (40 per arm), in order to ensure study power and allowing us to detect smaller effect sizes.

For detection of tumour volume percentage, we performed a power calculation based on the assumption that AI should be able to replace at least 20% of the IHC stains, in order to be cost effective. IHC is currently used in 100% of all prostate needle biopsies. Using a power of 80% and a one-sided significance level of 5%, this leads to 27 patients per arm.

#### CONFIDENT-B

Sample size calculations for the CONFIDENT-B trial are based on the assumption that the AI algorithm can detect all metastases for which currently IHC is used, which are mainly micro-metastases and isolated tumour cells (ITC). Approximately 15% of the SN specimens in the UMC Utrecht contain a micrometastasis or ITC.

A sample size of 166 patients (83 per arm) with a one-sided 5% significance level, therefore, results in a power of 80%. Again, as there are uncertainties on the assumptions on what amount of the metastases will be detected by AI, we decided to be conservative and include 180 patients (90 per arm).

Overall, we are only interested in one-sided outcomes, as it is not possible that more IHC will be performed in

the AI-assisted arm. IHC is performed to detect metastases, when they are macroscopically undetectable, rather than to confirm them when they are macroscopically visible. As AI would show only more metastases than the pathologist could macroscopically detect, this means that only a reduction of IHC is possible.

Sample sizes were calculated using the `power.prop.test` command in R V.4.2.2.<sup>20</sup>

### Statistical methods

For baseline comparisons between both arms, the appropriate measures (parametric or non-parametric) for categorical ( $\chi^2$  test/Fisher's exact) and continuous variables (T-test and Mann-Whitney U test) will be used. For the analysis of the primary outcome measure, we will compare the proportion of IHC use in both arms, and calculate adjusted relative risks, using a log-binomial model.<sup>21–23</sup>

Missing data for baseline characteristics and for the primary outcome are not to be expected, as they are obligatory items in the structured pathology reports.

We will determine sensitivity and specificity of the conclusion of the AI-assisted pathologists without the use of IHC. Subsequently, we will focus on the cases with metastases that the AI-assisted pathologist misses, categorise them (ie, macro-metastases, micro-metastases and ITC) and determine their clinical relevance (ie, clinical consequences if these metastases are being missed). Data analysis will be performed in R Statistical Software,<sup>21</sup> with a significance level set at  $p < 0.05$ .

### Data collection and management

All data (baseline and primary outcome measurements) will be retrieved from the structured pathology reports and will be managed and stored in Castor EDC.<sup>24</sup> For the secondary outcome measure of time spent by the pathologist on a slide, data will be collected on an interval basis for practical reasons, as timing every assessment for months (by stopwatch) was not deemed feasible. For the secondary outcome measure of AI-assisted work process for pathologists, a questionnaire will be distributed to the participating pathologists. The final secondary assessment measurement, the number of IHC stains that may have been omitted after AI-implementation, will be determined by the researchers based on the data from the structured pathology reports (combination of IHC and AI-assisted diagnoses of the pathologist).

### Ethical approval

Research within these trials is not subject to the (Dutch) Medical Research Involving Human Subjects Act (WMO), as participants are not subjected to procedures and as they are not required to follow rules. Therefore, the ethics committee (MREC NedMec) waived the need of ethical approval and informed consent.

### Risk of harm

Patients are not at risk of any harm for an inferior diagnosis (ie, missed tumour cells), as in both arms, IHC staining will be performed when no tumour cells are

visible, according to current clinical standards. As a rule in augmented intelligence, all cases will be evaluated by a pathologist, which further minimised the risk of a false diagnosis based on the AI algorithm. Taking all of the above into account, a data monitoring committee is not required and adverse events are not to be expected. In theory, the algorithm could be more of a disturbance than a help to pathologists (eg, when it frequently reports false positive or negative results, which have to be corrected by the pathologist). However, the algorithms used are IVDR approved, and thus have undergone extensive review for their intended purpose. Nonetheless, the experience and ease of use of pathologists working with the algorithm will be one of the secondary outcome measures.

### Informed consent and data access

Informed consent was waived by the local quality coordinator and data protection officer for the following reasons. First, in both arms, patients receive standard care, while maintaining diagnostic safety standards (pathologists' supervision and IHC in all negative cases). Second, patients are not subjected to any procedures. Third, all patient data will be anonymised to the researchers by the pathologist who assessed the slide.

The collected (anonymous) research data will be stored in Castor EDC to ensure data security. Data will be kept for a period of 15 years. Data access in Castor will be restricted to two researchers (RNF and CvD). Pathologists have access to the electronic patient files for the purpose of patient care. The researchers are not permitted access to these files. At no point will the data (both in Castor EDC and patient files) be accessed by the companies providing the algorithms (ie, Visiopharm and Paige).

### Patient and public involvement

None.

## DISCUSSION

The promising retrospective results of AI-assisted pathology have not yet resulted in prospective clinical implementation studies. This may be due to a lack of digital transition in the majority of pathology laboratories, but it may also be partly due to the lack of a good implementation model. Fortunately, however, new guidelines for AI trials have recently been proposed by the SPIRIT-AI and Consolidated Standards of Reporting Trials-AI steering groups, as well as roadmaps to routine use of AI in clinical practice.<sup>15 25 26</sup> Yet, to date, no pathology AI trials have been published in PubMed or Web of Science or, to the best of our knowledge, otherwise made public.

As a pathology laboratory with a fully digital workflow, we developed a clinical trial template for tumour detection models, as a first step to implement AI in daily pathology practice. We will start with an object localisation task (ie, tumour cells) as a reference standard is in place in the routine clinical workflow. For classification

tasks like tumour grading, a clinical trial design is more challenging, as no reference is in place in daily pathology practice and inter-laboratory and inter-pathologist variation is notorious.<sup>27-31</sup> Nevertheless, in future trials, implementing AI assistance in the grading process might also reduce this variation. For now, results of the CONFIDENT trials will provide the first assessment of the potential added value of AI in daily pathology practice. This evaluation will substantially contribute to a potential paradigm shift in tumour detection in pathology. The pragmatic template of the CONFIDENT trials may serve as example for other prospective AI implementation trials in diagnostic pathology.

**Contributors** PJvD conceived of the study. RNF and CvD initiated the study design and NS, TQN and NDTH helped with implementation. RNF and CvD provided statistical expertise in clinical trial design and conducted the primary statistical analysis. All authors contributed to refinement of the study protocol and approved the final manuscript.

**Funding** The Hanarth Foundation has provided funding to support this study.

**Competing interests** PJvD is a member of the scientific advisory board of Paige and Sectra.

**Patient and public involvement** Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

**Patient consent for publication** Not applicable.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

### ORCID ID

Rachel N Flach <http://orcid.org/0000-0002-4448-2895>

## REFERENCES

- Jiang Y, Yang M, Wang S, *et al*. Emerging role of deep learning-based artificial intelligence in tumor pathology. *Cancer Commun (Lond)* 2020;40:154–66. 10.1002/cac2.12012 Available: <https://onlinelibrary.wiley.com/toc/25233548/40/4>
- van der Laak J, Litjens G, Ciompi F. Deep learning in Histopathology: the path to the clinic. *Nat Med* 2021;27:775–84.
- Steiner DF, Chen P-H, Mermel CH. Closing the translation gap: AI applications in Digital pathology. *Biochim Biophys Acta Rev Cancer* 2021;1875.
- Robboy SJ, Weintraub S, Horvath AE, *et al*. Pathologist workforce in the United States: I. development of a predictive model to examine factors influencing supply. *Arch Pathol Lab Med* 2013;137:1723–32.
- Bulten W, Pinckaers H, van Boven H, *et al*. Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study. *Lancet Oncol* 2020;21:233–41.
- Ehteshami Bejnordi B, Veta M, Johannes van Diest P, *et al*. Diagnostic assessment of deep learning Algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* 2017;318:2199–210.
- Steiner DF, Nagpal K, Sayres R, *et al*. Evaluation of the use of combined artificial intelligence and pathologist assessment to review and grade prostate biopsies. *JAMA Netw Open* 2020;3.
- Ström P, Kartasalo K, Olsson H, *et al*. Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study. *Lancet Oncol* 2020;21:222–32.
- Raciti P, Sue J, Ceballos R, *et al*. Novel artificial intelligence system increases the detection of prostate cancer in whole slide images of core needle biopsies. *Mod Pathol* 2020;33:2058–66.
- Hekler A, Utikal JS, Enk AH, *et al*. Superior skin cancer classification by the combination of human and artificial intelligence. *Eur J Cancer* 2019;120:114–21.



- 11 Harrison JH, Gilbertson JR, Hanna MG, *et al.* Introduction to artificial intelligence and machine learning for pathology. *Arch Pathol Lab Med* 2021;145:1228–54.
- 12 Bulten W, Balkenhol M, Belinga J-J, *et al.* Artificial intelligence assistance significantly improves Gleason grading of prostate biopsies by Pathologists. *Mod Pathol* 2021;34:660–71.
- 13 Muehlematter UJ, Daniore P, Vokinger KN. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015-20): a comparative analysis. *Lancet Digit Health* 2021;3:e195–203.
- 14 Sung H, Ferlay J, Siegel RL, *et al.* Global cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2021;71:209–49.
- 15 Cruz Rivera S, Liu X, Chan A-W, *et al.* Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Lancet Digit Health* 2020;2:e549–60.
- 16 Campanella G, Hanna MG, Geneslaw L, *et al.* Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat Med* 2019;25:1301–9.
- 17 da Silva LM, Pereira EM, Salles PG, *et al.* Independent real-world application of a clinical-grade automated prostate cancer detection system. *J Pathol* 2021;254:147–58.
- 18 Visiopharm. Available: <https://visiopharm.com/app-center/app/metastasis-detection-ai>
- 19 Ehteshami Bejnordi B, Veta M, Johannes van Diest P, *et al.* Diagnostic assessment of deep learning Algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* 2017;318:2199.
- 20 R Core Team. R: A language and environment for statistical computing. 2022. Available: <https://www.r-project.org>
- 21 McNutt L-A, Wu C, Xue X, *et al.* Estimating the relative risk in cohort studies and clinical trials of common outcomes. *Am J Epidemiol* 2003;157:940–3.
- 22 Lumley T, Kronmal R, Ma S. Relative risk regression in medical research: models, contrasts, Estimators, and Algorithms [<http://www.bepress.com/uwbiostat/paper293>]. In: *UW Biostatistics Working Paper Series 293*. 2006.
- 23 Zou G. A modified Poisson regression approach to prospective studies with binary data. *Am J Epidemiol* 2004;159:702–6.
- 24 Castor EDC. Castor electronic data capture. [Online]. 2019. Available: <https://castoredc.com>
- 25 Liu X, Cruz Rivera S, Moher D, *et al.* Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Lancet Digit Health* 2020;2:e537–48.
- 26 Colling R, Pitman H, Oien K, *et al.* Artificial intelligence in Digital pathology: a roadmap to routine use in clinical practice. *J Pathol* 2019;249:143–50. 10.1002/path.5310 Available: <https://onlinelibrary.wiley.com/toc/10969896/249/2>
- 27 Kuijpers C, Sluijter CE, von der Thüsen JH, *et al.* Interlaboratory variability in the histologic grading of colorectal adenocarcinomas in a nationwide cohort. *Am J Surg Pathol* 2016;40:1100–8.
- 28 Kuijpers C, Sluijter CE, von der Thüsen JH, *et al.* Interlaboratory variability in the grading of dysplasia in a nationwide cohort of colorectal adenomas. *Histopathology* 2016;69:187–97.
- 29 Flach RN, Willemse P-P, Suelmann BBM, *et al.* Significant Inter- and Intralaboratory variation in Gleason grading of prostate cancer: A nationwide study of 35,258 patients in the Netherlands. *Cancers (Basel)* 2021;13.
- 30 van Dooijeweert C, van Diest PJ, Willems SM, *et al.* Significant Inter- and intra-laboratory variation in grading of Ductal carcinoma in situ of the breast: a nationwide study of 4901 patients in the Netherlands. *Breast Cancer Res Treat* 2019;174:479–88.
- 31 van Dooijeweert C, van Diest PJ, Willems SM, *et al.* Significant Inter- and intra-laboratory variation in grading of invasive breast cancer: A nationwide study of 33,043 patients in the Netherlands. *Int J Cancer* 2020;146:769–80.