

The use of artificial intelligence for enhanced detection of urothelial bladder cancer amongst cystoscopy and urine cytology tests: a systematic review and meta-analysis

Matthew Feyissa¹, Alex Ng¹, Kimberley Chan¹, Buraq Ahmed², Peter Davies¹, Mohammad Alomari¹, Yiannis Philippou¹, Muheilan Muheilan³, Jennifer Linehan⁴, Clayton Lau⁵, Niyati Lobo⁶, Yuhong Yuan^{7,8,9}, Jeremy Teoh^{10,11}, Nikhil Vasdev^{1,12}

¹ Hertfordshire and Bedfordshire Urological Cancer Centre, Department of Urology, Lister Hospital, East and North Hertfordshire NHS Teaching Trust, Stevenage

² Cambridge University, School of Clinical Medicine, Cambridge

³ Urology Department, The James Cook University Hospital, Middlesbrough

⁴ Urology Department, Providence Saint John's Cancer Institute, Santa Monica

⁵ Urology Department, City of Hope Hospital, Los Angeles

⁶ Urology department, Addenbrookes Hospital, Cambridge NHS Teaching Trust, Cambridge

⁷ Department of Medicine, Western University, London, Ontario

⁸ Department of Medicine, London Health Science Centre, London, Ontario

⁹ Department of Medicine, McMaster University, Hamilton, Ontario

¹⁰ Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong

¹¹ Department of Urology, Medical University of Vienna, Vienna,

¹² University of Hertfordshire Medical School, Hertfordshire

Word Count: 2556 words

Author Contribution: All authors were involved significantly in the makeup of this manuscript. MF, YP, MM, NV researched literature and conceived the study idea alongside all listed authors in this review, MF, AN, BA, PD carried out data extraction. PD, AN and BA carried out quality assessment. MF and AN carried out statistical analysis with the guidance of YY, JT, and NV. MF, AN, KC, BA, PD, MA, YP and MM wrote the manuscript. JL, CL, NL, YY, JT, NV reviewed, edited the manuscript and prepared it for submission.

Declaration of interests: No conflict of interest to declare

Data availability statement: All data used is made available with publication

Ethics: No ethical approval required

Funding: No funding sources

Corresponding Author: Prof Nikhil Vasdev, Urology Consultant, Associate Director of Robotic and Cancer Services, East and North Herts NHS Trust, Stevenage (nikhil.vasdev@nhs.net)

Level of Evidence: 2a

Guarantor: Prof Nikhil Vasdev

Abstract:

Objective: Bladder cancer remains a significant burden on healthcare systems worldwide. The aim of this review is to evaluate the diagnostic performance of artificial intelligence (AI) against conventional first line methods (cystoscopy and urine cytology) for bladder cancer.

Methods: A PROSPERO-registered (CRD420261291622) systematic review and meta-analysis. Studies were included if they assessed AI performance in definitive urothelial carcinoma detection via cystoscopy or urine cytology against a non-AI human comparator. Bivariate random-effects meta-analysis was performed to assess diagnostic performance with AUC calculated from HSROC curves.

Results: Nine studies were included (six cytology, three cystoscopies; 8,918 data points). AI demonstrated statistically significant greater sensitivity (0.927 vs 0.754), with a lower negative likelihood ratio (0.087 vs 0.254), suggesting stronger 'rule-out' performance. However, this came at the cost of higher false positives compared to conventional methods. Conventional methods (cystoscopy/cytology) demonstrated higher specificity (0.968 vs 0.841) and positive likelihood ratio (23.275 vs 5.849), reflecting stronger rule-in capability.

Conclusion: AI algorithms potentially have enhanced screening performance for bladder cancer compared to first line modalities. The utilisation of a hybrid model may improve outcomes and efficiencies. However, large-scale, prospective trials with standardised reporting and histological reference standards are required before AI can safely and equitably be deployed.

Keywords: bladder cancer, artificial intelligence, cystoscopy, urine cytology, diagnostics, meta-analysis

Level of Evidence: 2a

Background:

Bladder cancer remains a significant burden on healthcare systems worldwide, with approximately 573,000 new cases and 213,000 deaths reported annually.¹ Urothelial carcinoma carries a high recurrence rate, necessitating ongoing surveillance, which is the principal cost driver related to the disease.^{2,3} These features underscore the critical importance of accurate and timely diagnosis, which directly influences treatment, outcomes, and healthcare resource allocation.

In the United Kingdom, the standard diagnostic pathway is governed by NICE guideline NG12, which mandates urgent suspected cancer pathway referral for patients aged >45 with unexplained visible haematuria, or >60 years old with unexplained non-visible haematuria in combination with dysuria or a raised white cell count.³ Following referral, the NHS England Bladder Best Practice Timed Pathway directs patients towards flexible cystoscopy as the primary investigative tool.⁴ The role of urine cytology in bladder cancer diagnostic pathways is somewhat controversial with differing views globally relating to its use. In the UK, British Association of Urological Surgeons (BAUS) and NICE have removed its previous recommendation due to poor diagnostic yield.^{3,4} The European Association of Urology (EAU) and American Urological Association advises that urine cytology should be reserved for high-grade bladder cancer and not be used routinely in the various haematuria screening pathways for bladder cancer.^{5,6}

Where cystoscopy appearances are suspicious or histological confirmation is obtained at transurethral resection of bladder tumour (TURBT), further staging with CT of the chest, abdomen and pelvis occurs. This reflects an endoscopic-first approach to bladder cancer diagnosis that is distinct from the cross-sectional imaging-led paradigms evaluated in the majority of existing artificial intelligence (AI) literature.

Despite a growing body of evidence examining AI applications in bladder cancer, existing reviews haven't carried out meta-analyses specifically relating to cystoscopy and cytology screening. Yuan et al concluded superior diagnostic performance of AI over clinicians in detecting and staging bladder cancer; however, their meta-analysis was restricted exclusively to CT, MRI, and ultrasound imaging, enrolling 22 studies (3,176 patients) with a primary focus on distinguishing muscle-invasive from non-muscle-invasive disease.⁷

To the best of our knowledge, this review is the first to perform a combined systematic review and meta-analysis specifically investigating AI performance in both cystoscopy and urine cytological bladder cancer detection, benchmarked against a non-AI human comparator.

Methods:

The project study protocol was pre-registered on PROSPERO (CRD420261291622).⁸ OVID Medline, EMBASE, PubMed, Google Scholar, and CENTRAL were searched for relevant results up to February 2026. Full search strategy and terms can be found in supporting material, alongside PICO criteria. For this review, we extracted data from clinical trials and observational studies which assessed the performance of Artificial intelligence models in the detection of bladder cancer using cystoscopic images or histopathological slides. For this review, we classify urothelial bladder carcinoma as both non-muscle-invasive bladder cancer (NMIBC) and/or muscle-invasive bladder cancer (MIBC). Studies which assessed the ability of AI to differentiate between the NMIBC and MIBC rather than definitively assessing the presence of cancer were excluded.

Bias assessment:

Two authors (PD and AN) judged the methodological quality and risk of bias using the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool.⁹ The standard QUADAS-2 signalling questions were supplemented with additional questions specific to AI-based diagnostic accuracy studies, reflecting methodological considerations unique to machine learning evaluation. Studies were not excluded from the review on the basis of their methodological quality assessment; rather QUADAS-2 results were used to contextualise the interpretation of diagnostic accuracy findings and to identify methodological strengths/limitations across the included studies.

Data Extraction & Outcomes

Four authors (MF., PD., AN., and BA) extracted data relating to study characteristic, AI algorithm, conventional method of diagnosis, reference standard, total data points, True positive (TP), False negative (FN), True negative (TN), and False positive (FP) performance of diagnostic tests. All the data extracted by authors was taken from 'test data' rather than 'developmental data' of AI models if studies reported it. The outcome of this review was to assess AI diagnostic performance compared to their conventional cystoscopy or histopathological equivalents.

Statistical Analysis:

MF and AN (in agreement with senior authors NV., JT., and YY.), performed statistical analysis using a bivariate random-effects model implemented in MetaBayesDTA v.1.5.0. to calculate pooled sensitivity, specificity, diagnostic odds ratio, false positive and negative rates, 95% confidence intervals (CIs) and generate SROC curves.^{10,11}

Area under the summary receiver operating characteristic curve (AUC) was carried out using principles outlined by Drevon et al.¹² AUC typically ranges between 0.5 to 1, with a value of 1 suggesting perfect accuracy with discrimination of disease and non-disease, and a value of 0.5 suggesting poor diagnostic accuracy capabilities. Funnel plots and Cochran's Q and I^2 tests were used to assess risk of publication bias and heterogeneity respectively. Significant heterogeneity was defined as $I^2 > 50\%$.

Results:

A total of 1094 studies were identified via database searches (Embase 453, CENTRAL 450, PubMed/Medline 191) and screening of reference lists included in associated and retained studies (**Figure 1**). 3 unique studies were manually obtained from Google Scholar. 631 studies were removed for duplication and a further 426 papers excluded on screening of titles and abstracts. 5 unique studies were manually added to full text screening from wider search of the current literature base. In total, 35 papers underwent full-text analysis, and 26 studies were excluded for the following reasons: incomplete study, duplication, incorrectly described outcomes, wrong study design. A total of nine (n=9) studies were included in meta-analysis to investigate the diagnostic ability of artificial intelligence to diagnose

bladder cancer. Out of the 9 studies included, six (n=6) investigated the use of AI in interpreting urine cytology.^{13–18} Three (n=3) studies investigated images taken from cystoscopy.^{19–21}

In total 8,918 different data points/cases on artificial intelligence (4,281) and traditional diagnostic modalities (4,637) were extracted from the included studies for meta-analysis. Where possible ‘test’ rather than ‘developmental’ data for artificial intelligence was extracted. **Table 1** includes details of included studies; country of origin, publication date, study type, conventional diagnostic modality, AI algorithms, population size. Liu TJ et al reported data for both Whole Slide Imaging (WSI) and Microscopy therefore authors extracted both sets of reported data; these were tested on the same cohort of tumours so acted as a viable comparison for analysis.¹⁷

Two (n=2) studies which initially passed abstract screening were excluded from this review article due to investigating only the grading of bladder cancer (muscle invasive vs non-muscle invasive) and in one instance substituting low grade bladder cancer for a ‘normal’ histological diagnosis.^{22,23} Liu et al, contained some level of comparative data from both upper and lower tract tumours however did not report explicitly the performance of AI in detecting bladder cancer specifically and therefore was subsequently excluded.²⁴

Bias assessment:

The methodological quality of the included studies was assessed using the QUADAS-2 tool (**Figure 2**) across 9 study entries.⁹ Risk of bias was variable across all domains. Patient selection was the most problematic, with the majority of entries (7/9, 78%) rated as high risk of bias. The index test and reference standard domains were each rated as high risk in four entries (44%), with a further two entries (22%) rated as unclear in each. Flow and timing was evenly split, with four entries (44%) rated as high risk and four as low, with one unclear. Applicability concerns were also notable: patient selection was the most affected, with over half of entries (5/9, 56%) rated as high concern. Reference standard applicability had four entries (44%) rated as high concern, though the majority remained low (5/9, 56%). Index test applicability was the least concerning, with two-thirds of entries (6/9, 67%) rated as low concern.

Meta-Analysis:

Sensitivity and Specificity for individual studies is presented below (**Table 2 & 3, Figure 3 & 4**). Bi-variate analysis (**Table 4**) calculated a pooled sensitivity for AI diagnostics (0.927 (0.871-0.960)) which was significantly higher than conventional diagnostic methods (0.754 (0.633-0.845)). Pooled specificity revealed for AI diagnostics (0.841 (0.718-0.917)) was inferior (however not significantly) to conventional modalities (0.968 (0.856-0.993)) as confidence intervals did overlap.

The random effects model used to construct the HSROC curves (**Figure 5**) suggests the use of AI algorithms overall had a greater pooled sensitivity and overall stronger discriminatory performance when diagnosing bladder cancer compared to conventional modalities assessed within the same studies. This is reinforced by the higher (but insignificant) calculated AUC for AI 0.95 (95% CIs – 0.93 to 0.96) versus AUC of 0.89 (95% CIs – 0.86 to 0.93) for conventional modalities (**Table 4**).

However, conventional modalities (cystoscopy/cytology) showed a greater ability to 'rule in' disease as presented by a higher overall diagnostic ratio (91.706 (18.324-458.959)) compared to AI (67.512 (17.623-258.638)), a higher specificity and higher +ve likelihood ratio (23.275 vs 5.849). In spite of this, diagnostic AI algorithms may have a better ability to 'rule out' disease due to a significantly higher sensitivity and lower false negative rate (0.032 (0.007-0.144)), paired with a lesser –ve likelihood ratio (0.087 (0.043-0.174)) compared to conventional modalities (0.254 (0.165-0.389)). Visual examination of the funnel plot (**Figure 6**) found no evidence of publication bias. As expected, heterogeneity between studies (I^2 : 85-96.82%) was significantly high (**Table 4**).

Discussions:

This review to our knowledge, is the first joint diagnostic evaluation of AI against cystoscopy and urine cytological bladder cancer detection. The reference standard, AI and 'control' diagnostic performance was judged against varied from TURBT biopsy samples to expert cytotechnologist/histopathologist/urologist consensus panel opinion of cystoscopy or urine cytological slides (**Table 1**).

Our review suggests that the use of artificial intelligence appears to have the most potential in screening pathways as a 'rule out' test rather than a true diagnostic tool due to the presence of a significantly lower false negative rate and higher sensitivity, in addition to a lesser (but insignificant) -ve likelihood ratio. This pattern is consistent with AI systems operating at a lower classification threshold than conventional methods, shifting the operating point 'up and right' on the ROC curve to maximise true positive detection at the cost of an increased false positive rate, a trade-off that may be clinically acceptable in a screening context where missed cancers carry greater harm than unnecessary follow-up.²⁵

Despite having a higher AUC, due to having a 'reduced' pooled specificity, higher +ve likelihood ratio and lower diagnostic odds ratio there are limitations in suggesting AI as a definitive 'rule in' diagnostic tool for bladder cancer over already established conventional modalities. However, taken together, a hybrid diagnostic model could be utilised, in which AI functions as a first-pass screening tool to reduce missed diagnoses, with conventional expert review retained for disease confirmation. This is particularly useful amongst 'lesser' trained individuals and also healthcare systems burdened with a large volume waiting lists due to superior efficiency. Wu et al. demonstrated that AI augmentation of cystoscopy produced the greatest sensitivity gains in less experienced clinicians: trainees improved from 50.8% to 73.1% with CAIDS guidance, whilst expert urologists improved from 75.4% to 90.8%, with a system latency of just 12 seconds versus 35–45 minutes for human review.²⁰ In the cytological domain, Kaneko et al. evaluated 953 cell images in approximately 28 seconds, supporting the efficiency advantage of automated cytology pipelines.¹⁴

The majority of AI study design is retrospective in nature with non-consecutive sampling with tumour-enriched case-control datasets.^{13,14,17,18,20,21,23} This likely explains the significant performance of AI algorithms to 'rule out' bladder cancer. This reflects in the QUADAS-2 assessment which identified high risk of bias in patient selection, which is an inherent feature of AI development pipelines, whereby training cohorts must over-represent positive cases which provides an explanation behind a higher sensitivity across AI algorithms in this review. In this review, only Lebre et al. contributed prospective data.^{15,16} This absence of large volume prospective evidence substantially limits the definitive real-world conclusions that can be drawn relating to AI performance. This mirrors findings from He C et al.,

who reported low methodological quality and high risk of bias as near-universal features of AI bladder cancer diagnostic studies specifically.²⁶ This bias is structurally inherent to AI development: 'training datasets must be enriched with positive cases to produce functional models, making them non-representative of true clinical prevalence and simultaneously generating population, outcome, and publication bias that inflates apparent performance', a problem Jayakumar et al. (2022) identified as field-wide rather than study-specific.²⁷ Current performance estimates should therefore be treated as exploratory upper bounds rather than reliable predictors of real-world clinical utility.

Limitations of this review:

A material structural limitation affects interpretation of the pooled specificity advantage for conventional methods. It is worth noting first that the comparatively smaller number of studies eligible for this meta-analysis reflects strict inclusion criteria: definitive cancer detection with a clearly defined control and reference standard; explicitly excluding grading, staging, and molecular subtyping studies such as Kurata et al. and Tsuji et al.^{22,23} Studies focused on grading address a clinically important but pooling such outcomes with detection data would introduce unacceptable heterogeneity. This stance is consistent with the STREAM-URO framework (STandardized Reporting of Applications of Machine learning in UROlogy), a 26-item instrument mapped to the TRIPOD statement, which highlights the importance of clearly defining the ML problem prior to evaluating AI diagnostic tools.²⁸

A further structural bias affects the specificity estimates. In AI augmentation studies where the human reader remains the final arbiter, AI specificity is intrinsically constrained by the human reader's own threshold and cannot exceed it. This ceiling effect is starkest in the VISIOCYT1 studies, where enrolment was restricted to patients with negative standard cytology at baseline, making it structurally impossible for conventional cytology to return a false positive within that cohort.¹⁵ The resulting 100% control specificity is an artefact of study design rather than a reflection of cytological performance, artificially inflating the conventional specificity advantage in this meta-analysis. The pooled specificity difference between AI and conventional methods should therefore be interpreted with this systematic distortion in mind, and the true AI specificity disadvantage is possibly smaller than the data suggest. It

is likely this is a common theme across artificial intelligence studies that extends beyond this review article.

Challenges to implementing AI:

Significant barriers to widespread clinical adoption remain before AI can be responsibly integrated into routine NHS bladder cancer diagnostics. The infrastructure required to deploy deep learning systems in endoscopy suites/cytopathology laboratories carries substantial capital and ongoing maintenance costs likely to be prohibitive for many NHS trusts.²⁹ The computational demands of real-time inference models e.g. CAIDS and CystoNet, raise important environmental considerations. Large-scale deployment of GPU-intensive AI platforms generate significant energy consumption and carbon emissions, a factor that sits in direct tension with the NHS's net-zero commitments under the Greener NHS programme.^{30,31}

Equity of access is a further concern, as institutions with the financial and technical capacity to implement these systems are likely to be large academic centres, potentially widening rather than narrowing diagnostic disparities between tertiary and district general hospitals.³² The regulatory landscape for AI medical devices in the UK (governed by the MHRA) remains evolving, and no AI cystoscopy or cytology tool currently holds full regulatory clearance for routine clinical use.³³ Finally, the 'black-box' nature of many deep learning models limits clinician trust and interpretability and raises medico-legal questions regarding accountability when AI-assisted diagnoses are incorrect.³⁴

Conclusions:

AI algorithms potentially have enhanced screening performance for bladder cancer compared to first line modalities. The utilisation of a hybrid model may improve outcomes and efficiencies. However, large-scale, prospective trials with standardised reporting and histological reference standards are required before AI can safely and equitably be deployed.

Author Contribution: All authors were involved significantly in the makeup of this manuscript. MF, YP, MM, NV researched literature and conceived the study idea alongside all listed authors in this review, MF, AN, BA, PD carried out data extraction. PD, AN and BA carried out quality assessment. MF and AN carried out statistical analysis with the guidance of YY, JT, and NV. MF, AN, KC, BA, PD, MA, YP and MM wrote the manuscript. JL, CL, NL, YY, JT, NV reviewed, edited the manuscript and prepared it for submission.

Conflicting interests: No conflict of interest to declare

Data availability statement: All data used is made available with publication

Ethical approval: No ethical approval required

Funding: No funding sources

Corresponding Author: Prof Nikhil Vasdev, Urology Consultant, Associate Director of Robotic and Cancer Services, East and North Herts NHS Trust, Stevenage (nikhil.vasdev@nhs.net)

Guarantor: Prof Nikhil Vasdev

Acknowledgements: n/a

Trial registration: n/a

Bibliography:

1. Su X, Tao Y, Chen F, et al. Trends in the global, regional, and national burden of bladder cancer from 1990 to 2021: an observational study from the global burden of disease study 2021. *Sci Rep* 2025; 15: 7655.
2. Cox E, Saramago P, Kelly J, et al. Effects of Bladder Cancer on UK Healthcare Costs and Patient Health-Related Quality of Life: Evidence From the BOXIT Trial. *Clin Genitourin Cancer* 2020; 18: e418–e442.
3. Overview | Bladder cancer: diagnosis and management | Guidance | NICE, <https://www.nice.org.uk/guidance/ng2> (2015, accessed 30 March 2026).
4. England NHS. NHS England » Implementing timed urology cancer diagnostic pathway – bladder, penile, renal and testicular,

read/implementing-timed-urology-cancer-diagnostic-pathway-bladder-penile-renal-and-testicular/ (accessed 30 March 2026).

5. **Diagnosis and Treatment of Non-Muscle Invasive Bladder Cancer: AUA/SUO Guideline: 2024 Amendment | Journal of Urology. *J Urol*, <https://www.auajournals.org/doi/10.1097/JU.0000000000003846> (2024, accessed 26 May 2026).**
6. **Babjuk M, Burger M, Capoun O, et al. European Association of Urology Guidelines on Non-muscle-invasive Bladder Cancer (Ta, T1, and Carcinoma in Situ). *Eur Urol* 2022; 81: 75–94.**
7. **Yuan S, Yang K, Wu H, et al. Diagnostic performance of artificial intelligence in detecting bladder carcinoma. *World J Urol* 2025; 44: 63.**
8. **PROSPERO, <https://www.crd.york.ac.uk/PROSPERO/view/CRD420261291622> (accessed 30 March 2026).**
9. **QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies | Annals of Internal Medicine, <https://www.acpjournals.org/doi/10.7326/0003-4819-155-8-201110180-00009> (accessed 30 March 2026).**
10. **MetaBayesDTA v1.5.3, <https://crsu.shinyapps.io/MetaBayesDTA/> (accessed 30 March 2026).**
11. **Cerullo E, Sutton AJ, Jones HE, et al. MetaBayesDTA: codeless Bayesian meta-analysis of test accuracy, with or without a gold standard. *BMC Med Res Methodol* 2023; 23: 127.**
12. **Drevon D, Fursa SR, Malcolm AL. Intercoder Reliability and Validity of WebPlotDigitizer in Extracting Graphed Data. *Behav Modif* 2017; 41: 323–339.**
13. **Awan R, Benes K, Azam A, et al. Deep learning based digital cell profiles for risk stratification of urine cytology images. *Cytometry A* 2021; 99: 732–742.**
14. **Kaneko M, Tsuji K, Masuda K, et al. Urine cell image recognition using a deep-learning model for an automated slide evaluation system. *BJU Int* 2022; 130: 235–243.**
15. **Lebret T, Pignot G, Colombel M, et al. Artificial intelligence to improve cytology performances in bladder carcinoma detection: results of the VisioCyt test. *BJU Int* 2022; 129: 356–363.**
16. **Lebret T, Paoletti X, Pignot G, et al. Artificial intelligence to improve cytology performance in urothelial carcinoma diagnosis: results from validation phase of the French, multicenter, prospective VISIOCYT1 trial. *World J Urol* 2023; 41: 2381–2388.**
17. **Liu T-J, Yang W-C, Huang S-M, et al. Evaluating artificial intelligence-enhanced digital urine cytology for bladder cancer diagnosis. *Cancer Cytopathol* 2024; 132: 686–695.**
18. **Ou Y-C, Tsao T-Y, Chang M-C, et al. Evaluation of an artificial intelligence algorithm for assisting the Paris System in reporting urinary cytology: A pilot study. *Cancer Cytopathol* 2022; 130: 872–880.**
19. **Kim J, Ham WS, Koo KC, et al. Evaluation of the Diagnostic Efficacy of the AI-Based Software INF-M01 in Detecting Suspicious Areas of Bladder Cancer Using Cystoscopy Images. *J Clin Med* 2024; 13: 7110.**

20. Wu S, Chen X, Pan J, et al. An Artificial Intelligence System for the Detection of Bladder Cancer via Cystoscopy: A Multicenter Diagnostic Study. *JNCI J Natl Cancer Inst* 2021; 114: 220–227.
21. Yang R, Du Y, Weng X, et al. Automatic recognition of bladder tumours using deep learning technology and its clinical application. *Int J Med Robot* 2021; 17: e2194.
22. Tsuji K, Kaneko M, Harada Y, et al. A Fully Automated Artificial Intelligence System to Assist Pathologists' Diagnosis to Predict Histologically High-grade Urothelial Carcinoma from Digitized Urine Cytology Slides Using Deep Learning. *Eur Urol Oncol* 2024; 7: 258–265.
23. Kurata Y, Nishio M, Moribata Y, et al. Development of deep learning model for diagnosing muscle-invasive bladder cancer on MRI with vision transformer. *Heliyon* 2024; 10: e36144.
24. Liu Y, Jin S, Shen Q, et al. A Deep Learning System to Predict the Histopathological Results From Urine Cytopathological Images. *Front Oncol* 2022; 12: 901586.
25. Nahm FS. Receiver operating characteristic curve: overview and practical use for clinicians. *Korean J Anesthesiol* 2022; 75: 25–36.
26. He C, Xu H, Yuan E, et al. The accuracy and quality of image-based artificial intelligence for muscle-invasive bladder cancer prediction. *Insights Imaging* 2024; 15: 185.
27. Jayakumar S, Sounderajah V, Normahani P, et al. Quality assessment standards in artificial intelligence diagnostic accuracy systematic reviews: a meta-research study. *Npj Digit Med* 2022; 5: 11.
28. Kwong JCC, McLoughlin LC, Haider M, et al. Standardized Reporting of Machine Learning Applications in Urology: The STREAM-URO Framework. *Eur Urol Focus* 2021; 7: 672–682.
29. Fair and equitable use of AI in healthcare - Moorfields Eye Hospital, <https://www.moorfields.nhs.uk/research/nihr-moorfields-clinical-research-facility-crf/news-crf/fair-and-equitable-use-of-ai-in-healthcare> (accessed 30 March 2026).
30. Vafaei Sadr A, Bülow R, Von Stillfried S, et al. Operational greenhouse-gas emissions of deep learning in digital pathology: a modelling study. *Lancet Digit Health* 2024; 6: e58–e69.
31. NHS G. Greener NHS » Delivering a net zero NHS, <https://www.england.nhs.uk/greenernhs/a-net-zero-nhs/> (accessed 30 March 2026).
32. Chan SCC, Neves AL, Majeed A, et al. Bridging the equity gap towards inclusive artificial intelligence in healthcare diagnostics. *BMJ* 2024; 384: q490.
33. Medical devices: software and artificial intelligence (AI). *GOV.UK*, <https://www.gov.uk/government/publications/software-and-artificial-intelligence-ai-as-a-medical-device> (2025, accessed 30 March 2026).
34. Stogiannos N, Malik R, Kumar A, et al. Black box no more: a scoping review of AI governance frameworks to guide procurement and adoption of AI in medical imaging and radiotherapy in the UK. *Br J Radiol* 2023; 96: 20221157.