BJU Int 2026; **137:** 596–604 doi:10.1111/bju.70164

BJUI
BJU International

Review

Digital pathology-based artificial intelligence algorithms in prostate cancer: inside the 'black box'

Claire M. de la Calle[1,2], Alexander S. Baras[3] and Tamara L. Lotan[3,4,5]

[1]*Department of Urology, University of Washington,* [2]*Fred Hutchinson Cancer Center, Seattle, WA,* [3]*Department of Pathology, Johns Hopkins University,* [4]*The James Buchanan Brady Urological Institute, and* [5]*Department of Oncology, Johns Hopkins University School of Medicine, Baltimore, MD, USA*

Artificial intelligence (AI) algorithms leveraging digital pathology slides are currently transforming the way urological cancers are diagnosed and graded, and they add additional prognostic, predictive and molecular subtyping information beyond traditional pathological risk stratification. This review explores recent advances in histopathology-based AI systems for prostate cancer. We examine how these algorithms perform relative to pathologists for tumour diagnosis and grading, and the ways in which they surpass pathologists with respect to reducing inter-observer variability and providing quantified tumour metrics. We particularly focus on prognostic algorithms that have been benchmarked against 'gold standard' patient outcomes such as metastasis or death, and we highlight the emerging role of digital pathology-enabled AI for predicting response to therapy or underlying tumour molecular alteration status. Finally, we touch on the advantages of, and barriers to, implementation of digital pathology and histopathology-based AI algorithms in clinical practice. Through this synthesis of current literature, we underscore the emerging potential of AI for standardising pathological assessment, guiding clinical management, and improving patient outcomes in prostate cancer.

Keywords

artificial intelligence, pathology, prediction, prognosis, prostate cancer, whole slide images

Introduction

Tremendous progress has been made in artificial intelligence (AI) assisted digital pathology interpretations in the last decade. Numerous AI algorithms have been developed to rapidly and efficiently identify cancers on digitised haematoxylin and eosin (H&E) whole-slide images, grade cancers with greater accuracy than human pathologists, prognosticate oncological outcomes, predict response to therapy, and identify molecular subtypes not usually identifiable without immunohistochemistry or tumour sequencing. Among all urological tumour types, prostate cancer has provided an early test case for application of these AI-augmented digital pathology tools, and numerous algorithms have advanced from training and validation to regulatory clearance and clinical use. In this non-systematic review, we describe the current state of the research in this area, focusing on the role of AI-assisted prostate cancer diagnostics, oncological prognostication, and predictive response to therapy.

Diagnostic AI Algorithms for Prostate Cancer

With over a million prostate biopsies performed each year in the United States [1], this specimen type represents a large portion of urological pathologist's workload. Given many of these biopsies are negative for cancer, pathologists spend a significant of amount of time screening large numbers of benign slides, and this is compounded by the fact that the global pathologist workforce is known to be dwindling [2]. Accordingly, one of the earliest and best developed applications of AI-based digital pathology algorithms was the identification of prostate cancer (and discrimination from benign prostate tissue) on H&E slides. Litjens et al. [3] reported on one of the early prostate cancer-detecting algorithms and estimated that if pathologists skipped all the slides without cancer, this would eliminate ~30% of biopsy slides needing review. Many more algorithms have been developed in various cohorts, including large independent cohorts with external validation [4,5]. These algorithms have repeatedly been shown to have excellent discrimination between cancer and benign tissue, with areas under the curve (AUCs) nearing 1.0 in most studies, again confirming that digital reads could be used as a pre-screening tool significantly reducing the workload of the human pathologist [3,6,7]. Accordingly, the commercially available Paige Prostate algorithm (Paige.AI Inc., New York, NY, USA) was among the first approved by several governmental health agencies in the United States and Europe for use as a clinical tool to identify cancer-containing prostate biopsy cores in concert with a pathologist's review [8,9]. Ultimately, triaging negative

biopsies is likely to be the most widely utilised initial application of AI algorithms in urological pathology clinical practice.

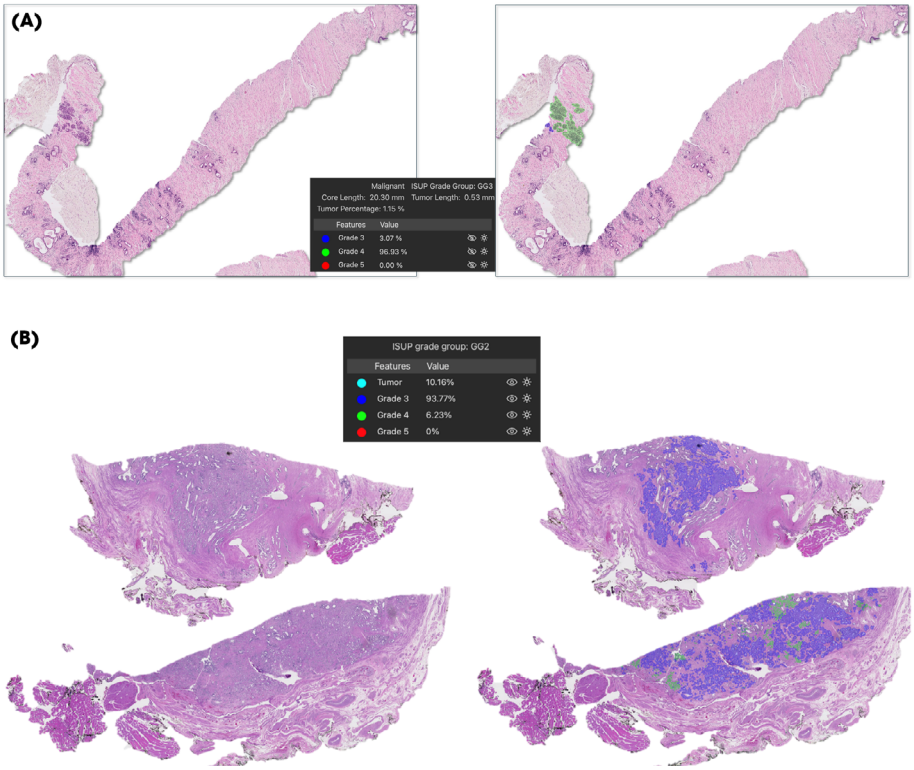
The next natural step was to develop algorithms that output a Gleason Grade Group (GG). Given the considerable intra- and inter-reader variability in Gleason grading [10–12], AI algorithms offer the opportunity to potentially streamline the grading process and reduce this variability. The first algorithms reported were generally trained on smaller cohorts using a single pathologist as the 'ground truth' comparator [13–17]. Yet, AI algorithms can only be as good as the training sets with which they are developed. As such, defining the 'ground truth' comparator with a majority or consensus vote from a large panel of expert genitourinary pathologists is essential [18], and ideally, standardised methodologies are employed for cases in which pathologists disagree. Using a consensus human pathology read as the comparator helps increase agreement between the AI read and the human read. For example, Nir et al. [19] generated multiple models, trained on several individual pathologists and also generated a model trained on a majority vote of all the pathologists with whom they worked. Agreement between the individual pathologists and the AI model ranged from 0.38 to 0.58 (quadratic weighted kappa agreement). However, using the majority vote of all the pathologists the agreement improved to 0.60. Ideally the databases used for AI algorithm training are also large enough to represent the full spectrum of Gleason grades, including the rarer very high grades and unusual histological patterns. It is equally important to use cohorts from multiple institutions to assess the generalisability of the algorithm, given inter-institutional variability in grading and H&E quality.

With the goal of developing highly reliable AI algorithms, in 2020, the Prostate cANcer graDe Assessment (PANDA) challenge consortium made publicly available 12 625 whole-slide images of prostate biopsies originating for several countries and tasked competitors to develop an AI algorithm for Gleason grading. On external validation, the algorithms achieved high agreement with panels of expert genitourinary pathologists (quadratic kappa agreement ~0.8) [20]. A few years later, Faryna et al. [21] compared the top five performing algorithms from the PANDA challenge – all of which are publicly available – with two commercially available algorithms Paige Prostate and AIRA Matrix (Thane, Maharashtra, India). They showed that, in a biopsy cohort generated via crowdsourcing thought to represent real-world variabilities, the commercially available algorithms numerically outperformed the publicly available algorithms by comparison to human pathologist reads; however, the highest performing publicly available algorithms performed quite well. The highest quadratic weighted kappa for agreement between the AI algorithm and the majority vote human pathology read was 0.9 for the AIRA Matrix algorithm, but this was not statistically significantly better than the Paige Prostate algorithm or some of the publicly available ones [21]. Overall, the agreements of the AI algorithms vs human pathologists reads reported in these direct comparisons [20,21], and in other studies [4,6,22–25], are higher than what we expect when comparing among most general pathologists. In fact, in a study conducted by Google Health, Steiner et al. [26] found that the concordance between general pathologists and genitourinary pathologists increased from 69.7% (without AI assistance) to 75.3% (with AI assistance). Interestingly, AI assistance also reduced mean review time and improved pathologist confidence. Similarly, Nagpal et al. [27] demonstrated that the AI algorithm performed better than general pathologists and performed similarly to genitourinary pathologists, suggesting that these AI algorithms can offer expert pathology reads in healthcare settings without subspecialists, potentially addressing inequities in healthcare access. Of course, it is also critical that algorithms be trained on diverse populations from community settings to ensure generalisability. In addition to a slide-level grade, some algorithms also output gland-level heatmaps of the AI grading that can be overlayed with the H&E slide in order to help the human pathologists rapidly identify the areas of most concern, confirm their own reads, and assist in pathologist training (Fig. 1) [1]. The glandular-level grading annotations can also help the human pathologist understand false positive or false negative AI reads, which can occur due to slide artefact, or areas with high-grade prostatic intraepithelial neoplasia or atypical small acinar proliferations.

Another advantage these AI algorithms offer over human reads is more objective and accurate cancer volume metrics, such as cancer length and exact percentage of each Gleason pattern (Fig. 1) [6]. Given the increasing focus on millimetres of Gleason pattern 4 as the most critical prognostic parameter within Gleason grading [28], this detailed quantification may be an important motivation for widespread adoption of slide digitisation and AI grading modules. Such algorithms can also identify and quantify or measure cribriform pattern 4 cancer [29,30]. Ambrosini et al. [30] developed an AI algorithm to identify cribriform histology, with an AUC of 0.81; however, a significant number of false positive cases were noted. The presence of larger cribriform glands is particularly associated with poor oncological outcomes, and most studies have settled on a 200 or 250 μm diameter cutoff for cribriform lesions that is difficult to measure visually [31,32], thus this automated read could add significantly to qualitative pathologist interpretations.

Overall, diagnostic AI algorithms developed for prostate cancer have been shown to reduce the subjective component of Gleason grading and decrease its variability. These algorithms also have the potential to streamline and expedite the histopathology slide reading process and provide expert

Fig. 1 Digital pathology AI algorithms for prostate cancer identification and grading can provide gland-level tumour annotations with precise quantifications of Gleason patterns and tumour extent. (**A**) Example of an AI-annotated prostate cancer biopsy (adapted from Reference [37]). (**B**) Example of an AI-annotated radical prostatectomy sample. ISUP, International Society of Urological Pathology.



grading in underserved areas. They also might be able to provide more refined risk stratifications by allowing more objective and precise quantification of the amount of each Gleason pattern, cancer volume metrics, and cribriform histology, potentially making these diagnostic AI algorithms useful prognostic tools as well.

Prognostic AI Algorithms for Prostate Cancer

The AI algorithms developed for the purpose of providing an AI-generated Gleason grade have also been investigated as prognostic tools. Using tissue from radical prostatectomy specimens, models for the prediction of metastasis based on an AI Gleason grade had similar AUCs to models incorporating pathologist-defined Gleason grade, providing reassurance that these AI Gleason grading algorithms adequately risk stratify patients for oncological outcomes [33]. Others have found that Gleason scoring AI algorithms might risk stratify patients better for recurrence [34] and improve prostate cancer-specific mortality predictions post-prostatectomy [35]. These algorithms might also provide more refined Gleason pattern 4 detection [36], which could be impactful for active surveillance patients with favourable-risk cancers. In an active surveillance cohort of exclusively GG1 patients (per the human pathologist), one AI algorithm upgraded 40% of the patients on initial biopsy who were ultimately upgraded by the human pathologist later during surveillance, suggesting that the algorithm's grade could help better determine which patients need closer surveillance [37].

It is increasingly recognised that cribriform histology and intraductal carcinoma are key drivers of oncological outcomes, and that the Gleason scoring system has several limitations in addition to inter- and intra-reader variability, such as misclassifying some biologically aggressive cancers as favourable risk, due to not incorporating intraductal carcinoma or measuring the extent of cribriform glands [38]. Leo et al. [39] developed an algorithm that can identify cribriform morphology, and they found that the volume of cribriform cancer identified by the AI algorithm was associated with biochemical recurrence (BCR) after radical prostatectomy, and it was most prognostic in the patients with GG2 disease. This is a significant finding as GG2 is a heterogeneous group, ranging from active surveillance candidates to patients with lethal disease. As such, AI algorithms that focus only on identifying the most aggressive

features of prostate cancer, agnostic of Gleason grade, might have more prognostic value than the AI Gleason grading algorithms.

Prognostic AI algorithms have also been developed and trained with the goal of predicting oncological outcomes, agnostic of Gleason grade or recognised aggressive tumour morphologies. A significant portion of patients undergoing definitive therapy for their localised prostate cancers will experience BCR [40], and BCR is tightly linked to metastases and prostate cancer-specific survival [41,42]. As such, prediction of BCR risk has been a focus of many prognostic algorithms. Using tissue microarrays obtained from radical prostatectomies, Pinckaers et al. [43] developed an AI digital pathology read trained to predict BCR after prostatectomy. On external validation, the algorithm was independently and strongly associated with BCR, despite adjusting the model for GG, pathological stage and margin status, and the algorithm stratified patients at low vs high risk of BCR reasonably, especially those with early BCR events. Cao et al. [44] also developed AI algorithms to predict BCR after prostatectomy but using prostate biopsy images. They found that the strongest model to predict BCR was the one incorporating digital pathology AI interpretations with clinical information, and on decision curve analysis this model seemed to provide net clinical benefit. Interestingly, they briefly note that in observing heat-maps of the H&E slides showing the areas most likely to contribute to BCR, the signal captured by the AI models seemed to be coming from tumour regions with features known to be associated with worse outcomes (vascular or nerve invasion, necrosis, etc.). However, they also suggested that the gland morphology of the non-cancerous tissue surrounding the cancer was distinct in patients who experienced BCR vs those who did not.

Eminaga et al. [45] developed a new grading system, risk stratifying patients into four risk groups, for the prediction of BCR. The novel grading system seemed to better risk stratify patients than Gleason scoring, but interestingly, there was no collinearity between the AI grading system and the Gleason scoring, highlighting the limitations of Gleason scoring and suggesting that perhaps signal from the tumour microenvironment needs to be considered as well. In a large radical prostatectomy dataset, Hu et al. [46] combined AI-derived radiomic and pathomic signatures, using preoperative prostate MRI data and radical prostatectomy slides and generated a model combining the radiomics, pathomics, and clinical data. The model was able to discriminate for BCR after surgery fairly well with an AUC of 0.860, and this model was superior to the individual components (i.e., radiomics or pathomics or clinical data alone), and also superior to the Cancer of the Prostate Risk Assessment (CAPRA) and CAPRA-postSurgical (CAPRA-S) nomograms. These results and the findings of other groups evaluating multimodal AI models integrating pathomics with radiomics, genomics, and other clinical data [47–49] strongly suggest that multimodal algorithms are likely to have the most clinical utility moving forward.

Leveraging the data from five NRG Oncology (acronym 'NRG' is derived from the names of the three groups: the National Surgical Adjuvant Breast and Bowel Project, the Radiation Therapy Oncology Group [RTOG], and the Gynecologic Oncology Group) randomised clinical trials (trials evaluating radiation therapy with or without androgen deprivation therapy [ADT]), Esteva et al. [50,51] developed a multimodal AI system called MMAI, now commercialised by Artera (Artera Inc., Los Altos, CA, USA). Using biopsy slides and clinical data (age, PSA, Gleason score, clinical T stage) from 5500 patients included in these trials, with prolonged follow-up, the group developed several models to predict BCR, metastases, and prostate cancer-specific mortality at 5- and 10-years after treatment. When compared to the National Comprehensive Cancer Network (NCCN) risk groups, the MMAI outperformed NCCN for each outcome, increasing the AUCs by 0.06 to 0.11. The group performed an ablation study, providing some insight into the added benefit of the AI digital pathology read over other clinical variables. For the prediction of metastases at 5 years, the AUC for the AI digital pathology read alone was 0.779, and this increased to 0.837 with the addition of the clinical variables, i.e., the full MMAI model. In comparison, the NCCN AUC for the same outcome was 0.72 in the validation group. These results are certainly promising, showing that the AI reads do provide additional information to risk stratify patients, in addition to traditional clinicopathological risk stratification; however, additional studies are needed to determine if this incremental benefit is clinically impactful. Future studies will be required to challenge the AI models with better risk stratifiers than NCCN risk groups alone, which has many known limitations. Notably, the trials used in the Esteva et al. studies [50,51], pre-date the MRI era. Comparisons will need to be made with clinical models that include targeted biopsy data, MRI data, PSA density, intraductal carcinoma, cribriform morphology etc., as clinicians and patients already use all of this information for clinical decision making.

The MMAI models from Artera were externally validated in other NRG/RTOG cohorts [52,53], and were shown to also risk stratify well patients who underwent radical prostatectomy [54,55]. In fact, the MMAI algorithm was strongly and independently associated with adverse pathological features and BCR after radical prostatectomy [55]. As a result of these studies, the NCCN has listed the MMAI ArteraAI Prostate as an adjunct tool for advanced risk stratification for all patients with localised prostate cancer [56]. Yet, how these models perform in low-risk patients, otherwise candidates for active surveillance remains to be studied. Given the models were developed for oncological outcomes rarely occurring in patients on active surveillance, it

is possible that they would not perform well in identifying low- or favourable intermediate-risk patients with cancers at risk of progression. Adapted algorithms for this patient population are likely to be needed.

The MMAI models have also been tested in cohorts with very advanced prostate cancer [57–60]. In a study by Wang et al. [57], the MMAI was found to be independently associated with overall survival and time to castration resistance in a cohort of patients with oligometastatic hormone-sensitive prostate cancer. In a cohort of patients with non-metastatic castrate-resistant prostate cancer (CRPC) from the SPARTAN trial (ClinicalTrials.gov identifier: NCT01946204), the MMAI score was associated with worse metastasis-free survival and overall survival [58]. Interestingly, in a cohort exclusively comprised of patients with lethal prostate cancer, 37% of the patients did not have high MMAI scores. This is likely driven at least somewhat by the low Gleason scores for most of these patients with low MMAI scores (~80% of the patients with low MMAI had GG ≤2), highlighting the fact that tissue-based biomarkers are inherently limited by tissue sampling. Further, given the higher prevalence of more aggressive forms of prostate cancer such as neuroendocrine differentiation in advanced disease, adapted AI algorithms trained specifically for patients with CRPC are likely to be more informative in this patient population.

Predictive AI Algorithms for Prostate Cancer

Very few truly predictive AI algorithms (i.e., able to predict response to a specific treatment based on AI digital pathology) have been developed for prostate cancer, and none have been tested as integrated biomarkers in a randomised prospective clinical trial to date. The best-known predictive AI algorithm based on digital pathology is Artera's MMAI, predicting the benefit of short-term ADT when given with radiation therapy in intermediate-risk patients to prevent metastases [61]. Similar to the development of the prognostic Artera MMAI, Spratt et al. [61] used patient data and biopsy slides from >5000 patients enrolled in NRG/RTOG trials to develop a new multimodal AI model and then validate it in another RTOG trial that randomised patients to radiation therapy with or without short-term ADT. The output of the model is either a positive prediction or a negative one. In the validation cohort, patients with a positive model prediction had a 10% reduction in distant metastases at 15 years with the use of ADT, while the patients with a negative model prediction did not demonstrate any significant benefit from ADT use [61]. This model is mentioned in the NCCN guidelines, however additional prospective testing as an integral marker in clinical trials is recommended [56]. Accordingly, the clinical utility of this predictive biomarker is also going to be prospectively studied in the Australian ASTuTE trial (Australian New Zealand Clinical Trials Registry, ACTRN12623000713695p) [62]. The Artera group recently reported on another predictive model that helps determine the need for long- vs short-term ADT for patients with high-risk or regional prostate cancer (MMAI Prostate long-term [LT]-ADT Predictive Model), not currently commercially available or incorporated into NCCN guidelines [63]. With the use of this model, 34% of men with high-risk disease could have avoided long-term ADT. Models predicting the benefit of novel antiandrogens are also likely to become commercially available in the near future [64]. Antiandrogen therapies are not a benign treatment as they associated with many long-lasting side effects [65], thus understanding which patients truly benefit from ADT is important. However, these predictive tests will need to be prospectively studied as integrated biomarkers in a randomised clinical trial before they can be considered to be truly validated for clinical use.

In additional *post hoc* analyses, some prognostic models have shown some promise in their ability to predict response to other treatments [66], such as response to apalutamide for non-metastatic CRPC [58] or metastasis-directed radiation therapy for oligometastatic disease [57] but to our knowledge, Artera's MMAI predicting the benefit of short-term ADT [61] is the only commercially available model. Interestingly, Artera's model is run on only one representative biopsy slide as selected by the human pathologist. In the future, predictive models may be run on all available slides as the cost of this additional service is low and would presumably further reduce sampling artefacts. Further, the imaging features the model is capturing to predict metastasis development are unknown, and we do not know how the Artera predictive model compares to other adjunct tests such as commercially available genomic classifiers regularly used in the clinic to help with decisions regarding ADT use, such as the Decipher score. Of note, other AI models have been shown to outperform the Decipher score in the prognostication of BCR [67,68]; however, to our knowledge such comparisons have not yet been made with the Artera models. These studies will be required to better understand the clinical utility of prognostic and predictive digital pathology AI models moving forward.

Molecular Subtype Classification AI Algorithms

An important emerging application for digital pathology-based AI algorithms is prediction of underlying tumour molecular subtype, or clinically-relevant molecular alteration status, using the diagnostic H&E image. In colorectal cancer, e.g., AI algorithms to predict mismatch repair deficiency (MMRd) from the H&E provide a promising method to screen for patients who may benefit from

immunotherapy [69], and similar work has been done for molecular subtyping in breast cancer [70]. Although the NCCN guidelines recommend germline testing for all patients with high- and very high-risk localised prostate cancer in order to identify potentially targetable germline alterations (including homologous DNA repair deficiency [HRD] or MMRd) [56], testing utilisation rates may be as low as 20% in academic settings, and even lower in population-based studies [71,72]. Somatic sequencing, another important tool to identify targetable alterations, is only currently recommended in the advanced disease setting in prostate cancer, in large part due to expense and complexity of testing. Thus, digital pathology-based AI algorithms that could triage cases at the time of diagnosis for underlying targetable germline and/or somatic genomic alterations in the high-risk setting would provide an additional rapid and inexpensive tool to identify subsets of patients with prostate cancer who could benefit most from confirmatory molecular testing. Accordingly, several AI algorithms have been developed as a proof-of-concept for molecular subtype identification in prostate cancer. Algorithms to identify *ERG* fusions, which are not currently targetable but represent the most common clonal prostate cancer molecular subtype in localised disease, have shown AUCs ranging from 0.73 to 0.89 in external validation cohorts [73–75]. *TP53* mutation [76] or *PTEN* deletion detection [74] have shown slightly decreased performance, potentially due to their frequently subclonal occurrence, which introduces sampling artefacts that may be challenging for AI algorithms. Algorithms predicting microsatellite instability [77], mRNA expression used in the Decipher score [78], and androgen receptor status [79] have also been developed but have variable reliability. To our knowledge, only one *BRCA1/2* mutation detection algorithm has been published to date [80], and additional large cohorts will likely be required to train for detection of this relatively rare, but highly actionable, alteration in primary prostate tumours.

Implementation of Digital Pathology: Impact on Pathology Workflows, Costs and Barriers

While there have been several studies examining the benefits and barriers for implementation of digital pathology in general, fewer studies have examined the real-world impact of AI models on pathology practice. Digital pathology has numerous benefits for pathology laboratory workflows and reporting, including dramatic improvements in archival slide retrieval times afforded by whole-slide images that facilitate comparison of newer patient tumour samples with prior material, and concomitant reduction in slide loss and misfiling. Digital slides are particularly beneficial for laboratories with multiple sites, reducing requirements for physical slide transportation and enabling remote case distribution so that practices can effectively subspecialise by organ system across their dispersed workforce [81]. While the financial investment required for digital pathology is significant – including file storage costs, laboratory information system integration, and slide scanner purchases – these costs are offset by increased laboratory efficiencies, such as reduced archival personnel and fewer required ancillary diagnostic tests due to ease of archival material retrieval [82]. In European studies, these benefits materialise into revenue generation by Year 3 after digital pathology implementation [81].

Though few studies have directly examined the additional impact of AI tools on this process, once the up-front costs of digital infrastructure are absorbed, incorporating menus of AI applications in image management software will presumably only further increase pathologist efficiency during digital slide review – a process that is slower than glass slide review in most studies [83]. Although there are no currently available reimbursement codes for digital pathology or AI slide interpretation in the United States, the former are under study, and the latter will likely be modelled on the codes currently available for quantitative immunohistochemistry assessment. Ultimately, regulatory hurdles may prove the most significant barriers to introduction of AI pathology tools, as the clearance process for algorithms beyond simple identification of cancer – such as grading, prognosis, or molecular subtype prediction algorithms – remains largely untested to date. Key to this process is a demonstration of model reliability and generalisability, as digitised slide images may be impacted by a multitude of artefacts (including out of focus images, coverslip bubbles or dirt, ink markings from the pathologist, thick or folded tissue sections, heterogeneity in H&E staining protocols, or variable image compression algorithms). To make models robust to these variables, heterogeneous training data, as well as image augmentation and generation techniques are used. To date, relatively few studies have directly examined the impact of these artefacts on algorithm performance in prostate cancer. A recent study found an increased false negative rate in prostate cancer identification may result from many of these artefacts [84]; however, additional computational tools may be developed to address them, improving model generalisability [85].

Conclusions and Future Directions

Digital pathology interpretations augmented by AI models are likely to continue to revolutionise the way we practice medicine in the next decade. AI algorithms based on digital pathology reads of H&E slides have been shown to be great tools at identifying prostate cancers, grading them, and providing both prognostic and predictive information. These algorithms have many benefits: they are inexpensive to perform, non-tissue destructive, and can be obtained rapidly to reduce the pathologist's workload. Such algorithms could

also contribute to our understanding of mechanisms of tumorigenesis, such as the importance of the tumour microenvironment or stromal compartment in predicting outcomes and response to treatment. Because they have the potential of being rapid to deploy, inexpensive, and more accessible than other tissue-based biomarkers, these AI models may also increase access to high-quality and more equitable care internationally.

Despite this significant progress, much work remains before the widespread use of AI for routine diagnosis, prognosis, and prediction in prostate cancer. First and foremost, more transparency or model explainability is needed to elucidate which components of the AI algorithms are driving published results. Are these AI algorithms simply better at quantifying Gleason pattern 4 or cribriform histology? Or are there signals coming from the tumour stroma? A better understanding of what these algorithms are reading out will not only increase their uptake in the clinic (i.e., make them feel less like a 'black box'), but it will also help to guide future research and novel targets for therapy. Second, more research is needed to understand the true clinical benefit of the AI component of the multimodal models over readily available clinical variables, e.g., using modern day cohorts that include MRI data for localised prostate cancer, or PSA doubling time for advanced prostate cancer. Third, future prognostic and predictive model development should be unsupervised and based on all available tissue, not just the representative slide as defined by the human pathologist, or a tissue microarray spot providing only a tiny fraction of the entire specimen, as important signal might be coming from the tissue around the tumour and one of the key goals of these AI models is to identify patterns not seen by the human eye. Predictive algorithms will require extensive testing in prospectively developed, biomarker-integrated clinical trials, rather than relying on *post hoc* analyses of trials designed for a different purpose. Finally, important issues remain to be solved – including the extensive infrastructure and high up-front cost of digital pathology, as well as privacy and legal concerns, and applicability in diverse clinical settings and patient populations – before we will fully realise the potential of histopathology-based AI systems in urological pathology.

Disclosure of Interests

Tamara L. Lotan has received research support from AIRA Matrix, Exact Sciences and Myriad Genetics for other studies and is a consultant for AstraZeneca. Claire M. de la Calle declares grants: National Cancer Institute (NCI) Early Surgeon Scientist Program (ESSP - P30CA015704). The Prostate Cancer Foundation (PCF) Young Investigator Award (24YOUN15). Alexander S. Baras has no disclosures.

References

1 **Schmeusser B**, **Levin B**, **Lama D**, **Sidana A**. Hundred years of transperineal prostate biopsy. *Ther Adv Urol* 2022; 14: 17562872221100590

2 **Walsh E**, **Orsi NM**. The current troubled state of the global pathology workforce: a concise review. *Diagn Pathol* 2024; 19: 163

3 **Litjens G**, **Sánchez CI**, **Timofeeva N et al.** Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Sci Rep* 2016; 6: 26286

4 **Satturwar S**, **Parwani AV**. Artificial intelligence-enabled prostate cancer diagnosis and prognosis: current state and future implications. *Adv Anat Pathol* 2024; 31: 136–44

5 **Pantanowitz L**, **Quiroga-Garza GM**, **Bien L et al.** An artificial intelligence algorithm for prostate cancer diagnosis in whole slide images of core needle biopsies: a blinded clinical validation and deployment study. *Lancet Digit Health* 2020; 2: e407–16

6 **Strom 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

7 **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

8 **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

9 **Raciti P**, **Sue J**, **Retamero JA et al.** Clinical validation of artificial intelligence-augmented pathology diagnosis demonstrates significant gains in diagnostic accuracy in prostate cancer detection. *Arch Pathol Lab Med* 2023; 147: 1178–85

10 **Allsbrook WC**, **Mangold KA**, **Johnson MH**, **Lane RB**, **Lane CG**, **Epstein JI**. Interobserver reproducibility of Gleason grading of prostatic carcinoma: general pathologist. *Hum Pathol* 2001; 32: 81–8

11 **Allsbrook WC**, **Mangold KA**, **Johnson MH et al.** Interobserver reproducibility of Gleason grading of prostatic carcinoma: urologic pathologists. *Hum Pathol* 2001; 32: 74–80

12 **Ozkan TA**, **Eruyar AT**, **Cebeci OO**, **Memik O**, **Ozcan L**, **Kuskonmaz I**. Interobserver variability in Gleason histological grading of prostate cancer. *Scand J Urol* 2016; 50: 420–4

13 **Monaco JP**, **Tomaszewski JE**, **Feldman MD et al.** High-throughput detection of prostate cancer in histological sections using probabilistic pairwise Markov models. *Med Image Anal* 2010; 14: 617–29

14 **Doyle S**, **Feldman M**, **Tomaszewski J**, **Madabhushi A**. A boosted Bayesian multiresolution classifier for prostate cancer detection from digitized needle biopsies. *IEEE Trans Biomed Eng* 2012; 59: 1205–18

15 **Gorelick L**, **Veksler O**, **Gaed M et al.** Prostate histopathology: learning tissue component histograms for cancer detection and classification. *IEEE Trans Med Imaging* 2013; 32: 1804–18

16 **Nguyen K**, **Sarkar A**, **Jain AK**. Prostate cancer grading: use of graph cut and spatial arrangement of nuclei. *IEEE Trans Med Imaging* 2014; 33: 2254–70

17 **Marginean F**, **Arvidsson I**, **Simoulis A et al.** An artificial intelligence-based support tool for automation and standardisation of Gleason grading in prostate biopsies. *Eur Urol Focus* 2021; 7: 995–1001

18 **Chen PC**, **Mermel CH**, **Liu Y**. Evaluation of artificial intelligence on a reference standard based on subjective interpretation. *Lancet Digit Health* 2021; 3: e693–5

19 **Nir G**, **Karimi D**, **Goldenberg SL et al.** Comparison of artificial intelligence techniques to evaluate performance of a classifier for automatic grading of prostate cancer from digitized histopathologic images. *JAMA Netw Open* 2019; 2: e190442

20 **Bulten W, Kartasalo K, Chen PC et al.** Artificial intelligence for diagnosis and Gleason grading of prostate cancer: the PANDA challenge. *Nat Med* 2022; 28: 154–63

21 **Faryna K, Tessier L, Retamero J et al.** Evaluation of artificial intelligence-based Gleason grading algorithms "in the wild". *Mod Pathol* 2024; 37: 100563

22 **Huang W, Randhawa R, Jain P et al.** Development and validation of an artificial intelligence-powered platform for prostate cancer grading and quantification. *JAMA Netw Open* 2021; 4: e2132554

23 **Ryu HS, Jin MS, Park JH et al.** Automated Gleason scoring and tumor quantification in prostate core needle biopsy images using deep neural networks and its comparison with pathologist-based assessment. *Cancers (Basel)* 2019; 11: 1860

24 **Schmidt B, Soerensen SJC, Bhambhvani HP et al.** External validation of an artificial intelligence model for Gleason grading of prostate cancer on prostatectomy specimens. *BJU Int* 2025; 135: 133–9

25 **Singhal N, Soni S, Bonthu S et al.** A deep learning system for prostate cancer diagnosis and grading in whole slide images of core needle biopsies. *Sci Rep* 2022; 12: 3383

26 **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: e2023267

27 **Nagpal K, Foote D, Tan F et al.** Development and validation of a deep learning algorithm for Gleason grading of prostate cancer from biopsy specimens. *JAMA Oncol* 2020; 6: 1372–80

28 **Scuderi S, Tin AL, Klug J et al.** The absolute volume of Gleason pattern 4 on radical prostatectomy is more strongly associated with advanced stage and biochemical recurrence than Gleason grade groups. *J Urol* 2025; 213: 722–9

29 **Silva-Rodríguez J, Colomer A, Sales MA, Molina R, Naranjo V.** Going deeper through the Gleason scoring scale: an automatic end-to-end system for histology prostate grading and cribriform pattern detection. *Comput Methods Programs Biomed* 2020; 195: 105637

30 **Ambrosini P, Hollemans E, Kweldam CF, Leenders GJLH, Stallinga S, Vos F.** Automated detection of cribriform growth patterns in prostate histology images. *Sci Rep* 2020; 10: 14904

31 **Salles DC, Vidotto T, Faisal FA et al.** Assessment of MYC/PTEN status by gene-protein assay in grade group 2 prostate biopsies. *J Mol Diagn* 2021; 23: 1030–41

32 **Chan E, McKenney JK, Hawley S et al.** Analysis of separate training and validation radical prostatectomy cohorts identifies 0.25 mm diameter as an optimal definition for "large" cribriform prostatic adenocarcinoma. *Mod Pathol* 2022; 35: 1092–100

33 **Oliveira LD, Lu J, Erak E et al.** Comparison of pathologist and artificial intelligence-based grading for prediction of metastatic outcomes after radical prostatectomy. *Eur Urol Oncol* 2025; 8: 9–13

34 **Sandeman K, Blom S, Koponen V et al.** AI model for prostate biopsies predicts cancer survival. *Diagnostics (Basel)* 2022; 12: 1031

35 **Wulczyn E, Nagpal K, Symonds M et al.** Predicting prostate cancer specific-mortality with artificial intelligence-based Gleason grading. *Commun Med* 2021; 1: 10

36 **Chen Y, Loveless IM, Nakai T et al.** Convolutional neural network quantification of Gleason pattern 4 and association with biochemical recurrence in intermediate-grade prostate tumors. *Mod Pathol* 2023; 36: 100157

37 **Ding CC, Su ZT, Erak E et al.** Predicting prostate cancer grade reclassification on active surveillance using a deep learning-based grading algorithm. *J Natl Cancer Inst* 2024; 116: 1683–6

38 **Nguyen JK, Harik LR, Klein EA et al.** Proposal for an optimised definition of adverse pathology (unfavourable histology) that predicts metastatic risk in prostatic adenocarcinoma independent of grade group and pathological stage. *Histopathology* 2024; 85: 598–613

39 **Leo P, Chandramouli S, Farré X et al.** Computationally derived cribriform area index from prostate cancer hematoxylin and eosin images is associated with biochemical recurrence following radical prostatectomy and is Most prognostic in Gleason grade group 2. *Eur Urol Focus* 2021; 7: 722–32

40 **Falagario UG, Abbadi A, Remmers S et al.** Biochemical recurrence and risk of mortality following radiotherapy or radical prostatectomy. *JAMA Netw Open* 2023; 6: e2332900

41 **Pound CR, Partin AW, Eisenberger MA, Chan DW, Pearson JD, Walsh PC.** Natural history of progression after PSA elevation following radical prostatectomy. *JAMA* 1999; 281: 1591–7

42 **Freedland SJ, Gao W, Lax A et al.** Real-world outcomes following biochemical recurrence after definitive therapy with a short prostate-specific antigen doubling time: potential role of early secondary treatment. *Prostate Cancer Prostatic Dis* 2024; 28: 809–16

43 **Pinckaers H, van Ipenburg J, Melamed J et al.** Predicting biochemical recurrence of prostate cancer with artificial intelligence. *Commun Med* 2022; 2: 64

44 **Cao L, He R, Zhang A et al.** Development of a deep learning system for predicting biochemical recurrence in prostate cancer. *BMC Cancer* 2025; 25: 232

45 **Eminaga O, Saad F, Tian Z et al.** Artificial intelligence unravels interpretable malignancy grades of prostate cancer on histology images. *NPJ Imaging* 2024; 2: 6

46 **Hu C, Qiao X, Huang R, Bao J, Wang X.** Development and validation of a multimodality model based on whole-slide imaging and Biparametric MRI for predicting postoperative biochemical recurrence in prostate cancer. *Radiol Imaging Cancer* 2024; 6: e230143

47 **Zhang YF, Zhou C, Guo S et al.** Deep learning algorithm-based multimodal MRI radiomics and pathomics data improve prediction of bone metastases in primary prostate cancer. *J Cancer Res Clin Oncol* 2024; 150: 78

48 **Ren J, Karagoz K, Gatza ML et al.** Recurrence analysis on prostate cancer patients with Gleason score 7 using integrated histopathology whole-slide images and genomic data through deep neural networks. *J Med Imaging* 2018; 5: 047501

49 **Fernandez-Mateos J, Cresswell GD, Trahearn N et al.** Tumor evolution metrics predict recurrence beyond 10 years in locally advanced prostate cancer. *Nat Cancer* 2024; 5: 1334–51

50 **Esteva A, Feng J, van der Wal D et al.** Prostate cancer therapy personalization via multi-modal deep learning on randomized phase III clinical trials. *NPJ Digit Med* 2022; 5: 71

51 **Esteva A, Feng J, van der Wal D et al.** Author correction: prostate cancer therapy personalization via multi-modal deep learning on randomized phase III clinical trials. *NPJ Digit Med* 2023; 6: 27

52 **Ross AE, Zhang J, Huang HC et al.** External validation of a digital pathology-based multimodal artificial intelligence architecture in the NRG/RTOG 9902 phase 3 trial. *Eur Urol Oncol* 2024; 7: 1024–33

53 **Spratt DE, Liu VYT, Jia AY et al.** Meta-analysis of individual patient-level data for a multimodal artificial intelligence biomarker in high-risk prostate cancer: results from six NRG/RTOG phase 3 randomized trials. *Eur Urol* 2024; 86: 369–71

54 **Li EV, Ren Y, Griffin J et al.** An artificial intelligence-digital pathology algorithm predicts survival after radical prostatectomy from the prostate, lung, colorectal, and ovarian cancer trial. *J Urol* 2025; 213: 600–8

55 **Bjartell A, Krzyzanowska A, Liu VYT et al.** Validation of a digital pathology-based multimodal artificial intelligence biomarker in a prospective, real-world prostate cancer cohort treated with prostatectomy. *Clin Cancer Res* 2025; 31: 1546–53

56 **NCCN.** NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines®) Prostate Cancer. Version 2.2025. https://www.nccn.org/professionals/physician_gls/pdf/prostate.pdf

57 **Wang JH, Deek MP, Mendes AA et al.** Validation of an artificial intelligence-based prognostic biomarker in patients with oligometastatic castration-sensitive prostate cancer. *Radiother Oncol* 2025; 202: 110618

58 **Feng FY, Smith MR, Saad F et al.** Digital pathology-based multimodal artificial intelligence scores and outcomes in a randomized phase III trial in men with nonmetastatic castration-resistant prostate cancer. *JCO Precis Oncol* 2025; 9: e2400653

59 **Markowski MC, Ren Y, Tierney M et al.** Digital pathology-based artificial intelligence biomarker validation in metastatic prostate cancer. *Eur Urol Oncol* 2025; 8: 755–62

60 **Parker CTA, Mendes L, Liu VYT et al.** External validation of a digital pathology-based multimodal artificial intelligence-derived prognostic model in patients with advanced prostate cancer starting long-term androgen deprivation therapy: a post-hoc ancillary biomarker study of four phase 3 randomised controlled trials of the STAMPEDE platform protocol. *Lancet Digit Health* 2025; 7: 100885

61 **Spratt DE, Tang S, Sun Y et al.** Artificial intelligence predictive model for hormone therapy use in prostate cancer. *NEJM Evidence* 2023; 2: EVIDoa2300023

62 **Wegener E, Ng M, Guerrieri M et al.** A multicentre implementation trial of an artificial intelligence-driven biomarker to inform shared decisions for androgen deprivation therapy in men undergoing prostate radiotherapy: the ASTuTE protocol. *BMC Cancer* 2025; 25: 250

63 **Armstrong AJ, Liu VYT, Selvaraju RR et al.** Development and validation of an artificial intelligence digital pathology biomarker to predict benefit of long-term hormonal therapy and radiotherapy in men with high-risk prostate cancer across multiple phase III trials. *J Clin Oncol* 2025; 43: 3494–504

64 **Parker C.** Multimodal artificial intelligence (MMAI) model to identify benefit from second generation androgen receptor pathway inhibitors (ARPI) in high-risk non-metastatic prostate cancer patients from STAMPEDE. 2025 https://meetings.asco.org/abstracts-presentations/253107

65 **Nguyen PL, Alibhai SM, Basaria S et al.** Adverse effects of androgen deprivation therapy and strategies to mitigate them. *Eur Urol* 2015; 67: 825–36

66 **Nakata W, Mori H, Tsujimura G, Tsujimoto Y, Gotoh T, Tsujihata M.** Pilot study of an artificial intelligence-based deep learning algorithm to predict time to castration-resistant prostate cancer for metastatic hormone-naïve prostate cancer. *Jpn J Clin Oncol* 2022; 52: 1062–6

67 **Leo P, Janowczyk A, Elliott R et al.** Computer extracted gland features from H&E predicts prostate cancer recurrence comparably to a genomic companion diagnostic test: a large multi-site study. *NPJ Precis Oncol.* 2021; 5: 35

68 **Fay M, Liao RS, Lone ZM et al.** Artificial intelligence-based digital histologic classifier for prostate cancer risk stratification: independent blinded validation in patients treated with radical prostatectomy. *JCO Clin Cancer Inform* 2025; 9: e2400292

69 **Kather JN, Pearson AT, Halama N et al.** Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat Med* 2019; 25: 1054–6

70 **Liu H, Zhang Y, Luo J.** Contrastive learning-based histopathological features infer molecular subtypes and clinical outcomes of breast cancer from unannotated whole slide images. *Comput Biol Med* 2024; 170: 107997

71 **Aguiar JA, Li EV, Siddiqui MR et al.** Utilization of genetic testing in men with advanced prostate cancer. *Prostate* 2023; 83: 516–23

72 **Kurian AW, Abrahamse P, Furgal A et al.** Germline genetic testing after cancer diagnosis. *JAMA* 2023; 330: 43–51

73 **Omar M, Xu Z, Rand SB et al.** Semi-supervised, attention-based deep learning for predicting TMPRSS2:ERG fusion status in prostate cancer using whole slide images. *Mol Cancer Res* 2024; 22: 347–59

74 **Erak E, Oliveira LD, Mendes AA et al.** Predicting prostate cancer molecular subtype with deep learning on histopathologic images. *Mod Pathol* 2023; 36: 100247

75 **Dadhania V, Gonzalez D, Yousif M et al.** Leveraging artificial intelligence to predict ERG gene fusion status in prostate cancer. *BMC Cancer* 2022; 22: 494

76 **Pizurica M, Larmuseau M, Van der Eecken K et al.** Whole slide imaging-based prediction of TP53 mutations identifies an aggressive disease phenotype in prostate cancer. *Cancer Res* 2023; 83: 2970–84

77 **Hu Q, Rizvi AA, Schau G et al.** Development and validation of a deep learning-based microsatellite instability predictor from prostate cancer whole-slide images. *NPJ Precis Oncol* 2024; 8: 88

78 **Weitz P, Wang Y, Kartasalo K et al.** Transcriptome-wide prediction of prostate cancer gene expression from histopathology images using co-expression-based convolutional neural networks. *Bioinformatics* 2022; 38: 3462–9

79 **Zhang J, Ding F, Guo Y et al.** AI-based prediction of androgen receptor expression and its prognostic significance in prostate cancer. *Sci Rep* 2025; 15: 3985

80 **Li Y, Xiong X, Liu X et al.** Predicting. *Ann Med* 2024; 56: 2399759

81 **Matias-Guiu X, Temprana-Salvador J, Garcia Lopez P et al.** Implementing digital pathology: qualitative and financial insights from eight leading European laboratories. *Virchows Arch* 2025; 487: 815–26

82 **Hanna MG, Reuter VE, Samboy J et al.** Implementation of digital pathology offers clinical and operational increase in efficiency and cost savings. *Arch Pathol Lab Med* 2019; 143: 1545–55

83 **Hanna MG, Reuter VE, Hameed MR et al.** Whole slide imaging equivalency and efficiency study: experience at a large academic center. *Mod Pathol* 2019; 32: 916–28

84 **Schomig-Markiefka B, Pryalukhin A, Hulla W et al.** Quality control stress test for deep learning-based diagnostic model in digital pathology. *Mod Pathol* 2021; 34: 2098–108

85 **Weng Z, Seper A, Pryalukhin A et al.** GrandQC: a comprehensive solution to quality control problem in digital pathology. *Nat Commun* 2024; 15: 10685

Correspondence: Tamara L. Lotan, Johns Hopkins University School of Medicine, 1550 Orleans Street, Baltimore, MD 21231, USA. e-mail: tlotan1@jhmi.edu

Abbreviations: ADT, androgen deprivation therapy; AI, artificial intelligence; AUC, area under the curve; BCR, biochemical recurrence; CAPRA, Cancer of the Prostate Risk Assessment; CRPC, castrate-resistant prostate cancer; GG, Grade Group; H&E, haematoxylin and eosin; MMAI, multimodal AI; MMRd, mismatch repair deficiency; NCCN, National Comprehensive Cancer Network; PANDA, Prostate cANcer graDe Assessment; RTOG, Radiation Therapy Oncology Group.

