

Review Article

Whole-mount histopathology as spatial ground truth for artificial intelligence in prostate cancer: a structured narrative review of techniques, models, and translational gaps

Yodit Aschenaki¹, Hamdi Khasawneh⁵, Trey Gustafson³, Raymond Hu⁴, Hanna Ketema¹, Mason Fraiman², Rasoul Sali¹, Christopher Dixon¹, David Y Zhang¹

¹Department of Computational Pathology, NovinoAI, 1443 NE 4th Ave, Fort Lauderdale, FL 33304, USA; ²Department of Biology, Siena University, 515 Loudon Rd, Loudonville, NY 12211, USA; ³University of Michigan Medical School, Ann Arbor, MI 48109-0340, USA; ⁴Brown University, Providence, RI 02903, USA; ⁵Department of Data Science and Artificial Intelligence, King Hussein School of Computing Sciences, Princess Sumaya University for Technology, Amman 11855, Jordan

Received June 8, 2026; Accepted July 8, 2026; Epub August 15, 2026; Published August 30, 2026

Abstract: Background. Prostate cancer is spatially heterogeneous, multifocal, and architecturally varied. Conventional biopsy and step-sectioned histopathology fragment the gland. This limits accurate assessment of dominant-lesion identity, tumor volume, multifocality, extraprostatic extension, seminal vesicle invasion, and surgical margin status. Whole-mount histopathology (WMH) preserves the prostate as a continuous cross-sectional unit and is increasingly used as the reference standard for radiology-pathology correlation and for validating artificial-intelligence (AI) models. Objective. This structured narrative review synthesizes evidence on the joint use of WMH and AI in prostate cancer pathology. For each clinical question, we examine what spatial information WMH preserves, what AI methods extract from that information, where the joint approach has been validated, and where validation gaps remain. The questions are index-lesion identification, tumor-volume estimation, multifocality, extraprostatic extension, seminal vesicle invasion, surgical margin assessment, Gleason grading, magnetic resonance imaging (MRI)-pathology registration, and surgical quality assurance. Findings. WMH improves index-lesion identification, volumetric reproducibility, and margin clarity, and provides a high-fidelity reference standard for training prostate-AI models. Registered to the patient's presurgical MRI, these whole-mount labels have also been used to train preoperative tools and have been studied in research cohorts for focal-therapy margin definition and extraprostatic-extension prediction, extending the value of WMH beyond retrospective staging. AI systems built on WMH show improved performance for tumor detection and Gleason classification. Most reported results, however, lack external validation, reporting against established medical-AI standards, and demonstrated robustness across scanners, stains, and patient populations. WMH is itself an imperfect ground truth, subject to orientation, registration, and interobserver variability that AI models inherit. Implications. WMH-grounded AI is a credible translational route, but routine adoption requires standardized large-format scanning, shared annotation protocols, external validation, and regulatory clearance for autonomous use. We identify field priorities, including cross-institutional WMH datasets, domain-shift mitigation, bias auditing, and reimbursement parity with AI-augmented radiology.

Keywords: Whole-mount histopathology, prostate cancer, artificial intelligence, deep learning, ground truth, radical prostatectomy

Introduction

Prostate cancer is one of the most prevalent malignancies among men worldwide [1]. Its clinical heterogeneity creates persistent challenges in diagnosis, staging, and treatment. Despite advances in screening and therapy, accurate characterization remains difficult be-

cause of the disease's complex biology and variable progression [2].

Histopathologic evaluation remains the diagnostic reference for prostate cancer staging and grading. Conventional approaches, however, rely on biopsy sampling and step-sectioned histology, which fragment the gland into smaller

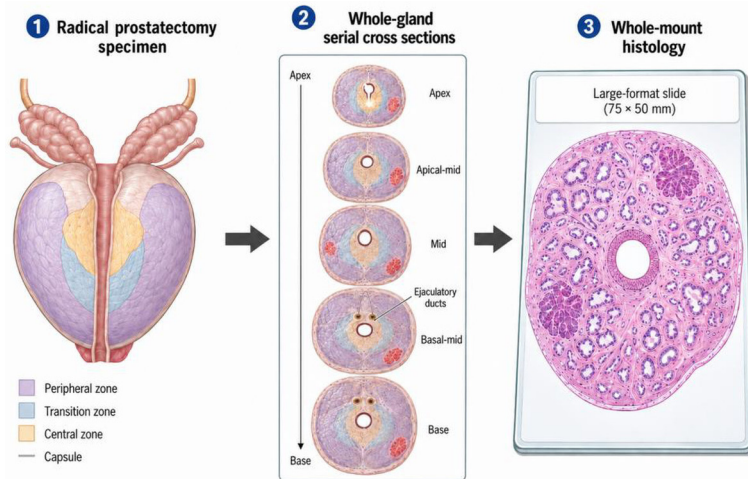


Figure 1. From radical prostatectomy specimen to whole-mount slide. The intact radical prostatectomy specimen (left) is serially cross-sectioned from apex to base perpendicular to the urethra, preserving the zonal anatomy (peripheral, transition, and central zones) and the capsule. Each whole-gland cross-section is mounted on a single large-format slide (75 × 50 mm), in contrast to the multiple fragmented blocks produced by conventional step-sectioning. The result is a continuous spatial map in which tumor foci, the capsule, ejaculatory ducts, and seminal vesicles remain in their true anatomical relationships, the property that makes whole-mount histopathology usable as a spatial reference standard for downstream artificial-intelligence analysis.

blocks. This sampling strategy can disrupt the spatial relationships between tumor foci, the prostatic capsule, surgical margins, and surrounding structures, relationships that influence grading and pathologic staging [3].

As management shifts toward precision oncology, a spatially complete understanding of tumor morphology and distribution has become increasingly important for diagnostic accuracy and therapeutic decision-making. Whole-mount histopathology (WMH) addresses several limitations of fragmented sampling by preserving large-format sections of the prostate gland (**Figure 1**). Whereas conventional histopathology mounts tissue on standard 75 × 25 mm glass slides, WMH uses larger 75 × 50 mm slides, so that an entire cross-section of the prostatectomy specimen can be examined on a single slide. The principal advantage of WMH is a more spatially coherent assessment of tumor distribution, tumor volume, surgical margins, and relationships to adjacent anatomical structures [4-6].

Advances in artificial intelligence (AI), particularly deep learning (DL)-based image analysis,

have transformed digital pathology [7]. Pathology AI models can analyze high-resolution histologic images to detect tumor regions, perform automated Gleason grading, and quantify morphologic features [8]. When applied to digitized whole-mount slides, these algorithms enable large-scale, reproducible analysis of the complex spatial patterns within prostate cancer [9].

This review examines the combined role of WMH and AI in prostate cancer evaluation. It details how WMH provides a spatially complete representation of tumor architecture, and how AI methods can use that information to improve tumor characterization, staging, surgical quality assessment, and education.

Scope and contribution

Prior reviews have addressed WMH as a surgical-pathology technique [5, 6] and have separately surveyed AI in prostate cancer at the methods level [10-12]. Benchmark studies such as the PANDA challenge have cataloged AI performance on biopsy-derived material [13], and radiology-pathology correlation work has begun to use whole-mount sections as alignment ground truth [9, 14]. This review complements that literature with a distinct lens: WMH is treated not only as a clinical artifact but as the spatial ground truth that determines what AI can and cannot validly measure on prostatectomy specimens. Where prior reviews emphasize either the technique or the models, this review systematically pairs the two. For each clinical question (index lesion, tumor volume, multifocality, extraprostatic extension (EPE), seminal vesicle invasion (SVI), margins, Gleason grade, and magnetic resonance imaging (MRI)-pathology registration), we examine what spatial information WMH preserves, what AI can extract from it, where the joint approach has been validated, and where validation is still missing. We also identify the technical, regulatory, and operational bottlenecks that constrain real-world deployment in community

Whole-mount histology as spatial ground truth

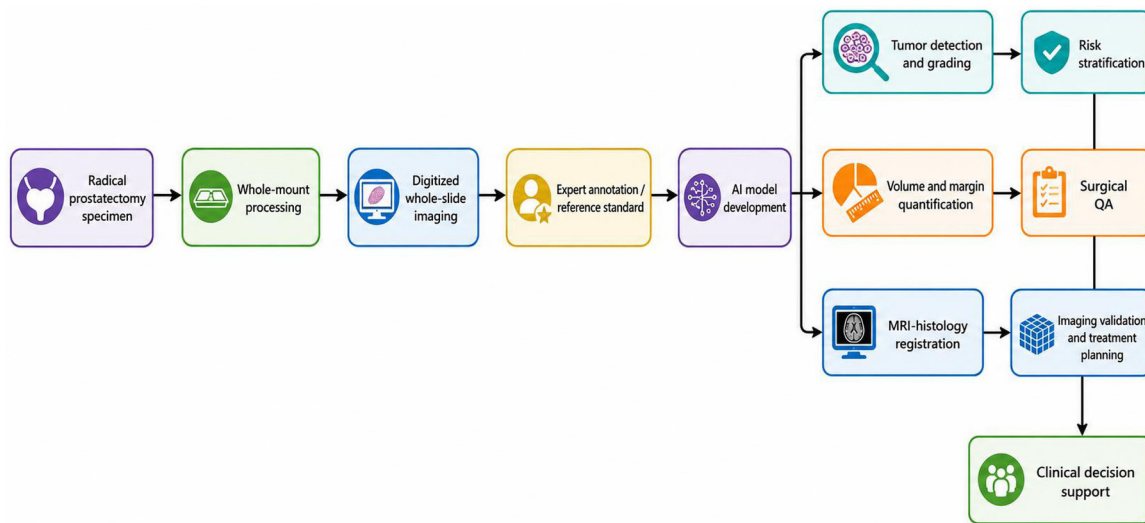


Figure 2. The whole-mount-histopathology-grounded artificial-intelligence pipeline. A radical prostatectomy specimen undergoes whole-mount processing and digitized whole-slide imaging. Expert annotation on the spatially complete whole-mount image creates the reference standard used for artificial-intelligence model development. Trained models feed three coupled output streams, namely tumor detection and grading, volume and margin quantification, and magnetic-resonance-imaging-histology registration, which in turn support risk stratification, surgical quality assurance, imaging validation, and ultimately multidisciplinary clinical decision support. The diagram emphasizes that every downstream artificial-intelligence output is only as valid as the spatial ground truth established at the annotation step.

pathology, the constituency most likely to benefit if WMH-grounded AI becomes routine.

Methods

This is a structured narrative review conducted in accordance with SANRA guidance for transparent narrative reviews [15]; it is narrative rather than systematic, and no protocol was registered. Literature was identified through PubMed and supplementary citation screening between January 2010 and May 2026, using combinations of the terms *whole-mount histopathology*, *whole-mount prostatectomy*, *digital pathology*, *artificial intelligence*, *deep learning*, *Gleason grading*, *extraprostatic extension*, *seminal vesicle invasion*, *surgical margin*, *MRI-pathology correlation*, and *foundation model*. Priority was given to (A) peer-reviewed articles from the last seven years, (B) higher-impact pathology, urology, and medical-AI venues, and (C) reporting-standards and regulatory documents relevant to medical AI [16-19]. Reference lists of included articles were screened iteratively. Reporting of evidence syntheses was informed by PRISMA principles where applicable [20].

Tumor characterization and AI applications

Tumor characterization in the radical-prostatectomy (RP) specimen underpins risk stratification, treatment planning, and prognostic assessment. Accurate identification of the dominant lesion, estimation of tumor burden, and characterization of multifocal distribution influence decisions on active surveillance, focal therapy, nerve-sparing surgery, and postoperative management. Conventional approaches, such as biopsy, MRI interpretation, and standard step-sectioned histology, are inherently limited by partial sampling and fragmented tissue architecture, which may obscure the true extent and spatial organization of disease (Figure 2) [21].

Compared with fragmented sectioning, WMH gives clearer visualization of tumor location, extent, and relationships to adjacent structures, and it facilitates correlation with biopsy findings and imaging abnormalities. For this reason, WMH is increasingly used as a reference standard in studies that link pathology with MRI and computational modeling [5, 14]. WMH should not, however, be treated as an infallible ground truth. Its reliability depends

on specimen orientation, tissue processing, annotation protocol, and the quality of pathology-imaging registration, and interobserver variability remains relevant for complex characterization tasks. This underscores the need for standardized annotation when WMH is used to train or validate AI models [22].

When combined with AI, WMH provides a foundation for more reproducible and quantitative tumor characterization. Models trained or validated on WMH datasets can support tumor detection, segmentation, grading, and spatial mapping across the gland [4, 9]. The strength of this framework lies not in automation alone but in the ability of WMH to provide spatially coherent labels against which AI outputs can be compared. WMH and AI together therefore offer a route to improving the accuracy, consistency, and interpretability of prostate cancer characterization, while still requiring careful attention to validation and generalizability.

Index-lesion identification and spatial mapping

Identification of the index lesion is central to prostate cancer characterization because it informs prognostication, focal-therapy planning, radiologic targeting, and pathologic correlation. In clinical practice, the index lesion is typically inferred from biopsy cores, MRI findings, and histologic review. These approaches are useful but susceptible to sampling error and observer-dependent interpretation, particularly with multifocal disease or heterogeneous intraprostatic tumor distribution [21].

Whole-mount RP specimens allow the dominant lesion to be defined with greater confidence through integrated assessment of tumor size, grade, laterality, and spatial orientation across the entire gland. This is especially important because the index lesion often determines the dominant pathologic burden even when multiple foci are present. In a whole-mount analysis of 100 RP specimens, tumors were multifocal in 78% of cases and bilateral in 86%, yet tumor volume, Gleason score, and pathologic stage were almost invariably defined by the index lesion, while most satellite foci were small and low grade [23].

DL models, especially convolutional neural networks (CNNs), can detect and classify tumor regions on whole-mount images and learn

complex features of tumor morphology and spatial distribution. DL pipelines have also been used to align histopathology images with MRI, integrating multimodal data for tumor mapping and localization [4, 14]. In external-validation work, an AI Gleason-grading model was evaluated against expert genitourinary pathologists on digitized whole-mount RP specimens and supported the potential utility of automated grading and lesion characterization, although performance should be read within the external-validation context rather than as evidence of solved localization [38]. An explainable CNN for prostate tumor segmentation on multiparametric MRI (mpMRI), validated against whole-mount histology, reported mean Dice similarity coefficients (DSC) of 0.31 to 0.32 relative to radiologist- and histology-based ground truth [30]. This is well below the threshold (commonly taken as a DSC of approximately 0.70) conventionally regarded as indicating clinically useful tumor segmentation. That result illustrates the difficulty of cross-modal lesion localization on prostate mpMRI and motivates further work; it does not demonstrate accurate tumor localization. The authors' use of gradient-weighted class activation mapping (Grad-CAM) overlays nonetheless provides interpretability that is valuable for training and trust-building, independent of the segmentation metric [30].

From a clinical perspective, more precise index-lesion identification improves alignment between imaging targets, pathology findings, and treatment strategies.

Tumor-volume estimation and quantitative burden

Tumor volume is an established predictor of adverse pathologic features and biochemical recurrence (BCR) after RP and remains an important component of postoperative risk assessment [25]. Traditional tumor-volume estimation has relied on manual measurement that is labor-intensive and subject to interobserver variability. Differences in contouring methods, reconstruction assumptions, and institutional practice contribute to inconsistency in how tumor burden is measured and reported [31, 39].

WMH improves volumetric assessment by allowing more accurate mapping of tumor

boundaries than fragmented step-sectioned histology. This anatomical integrity reduces sampling distortion and improves reproducibility. AI further enhances this capability through automated segmentation and quantitative analysis. An AI cancer-mapping platform (Unfold AI; Avenda Health) was evaluated in 204 men for AI-based estimation of prostate tumor volume and prediction of focal-therapy outcomes, illustrating the use of whole-mount-anchored quantitative tools for treatment planning [26]. This matters preoperatively because imaging systematically understates disease extent. In patient-specific-mold studies that correlate MRI directly with whole-mount pathology, cancer foci averaged roughly 11 mm larger in maximal diameter and about three-fold greater in volume than their T2-weighted MRI segmentation, so volume estimates that ignore this bias risk inadequate focal-therapy margins and undertreatment [40]. Building on whole-mount-validated tumor maps, an AI model that produces three-dimensional cancer-estimation maps was tested retrospectively in 50 RP specimens with whole-mount ground truth. Its focal-treatment margins captured more cancer-bearing tissue than conventional MRI regions of interest (97% vs 37% voxel sensitivity), and its predicted negative-margin probabilities agreed closely with observed rates. The same margins, however, did not significantly outperform fixed 10 mm expansion margins [24]. DL image-analysis models have separately shown agreement with expert-pathologist assessments of tumor volume and percent tumor involvement on whole-mount slides in validation studies [35]. A prior report describing the NovinoAI digital-pathology information system found that pathologist estimates of tumor volume in RP specimens varied by up to 30% relative to AI-assisted quantification [41].

These findings indicate that WMH is useful both for more accurate direct measurement of tumor burden and as a high-fidelity reference standard for training and validating AI models that infer tumor extent from imaging or other incomplete data sources. This value is not limited to measuring tumor volume after RP. As detailed in Section 2.5, the same whole-mount-validated models are being developed to improve preoperative estimation of whole-gland disease burden from biopsy and imaging inputs, the setting in which they could inform

treatment selection. As these workflows mature, AI-based quantification grounded in WMH may support more standardized reporting and strengthen recurrence prediction when integrated with grade, margin status, and stage.

Multifocal disease and spatial heterogeneity

Prostate cancer is frequently multifocal, with multiple discontinuous tumor foci that vary in size, grade, and spatial distribution. Conventional biopsy-based assessment may underestimate this complexity, because limited sampling can miss smaller satellite lesions or fail to capture the relationship between dominant and secondary foci. Traditional workflows may therefore incompletely represent the true burden and heterogeneity of disease [23].

WMH enables visualization of multifocal distribution by revealing discontinuous tumor foci that may otherwise remain undetected. This allows pathologists to determine whether lesions are separate, contiguous, or part of a broader heterogeneous pattern, and to characterize lesion-to-lesion relationships and zonal distribution more reliably than fragmented sampling [5].

AI systems trained on whole-mount datasets can learn spatial patterns associated with multifocal disease and may detect subtle satellite lesions [4]. Explainable methods support interpretability here: Grad-CAM can highlight image regions contributing to model predictions, improving transparency in how tumor foci are identified [30]. Technical systems that improve reproducibility of histopathology orientation and registration may further support consistent digital workflows for spatial analysis of whole-mount data [14]. Clinically, improved characterization of multifocality may influence surgical planning, focal-therapy candidacy, and interpretation of recurrence risk; whole-mount-validated AI tumor maps have been used in research to define focal-therapy margins that captured more of the histologic tumor extent than conventional imaging contours alone [24]. Nonetheless, better visualization of multifocal architecture does not automatically mean that all clinically relevant foci can be detected preoperatively or modelled reliably across institutions [27].

Staging: extraprostatic extension, seminal vesicle invasion, and lymph-node assessment

While tumor characterization focuses on the properties of the tumor itself, staging determines how far the tumor has spread. In prostate cancer, staging uses the Tumor-Node-Metastasis (TNM) system. Pathologic staging following RP provides the most accurate assessment of tumor spread, allowing determination of EPE, SVI, and lymph-node involvement, features critical to prognostication, recurrence-risk estimation, and therapy decisions [3]. Traditional staging relies on a combination of biopsy results, imaging, and step-sectioned histology, but these approaches are limited by partial sampling and fragmented tissue architecture. WMH improves staging accuracy by preserving the spatial continuity of the gland and maintaining the relationship between tumor foci, the capsule, and surrounding structures. Combined with AI, this can support more reproducible and accurate staging [4].

Extraprostatic extension: EPE occurs when prostate cancer penetrates the prostatic capsule and spreads into surrounding tissue. Detecting EPE is important because it influences surgical strategy: when EPE can be confidently excluded, nerve-sparing RP may preserve erectile and urinary function, whereas if EPE cannot be ruled out a wider resection may be required for complete tumor removal. mpMRI, the current primary technique for preoperative EPE evaluation, has limited and variable sensitivity between radiologists, which constrains objective and reproducible preoperative assessment [28]. On the RP specimen, WMH improves this assessment by preserving the continuous interface between the prostatic capsule and surrounding tissue, allowing pathologists to follow the capsule around the entire gland and directly observe where tumor cells extend [4]. In current practice this role is post-surgical: WMH refines pathologic (pT) staging of EPE and the resulting risk stratification, rather than guiding an operation already performed on that patient. AI systems trained on WMH datasets may improve the reproducibility of pathologic EPE classification and reduce interobserver variability. Critically, the inverse task, preoperative prediction of capsular status, is no longer only aspirational, and the bridge is spatial registration. Because RP whole-mounts

can be aligned to the patient's own presurgical MRI, the capsule-referenced labels they provide become the training target for models that act before surgery. EPENet, a DL model developed on mpMRI with EPE ground truth transferred from registered whole-mount histopathology, detected EPE with higher sensitivity than radiologists (80% vs 50%) in an independent test set drawn from a 123-patient registered cohort, albeit at lower specificity [29]. DL models reading diagnostic biopsy hematoxylin and eosin (H&E) slides have separately predicted RP EPE from preoperative tissue alone [42]. A 2025 systematic review and meta-analysis reported a higher pooled area under the curve for AI-enabled imaging than for radiologists in preoperative EPE prediction, although pooled sensitivity and specificity did not differ significantly [43]. These tools are still constrained by modest specificity and largely single-cohort validation, but they illustrate how WMH-derived spatial ground truth can be translated toward preoperative prediction rather than serving only retrospective pT staging.

Seminal vesicle invasion: SVI is among the most significant prognostic factors in prostate cancer; affected patients have higher rates of recurrence, metastatic progression, and cancer-specific mortality. Accurate identification of SVI requires distinguishing true invasion of the seminal vesicle from extension toward the prostate base. This distinction is challenging when limited tissue sections are examined, since small foci of invasion near the ejaculatory ducts may be missed if this region is inadequately sampled [3]. Whole-mount sections improve SVI assessment by preserving the prostate base, ejaculatory ducts, and seminal vesicles in one continuous tissue plane. This makes it easier to determine whether the tumor is truly invading the vesicle wall rather than lying close to it, and it enhances detection of small or discontinuous foci that standard sampling may miss [4]. As discussed in Section 2.1, AI segmentation of prostate tumor on mpMRI validated against whole-mount histology has so far yielded low cross-modal overlap (DSC 0.31 to 0.32). This reinforces that WMH is valuable primarily as the reference standard against which such models are measured, rather than as evidence that cross-modal localization is solved [30].

Neurovascular bundle involvement and nerve-sparing assessment: Neurovascular bundle involvement is closely tied to surgical quality and is discussed in Section 3.2 to avoid duplication. In brief, because WMH retains zonal orientation, it can reveal whether nerve-sparing dissections produced capsular incisions or close margins in high-risk regions [28].

Lymph-node assessment: Lymph-node involvement is associated with systemic progression, and conventional size-based imaging has limited sensitivity, because metastatic prostate cancer can occur in normal-sized nodes. WMH characterizes the prostate specimen only after surgery and therefore cannot itself guide intraoperative decisions in the same patient. The spatial tumor features it captures, however, provide a high-fidelity training target for AI models. When such models are applied to preoperative inputs, such as biopsy pathology, imaging, and prostate-specific antigen (PSA), they may refine estimation of nodal-metastasis risk and help identify, before surgery, the patients most likely to benefit from pelvic lymph-node dissection performed at the time of RP [4].

Together, WMH and AI may help standardize staging, reduce interobserver variability, and improve correlations between pathology, imaging, and clinical outcomes as digital-pathology infrastructure expands.

Preoperative translation: WMH-trained AI for focal therapy and treatment planning

A recurring and legitimate objection to WMH is that, because it is obtained only from the RP specimen, it cannot itself benefit the patient it was taken from: the operation is already complete, and the applications described above are, on their face, post-prostatectomy. Resolving this objection requires distinguishing WMH as a *direct intervention* from WMH as the *spatial ground truth* that makes preoperative AI tools trustworthy. It is the latter role, not retrospective staging, that can support preoperative, patient-facing value, and a growing body of research now illustrates this chain.

The enabling step is registration. DL pipelines such as ProsRegNet and RAPHIA spatially align whole-mount histopathology with the patient's presurgical mpMRI and transfer the histologic cancer labels onto the imaging, producing a

voxel-level “answer key” that would otherwise be unobtainable in living patients [14, 44]. Models trained against this whole-mount-derived key learn to recognize disease on inputs that are available before surgery, namely MRI and biopsy, so the postoperative specimen becomes the teacher for preoperative prediction rather than an end in itself.

This chain has been most fully realized for focal-therapy planning, where it speaks directly to overtreatment and undertreatment. MRI systematically underestimates tumor geometry relative to whole-mount, by roughly 11 mm in maximal diameter and threefold in volume in mold-based correlation studies, so focal-treatment margins drawn from imaging alone may transect undetected tumor, driving recurrence or, when compensated by overly wide ablation, unnecessary damage to continence and potency structures [40]. In a retrospective validation of 50 RP cases with whole-mount ground truth, an AI model generating three-dimensional cancer-estimation maps produced focal-treatment margins that captured more cancer-bearing tissue than conventional MRI regions of interest and gave accurate estimates of negative-margin probability, although it did not significantly outperform fixed 10 mm expansion margins [24]. Because those margins were derived from preoperative imaging and biopsy inputs, the study illustrates how whole-mount labels can train tools intended to inform planning before treatment, even though the evidence so far is retrospective and single-institution. A separate 204-man evaluation of one such platform assessed AI tumor-volume estimation together with prediction of focal-therapy outcomes [26]. These are preoperative decisions, such as who is a focal-therapy candidate and where the ablation boundary should fall, that such models may help inform, with the whole-mount serving to train and validate them.

The same logic extends to candidacy and treatment-selection decisions that hinge on local stage. EPENet, trained on mpMRI with EPE labels registered from whole-mount, predicted EPE preoperatively with higher sensitivity than radiologists, informing the nerve-sparing-versus-wide-excision decision before the operation [29]. DL analysis of diagnostic biopsy H&E has predicted RP EPE from preoperative tissue alone [42], and pooled analysis reported higher

area under the curve for AI-enabled imaging than for conventional reads, though without a significant difference in sensitivity or specificity [43]. In each case the whole-mount is the reference that makes the preoperative prediction credible.

These tools are not yet mature. Reported specificity is often modest, most validation is single-institution and retrospective, and prospective trials demonstrating improved patient outcomes, not just improved concordance with pathology, are still needed. But the trajectory is concrete rather than aspirational. WMH does not treat the patient from whom it was taken; it is the spatial ground truth that may make preoperative focal-therapy planning, lesion targeting, and treatment selection more accurate for the next patient, a benefit that prospective trials still need to confirm.

Surgical evaluation and quality assurance

The spatially complete representation of the prostate gland through WMH enables pathologists and institutions to evaluate tumor volume, positive surgical margins, the oncological adequacy of nerve-sparing surgery (the relationship of tumor and margins to the neurovascular bundle, rather than functional outcomes such as erectile recovery, which are assessed clinically), and Gleason grading in a structured, reproducible manner. Beyond diagnosis, WMH has been used in studies of quality assurance (QA), surgical-performance evaluation, and institutional pathology audit.

Positive surgical margins

Alongside EPE and SVI assessment, evaluation of surgical-margin status is critical. Positive surgical margins (PSM) indicate that cancer cells extend to the inked surface of the prostate, suggesting a risk of residual disease. In a study of minimally invasive RP, PSM were present in 30.6% of cases and were associated with higher tumor volume and more advanced stage, with BCR occurring in 10.7% of margin-positive patients compared with 5.1% of those with negative margins. Although not all patients with PSM recur, the finding indicates disease aggressiveness and the need for closer follow-up and potential adjuvant therapy [45].

Unlike standard step-sectioned histology, which may contain gaps or misalignment, WMH allows continuous localized visualization of tumor proximity to resection margins throughout the entire gland. This format facilitates identification of PSM and more accurate measurement of their length and location [31]. The ability to examine the full perimeter of the gland, including apex, base, posterolateral, and capsular zones, provides a comprehensive view of whether tumor cells intersect inked margins or remain confined. This matters because both PSM length and location are determinants of BCR and clinical significance [46, 47]. WMH also reduces complications of histologic interpretation such as cautery effects, ink smearing, tissue distortion, and tangential cuts introduced as surgical artifacts, helping pathologists more confidently distinguish true margin involvement from artifact and reducing misclassification of margins as positive [25].

Margin assessment nonetheless remains labor-intensive and subject to interobserver variability, influenced by surgical technique, tumor characteristics, patient factors, and specimen handling. Even subtle distinctions such as tumor “touching” versus “almost touching” ink can alter recurrence risk, so standardized, objective criteria are needed [3, 36, 48]. DL models may address some of these limitations through pixel-level margin detection, mapping tumor relative to inked surfaces, and quantifying margin length and multifocality. AI systems may also help distinguish true tumor involvement from surgical artifact and assess tumor-to-margin distance, supporting more precise orientation and risk assessment [9].

Nerve-sparing outcomes

Several reports describe WMH in evaluating the outcomes of nerve-sparing procedures. Because WMH retains zonal orientation, it can reveal whether neurovascular-bundle-preserving dissections resulted in capsular incisions or close margins in high-risk regions, and it can be used to assess whether nerve-sparing efforts were oncologically appropriate based on tumor proximity to the surgical plane [28]. AI-based spatial analysis may further enhance this evaluation by quantifying tumor-to-capsule distances and mapping tumor distribution rela-

tive to critical anatomical structures across digitized whole-mount slides.

Gleason grading quality assurance

Gleason grading remains a cornerstone of prostate cancer prognosis and treatment planning [32]. Accurate assignment of Grade Groups informs surgical decisions, adjuvant-therapy considerations, and active-surveillance eligibility. Even small high-grade foci, such as tertiary Gleason pattern 5 or cribriform glands, and features like intraductal carcinoma of the prostate (IDC-P), can substantially affect BCR, EPE, and cancer-specific survival [49, 50]. Despite its clinical importance, conventional Gleason grading shows considerable interobserver variability. Pathologists may disagree on primary and secondary patterns, on the reporting of tertiary patterns, and on the integration of heterogeneous architectural features, which reduces reproducibility and may affect prognostic accuracy [28, 37, 51].

AI offers a route to mitigate these challenges. DL models trained on digitized whole-mount slides can support Gleason grading by analyzing architectural patterns, nuclear and cytoplasmic features, and spatial context across the entire gland [9, 37]. AI Gleason and tumor-detection tools have shown agreement with expert pathologists and improved grading consistency in validation and benchmark studies, including clinical-grade weakly supervised models and large multi-reader and multi-institutional evaluations [13, 33, 34, 52-55]. While much of this evidence derives from prostate-biopsy material, external-validation work has begun to assess such tools on whole-mount RP slides [38]. By combining whole-gland spatial analysis with algorithmic consistency, these models provide a more objective and reproducible framework for tumor grading, though performance figures should always be read against the cohort, reference standard, and external-validation status of each study.

Quality-assurance applications

AI systems developed using WMH data have also been applied to surgical QA. Models have been developed to detect margins, grade tumors across whole-mount slides, measure tumor-to-margin distances automatically, quantify margin involvement, and distinguish multi-

focal disease; these tools are being evaluated for their potential to support structured review of surgical specimens and provide standardized reports at scale [35, 38]. Beyond individual case review, WMH has been described as a useful resource for institutional and longitudinal QA. By preserving the spatial distribution of tumors and margin involvement across patient cohorts, WMH supports tracking of surgical trends, margin-positivity rates, and grade distribution over time, and WMH-based datasets have been proposed for integration into dashboards and analytic systems that allow benchmarking of surgeon-level outcomes or departmental QA efforts [35, 36]. Synthetic datasets have also been introduced as a training and audit resource, allowing institutions to train AI models and human reviewers using controlled data that represent specific edge cases, complications, or rare morphologies [37].

Educational applications

WMH is a valuable format in teaching materials for pathology, urology, and radiology trainees, because it preserves the architecture of the prostate in its full cross-sectional form. This enables learners to review the continuity of anatomical zones, tumor growth patterns, and proximity to surrounding structures more clearly than in fragmented or partial sections [4]. Features such as cribriform glands, tertiary Gleason patterns, and discontinuous tumor foci may be more readily visible on whole-mount slides owing to preserved gland-wide continuity, which is particularly relevant for training in Gleason grading, Grade Group assignment, and evaluation of multifocality [37].

WMH has also been used as a reference for mpMRI radiologic-pathologic correlation in education, allowing trainees to align imaging findings with histopathologic reality and better understand how radiologic features correspond to underlying tissue architecture [14]. AI-driven tools add further educational value: explainable methods such as Grad-CAM heatmaps make model reasoning more visually accessible and can help learners understand which image regions drive predictions, supporting teaching of tumor recognition and spatial reasoning [30]. Synthetic whole-mount data generated with generative models have additionally been explored as a means of expanding and

standardizing training resources, offering exposure to rare or high-grade patterns where real data are scarce, though further validation is needed to ensure synthetic cases reflect real clinical material [37]. The principal clinical and translational applications of WMH and AI are summarized in **Table 1**.

Limitations and implementation barriers

Whole-mount sectioning of prostatectomy specimens provides unparalleled anatomical context and margin visualization, yet it remains a labor-intensive and costly procedure (**Table 2**). The increased size of tissue sections demands more complex fixation, embedding, and staining, resulting in prolonged turnaround times and additional QA steps, and comparative workflow analyses show that whole-mount preparation may increase processing time relative to standard sectioning [4, 6]. The specialized large-format slides and high-resolution scanners required for whole-mount imaging substantially raise operational expenses [6]. As such, these methods, while offering diagnostic precision, are often impractical for smaller laboratories or community pathology centers with limited infrastructure [56].

Limits of WMH itself as a ground truth

WMH is a high-fidelity but not infallible ground truth. Specimen orientation, fixation, large-format embedding, and pathologist annotation introduce systematic and observer-dependent variability. Interobserver agreement for Gleason pattern-4 subtypes (cribriform versus fused glands) and for the tumor-capsule interface remains imperfect even on whole-mount slides [31, 50]. The reliability of whole-mount RP histopathology as the ground truth for AI has been examined directly and found to depend on registration quality and reader agreement [22]. AI models trained or validated against WMH inherit these uncertainties, which is precisely why the framing of WMH as a *reference standard* must be qualified rather than absolute.

Domain shift across scanners, stains, and institutions

Performance of digital-pathology models trained on H&E whole-mount slides from one institution frequently drops when deployed elsewhere, because of scanner differences, sta-

in variation, and population mix [58]. WMH-trained models additionally face a 75 × 50 mm scanning constraint, since most clinical-grade scanners are built for 75 × 25 mm slides, which can fragment the very spatial continuity WMH is intended to preserve. Stain normalization, color augmentation, and foundation-model pre-training mitigate but do not eliminate the gap [59].

External validation and reporting standards

Many WMH-AI studies report internal cross-validation only. Reporting frameworks for medical-AI research, including TRIPOD-AI [16], CLAIM [17], DECIDE-AI [18], and STARD-AI [19], call for prospective external validation, multi-site cohorts, and predefined performance thresholds. Compliance in the prostate-AI literature is uneven, and few WMH-based studies report conformance. Transparent evidence synthesis additionally benefits from PRISMA-aligned reporting where systematic components are present [20].

Algorithmic bias

Prostate-AI training cohorts remain dominated by patients of Western-European ancestry, with underrepresentation of patients of African ancestry, a clinically critical gap given known differences in prostate-cancer incidence and mortality [2]. Bias audits are not yet standard practice, and equitable performance across age and ancestry should be reported explicitly.

Interpretability beyond saliency

Saliency methods such as Grad-CAM illustrate where a model attended but do not explain why a feature is causally informative [30]. Concept-level and pathologist-collaborative interfaces remain research-stage, and regulatory review increasingly expects mechanistic plausibility or strong external validation in lieu of mechanism.

Regulatory, interoperability, and reimbursement gaps

Clinical deployment of AI in prostate pathology is constrained by regulatory and interoperability challenges. Although multiple studies report diagnostic performance comparable to expert pathologists in some respects [12, 38], few AI systems have achieved formal regulatory clear-

Whole-mount histology as spatial ground truth

Table 1. Clinical and translational applications of whole-mount histopathology and artificial intelligence in prostate cancer

Domain	Role of whole-mount histopathology	Role of artificial intelligence/digital pathology	Representative clinical value	Representative sources
Index-lesion identification	Preserves full-gland spatial anatomy; supports localization of dominant lesions across the entire prostate	Automates tumor detection, segmentation, and lesion mapping on digitized whole-mount slides	Supplies the whole-mount-histopathology ground truth behind artificial-intelligence three-dimensional tumor maps that help define preoperative focal-therapy margins and magnetic-resonance-imaging targets	[23, 24]
Tumor-volume estimation	Enables more accurate contouring of tumor extent than fragmented sectioning	Supports automated segmentation and quantitative burden assessment	Strengthens postoperative risk stratification; whole-mount-validated artificial-intelligence volume maps also inform preoperative focal-therapy margins	[24-26]
Multifocal-disease characterization	Reveals discontinuous tumor foci, bilateral disease, and lesion-to-lesion relationships	Learns spatial patterns of multifocality; may detect subtle satellite lesions	Refines assessment of true intraprostatic disease burden	[23, 27]
Extraprostatic extension	Preserves the capsule-tumor interface and surrounding tissue planes	May support reproducible classification of capsular penetration	Refines pathologic (pT) staging and post-prostatectomy risk stratification; registered whole-mount labels also train models that predict extraprostatic extension preoperatively from magnetic resonance imaging or biopsy	[4, 28, 29]
Seminal vesicle invasion	Maintains continuity of prostate base, ejaculatory ducts, and seminal vesicles	Supports lesion localization and cross-modal correlation with imaging	Improves prognostication and postoperative therapy selection	[4, 30]
Surgical margin assessment	Permits continuous visualization of tumor proximity to inked surfaces across the gland	Enables pixel-level margin detection, margin-length quantification, and artifact discrimination	Improves risk assessment for residual disease and biochemical recurrence	[9, 31]
Gleason grading/Grade Groups	Provides whole-gland context for heterogeneous architectural patterns	Standardizes grading and detects high-grade foci with reduced interobserver variability	Supports prognostic classification and treatment planning	[32-34]
Radiology-pathology correlation	Serves as a reference standard for magnetic-resonance-imaging registration and lesion validation	Enables automated magnetic-resonance-imaging-histopathology alignment and explainable spatial mapping	Enhances imaging validation and multidisciplinary interpretation	[14, 30]
Quality assurance	Preserves surgeon- and specimen-level spatial outcome patterns	Facilitates structured analytics, margin dashboards, and standardized review	Supports institutional benchmarking and pathology audit workflows	[35, 36]
Education and training	Demonstrates whole-gland anatomy, zonal relationships, and tumor spread	Provides artificial-intelligence overlays, heatmaps, and synthetic training datasets	Improves trainee recognition of morphology and cross-disciplinary learning	[4, 37]

Whole-mount histology as spatial ground truth

Table 2. Implementation barriers and potential mitigations for whole-mount-histopathology-grounded artificial intelligence

Barrier	Practical impact	Why it matters	Potential mitigation	Representative sources
Increased labor and processing time	Larger sections require more complex fixation, embedding, staining, and review	Reduces feasibility in high-volume or resource-limited laboratories	Selective whole-mount use, workflow optimization, centralized processing hubs	[4, 6]
Higher cost	Large-format slides, scanners, storage, and computational infrastructure increase expenses	Limits adoption outside academic centers	Shared digital infrastructure and reimbursement-aligned deployment	[56, 57]
Specialized equipment and expertise	Requires large-format microtomes, scanning systems, and trained histotechnologists	Technical errors can reduce slide quality and diagnostic utility	Dedicated training programs and standardized technical protocols	[6]
Annotation variability	Expert labels may differ between pathologists and institutions	Weakens artificial-intelligence training consistency and external validity	Consensus annotation frameworks and multi-reader adjudication	[22, 31]
Scanner and file-format interoperability	Proprietary systems complicate data exchange and deployment	Impairs integration with laboratory information systems and multi-site studies	Open standards, vendor-neutral workflows, harmonized metadata	[58]
Generalizability of artificial-intelligence models	Performance may drop across scanners, stains, institutions, and populations	Restricts safe clinical translation	External validation and domain-adaptation strategies	[58, 59]
Regulatory barriers	Few systems cleared for routine autonomous use	Slows real-world implementation	Prospective validation, human-in-the-loop deployment, regulatory-grade evidence	[16, 18]
Data privacy and ethics	Large annotated datasets require secure governance and transparency	Raises concerns about bias, consent, and equitable access	Federated learning, de-identification, bias auditing	[60]
Outcome-linked validation gaps	Many tools are technically accurate but not yet tied to patient outcomes	Clinical utility remains uncertain	Longitudinal studies linking artificial-intelligence metrics to recurrence and treatment response	[12]

ance for routine diagnostic use, and clearance pathways differ (for example, 510(k) clearance, De Novo authorization, and Breakthrough Device designation in the United States, alongside CE marking in Europe). Specific cleared or designated prostate-pathology products and their intended-use statements should be checked at submission, as device status changes rapidly. Variability among imaging vendors, proprietary file formats, and inconsistent data standards hinder integration into laboratory information systems [58]. Emerging governance instruments, including risk-based AI regulation and predetermined-change-control frameworks for adaptive models, are directly relevant and should be cited specifically rather than gestured at [60]. Finally, digital-pathology and AI reimbursement remains incomplete relative to AI-augmented radiology; combined with the capital cost of large-format whole-mount-capable scanners, this is a dominant practical reason WMH-AI is concentrated in academic centers [57].

Data sharing, labeling variation, and ethics

AI models depend on large, annotated datasets from diverse populations; inconsistent labeling conventions and limited availability of expert-verified data remain significant obstacles. The PANDA challenge, encompassing thousands of digitized prostate slides, revealed substantial inter-laboratory variation that affected algorithm performance despite standardized protocols [13, 61]. Ethical considerations further complicate deployment, particularly regarding patient data privacy, model transparency, and algorithmic bias toward over-represented demographics [10, 60]. Balancing innovation with equity, interpretability, and regulatory compliance is essential to ensure that both whole-mount and AI-assisted approaches improve, rather than complicate, prostate cancer diagnosis.

Conclusion and future directions

Prostate cancer is defined by architectural heterogeneity, multifocality, and spatial complexity. Conventional diagnostic workflows fragment this spatial reality, introducing sampling error, grade discordance, and staging problems. WMH offers an anatomically faithful representation of the prostate gland, preserving tumor boundaries, capsular interfaces, margin

surfaces, and seminal-vesicle relationships. AI, particularly DL-based image analysis, can use this foundation to transform spatial structure into scalable, quantitative insight: as WMH improves accuracy through anatomical continuity, AI can improve reproducibility, objectivity, and efficiency. This combined framework links pathology, radiology, and surgical oncology by enabling tumor mapping, more standardized grading, objective margin assessment, and quantitative staging metrics, provided each AI claim is calibrated to the quality of the underlying spatial ground truth.

A critical near-term direction is consolidating and prospectively validating the translation of WMH-derived knowledge into preoperative, biopsy- and imaging-based workflows. Early validated examples already exist, such as focal-therapy margin definition and EPE prediction built on whole-mount-registered labels (Section 2.5) [24, 29]. AI models trained on spatially complete WMH datasets can be extended to infer whole-gland tumor burden, index-lesion dominance, and multifocal architecture from limited biopsy cores combined with imaging, reducing grade discordance between biopsy and prostatectomy and improving preoperative risk stratification [4, 9]. Registration pipelines that align histopathology with imaging are the mechanism that makes preoperative prediction of EPE, SVI, and tumor volume possible [14, 44]. A second direction is the integration of general-purpose pathology foundation models, including large self-supervised encoders such as UNI [62], CONCH [63], Virchow [64], and Prov-GigaPath [65], as feature backbones for WMH-based downstream tasks; how these models behave on large-format whole-mount inputs, and whether WMH pretraining improves robustness to domain shift, is an open and high-priority question. Pixel-level segmentation grounded in WMH can quantify tumor-to-margin distance and distinguish true involvement from artifact; integrated with longitudinal outcomes, such metrics could enhance prognostic modeling beyond binary margin status [9, 35]. Prospective, externally validated studies across diverse populations, reported against established medical-AI standards, will be essential to establish clinical utility and reproducibility [16, 18].

For WMH and AI to achieve widespread adoption, multi-institutional validation is necessary.

Variability in staining protocols, scanner hardware, annotation standards, and case mix continues to challenge generalizability; large-scale collaborative datasets, standardized segmentation frameworks, and external-validation cohorts will be required to ensure robustness across institutions and demographic groups [58, 59]. Regulatory pathways and reimbursement parity, alongside deliberate attention to disparities between academic and community settings, remain prerequisites for routine deployment [57]. Beyond immediate applications, three-dimensional reconstruction of whole-gland histology combined with imaging and molecular profiling may enable spatial modeling of tumor evolution and, ultimately, a more comprehensive computational understanding of prostate cancer biology.

Conclusion

Whole-mount histopathology enhances staging precision by preserving the prostate gland as a continuous anatomic structure, allowing comprehensive visualization of tumor boundaries, capsular relationships, seminal vesicles, and margin surfaces. Artificial intelligence builds on this foundation by standardizing interpretation, quantifying morphologic features, and enabling scalable integration with imaging and surgical data. Therefore, when AI claims are calibrated to the fidelity of the spatial ground truth, WMH and AI address fundamental limitations in sampling, variability, and spatial fragmentation that affect prostate cancer diagnosis and classification. Realizing the full potential of this combined approach will require rigorous external validation, regulatory clarity, equitable access strategies, and thoughtful integration into clinical workflows.

Acknowledgements

AI-Assisted Writing Statement. During the preparation of this manuscript, the authors used a generative artificial-intelligence assistant, Claude (Anthropic; Claude Opus 4 model family, accessed 2025-2026), as an AI-assisted writing tool. Its use was confined to language editing and polishing, namely improving grammar, clarity, concision, and consistency of style in text first drafted by the authors, and to the clerical formatting of references into the journal's citation style. In addition, ChatGPT 5.2 (OpenAI)

was used to generate **Figures 1** and **2**. These tools were not used to conceive the review, select or screen the literature, generate scientific claims, produce or compute any numerical result, or interpret the evidence. All literature selection, analysis, synthesis, claims, numerical values, citations, the conception and content of all figures, and conclusions were produced, checked against the primary sources, and approved by the authors. The authors reviewed and edited all AI-assisted output and take full responsibility for the integrity, accuracy, and content of the work. No AI system is listed as an author, and no AI system meets the criteria for authorship.

Disclosure of conflict of interest

The authors are affiliated with NovinoAI, a computational-pathology company. This review discusses commercial pathology-AI products (including Unfold AI/Avenda Health and externally validated commercial Gleason-grading systems) on the same evidentiary basis as non-commercial methods and does not endorse any product. Two cited works include NovinoAI-affiliated authors and are disclosed here as self-citations: Cui and Zhang (2021) and Zhu and colleagues (2024). The authors declare no other competing interests.

Abbreviations

AI, Artificial intelligence; BCR, Biochemical recurrence; CNN, Convolutional neural network; DL, Deep learning; DSC, Dice similarity coefficient; EPE, Extraprostatic extension; Grad-CAM, Gradient-weighted class activation mapping; H&E, Hematoxylin and eosin; IDC-P, Intraductal carcinoma of the prostate; mpMRI, Multiparametric magnetic resonance imaging; MRI, Magnetic resonance imaging; PI-RADS, Prostate Imaging Reporting and Data System; PSA, Prostate-specific antigen; PSM, Positive surgical margin; QA, Quality assurance; RP, Radical prostatectomy; SVI, Seminal vesicle invasion; TNM, Tumor-Node-Metastasis; WMH, Whole-mount histopathology; WSI, Whole-slide imaging.

Address correspondence to: Dr. David Y Zhang, Department of Computational Pathology, NovinoAI, 1443 NE Ave, Fort Lauderdale, FL 33304, USA. E-mail: david.zhang.nai@novinoai.com

References

- [1] Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I and Jemal A. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2024; 74: 229-263.
- [2] Jubber I, Ong S, Bukavina L, Black PC, Comp  rat E, Kamat AM, Kiemeney L, Lawrentschuk N, Lerner SP, Meeks JJ, Moch H, Necchi A, Panebianco V, Sridhar SS, Znaor A, Catto JWF and Cumberbatch MG. Epidemiology of bladder cancer in 2023: a systematic review of risk factors. *Eur Urol* 2023; 84: 176-190.
- [3] Srigley JR. Key issues in handling and reporting radical prostatectomy specimens. *Arch Pathol Lab Med* 2006; 130: 303-317.
- [4] Duan L, Liu Z, Wan F and Dai B. Advantage of whole-mount histopathology in prostate cancer: current applications and future prospects. *BMC Cancer* 2024; 24: 448.
- [5] Cimadamore A, Cheng L, Lopez-Beltran A, Mazzucchelli R, Lucian   R, Scarpelli M, Montorsi F and Montironi R. Added clinical value of whole-mount histopathology of radical prostatectomy specimens: a collaborative review. *Eur Urol Oncol* 2021; 4: 558-569.
- [6] Montironi R, Lopez Beltran A, Mazzucchelli R, Cheng L and Scarpelli M. Handling of radical prostatectomy specimens: total embedding with large-format histology. *Int J Breast Cancer* 2012; 2012: 932784.
- [7] Cui M and Zhang DY. Artificial intelligence and computational pathology. *Lab Invest* 2021; 101: 412-422.
- [8] Zhu M, Sali R, Baba F, Khasawneh H, Ryndin M, Leveillee RJ, Hurwitz MD, Lui K, Dixon C and Zhang DY. Artificial intelligence in pathologic diagnosis, prognosis and prediction of prostate cancer. *Am J Clin Exp Urol* 2024; 12: 200-215.
- [9] Duenweg SR, Brehler M, Bobholz SA, Lowman AK, Winiarz A, Kyereme F, Nencka A, Iczkowski KA and LaViolette PS. Comparison of a machine and deep learning model for automated tumor annotation on digitized whole slide prostate cancer histology. *PLoS One* 2023; 18: e0278084.
- [10] Agrawal S and Vagha S. A comprehensive review of artificial intelligence in prostate cancer care: state-of-the-art diagnostic tools and future outlook. *Cureus* 2024; 16: e66225.
- [11] Goldenberg SL, Nir G and Salcudean SE. A new era: artificial intelligence and machine learning in prostate cancer. *Nat Rev Urol* 2019; 16: 391-403.
- [12] Eminaga O, Abbas M, Kunder C, Tolkach Y, Han R, Brooks JD, Nolley R, Semjonow A, Boegemann M, West R, Long J, Fan RE and Betten-
dorf O. Critical evaluation of artificial intelligence as a digital twin of pathologists for prostate cancer pathology. *Sci Rep* 2024; 14: 5284.
- [13] Bulten W, Kartasalo K, Chen PC, Str  m P, Pinckaers H, Nagpal K, Cai Y, Steiner DF, van Boven H, Vink R, Hulsbergen-van de Kaa C, van der Laak J, Amin MB, Evans AJ, van der Kwast T, Allan R, Humphrey PA, Gr  nberg H, Samaratunga H, Delahunt B, Tsuzuki T, H  kkinen T, Egevad L, Demkin M, Dane S, Tan F, Valkonen M, Corrado GS, Peng L, Mermel CH, Ruusuvaari P, Litjens G and Eklund M; PANDA challenge consortium. Artificial intelligence for diagnosis and Gleason grading of prostate cancer: the PANDA challenge. *Nat Med* 2022; 28: 154-163.
- [14] Shao W, Vesal S, Soerensen SJC, Bhattacharya I, Golestani N, Yamashita R, Kunder CA, Fan RE, Ghanouni P, Brooks JD, Sonn GA and Rusu M. RAPHIA: a deep learning pipeline for the registration of MRI and whole-mount histopathology images of the prostate. *Comput Biol Med* 2024; 173: 108318.
- [15] Baethge C, Goldbeck-Wood S and Mertens S. SANRA-a scale for the quality assessment of narrative review articles. *Res Integr Peer Rev* 2019; 4: 5.
- [16] Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, Ghassemi M, Liu X, Reitsma JB, van Smeden M, Boulesteix AL, Camaradou JC, Celi LA, Denaxas S, Denniston AK, Glocker B, Golub RM, Harvey H, Heinze G, Hoffman MM, Kengne AP, Lam E, Lee N, Loder EW, Maier-Hein L, Mateen BA, McCradden MD, Oakden-Rayner L, Ordish J, Parnell R, Rose S, Singh K, Wynants L and Logullo P. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* 2024; 385: e078378.
- [17] Tejani AS, Klontzas ME, Gatti AA, Mongan JT, Moy L, Park SH and Kahn CE Jr; CLAIM 2024 Update Panel. Checklist for artificial intelligence in medical imaging (CLAIM): 2024 update. *Radiol Artif Intell* 2024; 6: e240300.
- [18] Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, Denniston AK, Faes L, Geerts B, Ibrahim M, Liu X, Mateen BA, Mathur P, McCradden MD, Morgan L, Ordish J, Rogers C, Saria S, Ting DSW, Watkinson P, Weber W, Wheatstone P and McCulloch P; DECIDE-AI expert group. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *Nat Med* 2022; 28: 924-933.
- [19] Sounderajah V, Ashrafian H, Golub RM, Shetty S, De Fauw J, Hooft L, Moons K, Collins G, Mo-

- her D, Bossuyt PM, Darzi A, Karthikesalingam A, Denniston AK, Mateen BA, Ting D, Treanor D, King D, Greaves F, Godwin J, Pearson-Stuttard J, Harling L, McInnes M, Rifai N, Tomasev N, Normahani P, Whiting P, Aggarwal R, Vollmer S, Markar SR, Panch T and Liu X; STARD-AI Steering Committee. Developing a reporting guideline for artificial intelligence-centred diagnostic test accuracy studies: the STARD-AI protocol. *BMJ Open* 2021; 11: e047709.
- [20] Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA, Stewart LA, Thomas J, Tricco AC, Welch VA, Whiting P and McKenzie JE. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ* 2021; 372: n160.
- [21] Chen MY, Woodruff MA, Dasgupta P and Rukin NJ. Variability in accuracy of prostate cancer segmentation among radiologists, urologists, and scientists. *Cancer Med* 2020; 9: 7172-7182.
- [22] Jager A, Postema AW, van der Linden H, Nooijen PTGA, Bekers E, Kweldam CF, Daures G, Zwart W, Mischi M, Beerlage HP and Oddens JR. Reliability of whole mount radical prostatectomy histopathology as the ground truth for artificial intelligence assisted prostate imaging. *Virchows Arch* 2023; 483: 197-206.
- [23] Karavitakis M, Winkler M, Abel P, Livni N, Beckley I and Ahmed HU. Histological characteristics of the index lesion in whole-mount radical prostatectomy specimens: implications for focal therapy. *Prostate Cancer Prostatic Dis* 2011; 14: 46-52.
- [24] Priester A, Fan RE, Shubert J, Rusu M, Vesal S, Shao W, Khandwala YS, Marks LS, Natarajan S and Sonn GA. Prediction and mapping of intraprostatic tumor extent with artificial intelligence. *Eur Urol Open Sci* 2023; 54: 20-27.
- [25] Yuk HD, Byun SS, Hong SK and Lee H. The tumor volume after radical prostatectomy and its clinical impact on the prognosis of patients with localized prostate cancer. *Sci Rep* 2022; 12: 6003.
- [26] Brisbane WG, Priester AM, Nguyen AV, Topoozian M, Mota S, Delfin MK, Gonzalez S, Grunden KP, Richardson S, Natarajan S and Marks LS. Focal therapy of prostate cancer: Use of artificial intelligence to define tumour volume and predict treatment outcomes. *BJU Compass* 2024; 6: e456.
- [27] Stabile A, Dell'Oglio P, De Cobelli F, Esposito A, Gandaglia G, Fossati N, Brembilla G, Cristel G, Cardone G, Deho' F, Losa A, Suardi N, Gaboardi F, Del Maschio A, Montorsi F and Briganti A. Association Between Prostate Imaging Reporting and Data System (PI-RADS) score for the index lesion and multifocal, clinically significant prostate cancer. *Eur Urol Oncol* 2018; 1: 29-36.
- [28] Bondarenko HD, Zanaty M, Harmouch SS, Negrean C, Pompe RS, Liberman D, Bhojani N, Karakiewicz PI, Zorn KC and El-Hakim A. External validation of the novel International Society of Urological Pathology (ISUP) Gleason grading groups in a large contemporary Canadian cohort. *Can Urol Assoc J* 2018; 12: 390-394.
- [29] Moroianu ŞL, Bhattacharya I, Seetharaman A, Shao W, Kunder CA, Sharma A, Ghanouni P, Fan RE, Sonn GA and Rusu M. Computational detection of extraprostatic extension of prostate cancer on multiparametric MRI using deep learning. *Cancers (Basel)* 2022; 14: 2821.
- [30] Gunashekar DD, Bielak L, Hägele L, Oerther B, Benndorf M, Grosu AL, Brox T, Zamboglou C and Bock M. Explainable AI for CNN-based prostate tumor segmentation in multi-parametric MRI correlated to whole mount histopathology. *Radiat Oncol* 2022; 17: 65.
- [31] Sadimin ET, Khani F, Diolombi M, Meliti A and Epstein JI. Interobserver reproducibility of percent Gleason pattern 4 in prostatic adenocarcinoma on prostate biopsies. *Am J Surg Pathol* 2016; 40: 1686-1692.
- [32] Epstein JI, Egevad L, Amin MB, Delahunt B, Srigley JR and Humphrey PA; Grading Committee. The 2014 International Society of Urological Pathology (ISUP) consensus conference on Gleason grading of prostatic carcinoma: definition of grading patterns and proposal for a new grading system. *Am J Surg Pathol* 2016; 40: 244-52.
- [33] Nagpal K, Foote D, Tan F, Liu Y, Chen PC, Steiner DF, Manoj N, Olson N, Smith JL, Mohtashami A, Peterson B, Amin MB, Evans AJ, Sweet JW, Cheung C, van der Kwast T, Sangoi AR, Zhou M, Allan R, Humphrey PA, Hipp JD, Gade-palli K, Corrado GS, Peng LH, Stumpe MC and Mermel CH. Development and validation of a deep learning algorithm for Gleason grading of prostate cancer from biopsy specimens. *JAMA Oncol* 2020; 6: 1372-1380.
- [34] Bulten W, Balkenhol M, Belinga JA, Brilhante A, Çakır A, Egevad L, Eklund M, Farré X, Geronatsiou K, Molinié V, Pereira G, Roy P, Saile G, Salles P, Schaafsma E, Tschui J, Vos AM; ISUP Pathology Imagebase Expert Panel, van Boven H, Vink R, van der Laak J, Hulsbergen-van der Kaa C and Litjens G. Artificial intelligence assistance significantly improves Gleason grading of prostate biopsies by pathologists. *Mod Pathol* 2021; 34: 660-671.

- [35] Kwak TY, Lee CH, Park WY, Ku JY, Jeong CW, Hwang EC, Choi SH, Cho J, Chang H, Kim KH, Kang BJ, Kim SW and Ha HK. Clinical implications of deep learning based image analysis of whole radical prostatectomy specimens. *Sci Rep* 2025; 15: 11006.
- [36] Grypari IM, Zolota V and Tzelepi V. Radical or not-so-radical prostatectomy: do surgical margins matter? *Cancers (Basel)* 2021; 14: 13.
- [37] Agosti V and Munari E. Histopathological evaluation and grading for prostate cancer: current issues and crucial aspects. *Asian J Androl* 2024; 26: 575-581.
- [38] Schmidt B, Soerensen SJC, Bhambhani HP, Fan RE, Bhattacharya I, Choi MH, Kunder CA, Kao CS, Higgins J, Rusu M and Sonn GA. External validation of an artificial intelligence model for Gleason grading of prostate cancer on prostatectomy specimens. *BJU Int* 2025; 135: 133-139.
- [39] Huang W, Randhawa R, Jain P, Iczkowski KA, Hu R, Hubbard S, Eickhoff J, Basu H and Roy R. Development and validation of an artificial intelligence-powered platform for prostate cancer grading and quantification. *JAMA Netw Open* 2021; 4: e2132554.
- [40] Priester A, Natarajan S, Khoshnoodi P, Margolis DJ, Raman SS, Reiter RE, Huang J, Grundfest W and Marks LS. Magnetic resonance imaging underestimation of prostate cancer geometry: use of patient specific molds to correlate images with whole mount pathology. *J Urol* 2017; 197: 320-326.
- [41] Sali R, Zhang V, Baba F, Zhu M, Zhang W, Zhang D. An innovative AI-powered digital pathology information system. *Ann Clin Pathol* 2025; 12: 1178.
- [42] Yu P, Liu N, Feng D, Jing Y, Xia M, Guo H, Yuan Y, Guo W, Alatan Y, Nie S, Zhao J, Su H, Miao Y and Miao Q. Deep learning model for predicting extraprostatic extension of prostate cancer based on H&E-stained biopsy digital images. *Ann Med* 2025; 57: 2547094.
- [43] Zhang X, Qi Y, Wang X, Chen H and Li J. Artificial intelligence-enabled imaging for predicting preoperative extraprostatic extension in prostate cancer: systematic review and meta-analysis. *J Med Internet Res* 2025; 27: e80981.
- [44] Shao W, Banh L, Kunder CA, Fan RE, Soerensen SJC, Wang JB, Teslovich NC, Madhuripan N, Jawahar A, Ghanouni P, Brooks JD, Sonn GA and Rusu M. ProsRegNet: a deep learning framework for registration of MRI and histopathology images of the prostate. *Med Image Anal* 2021; 68: 101919.
- [45] Sachdeva A, Veeratterapillay R, Voysey A, Kelly K, Johnson MI, Aning J and Soomro NA. Positive surgical margins and biochemical recurrence following minimally-invasive radical prostatectomy - an analysis of outcomes from a UK tertiary referral centre. *BMC Urol* 2017; 17: 91.
- [46] John A, Lim A, Catterwell R, Selth L and O'Callaghan M. Length of positive surgical margins after radical prostatectomy: does size matter? - A systematic review and meta-analysis. *Prostate Cancer Prostatic Dis* 2023; 26: 673-680.
- [47] John A, Milton T, Gupta A, Nguyen MT, Stretton B, Hewitt J, Virgin J, Kovoov J, Catterwell R, Selth L and Callaghan MO. Impact of positive surgical margin location after radical prostatectomy: a network meta-analysis. *World J Urol* 2025; 43: 134.
- [48] Izard JP, True LD, May P, Ellis WJ, Lange PH, Dalkin B, Lin DW, Schmidt RA and Wright JL. Prostate cancer that is within 0.1 mm of the surgical margin of a radical prostatectomy predicts greater likelihood of recurrence. *Am J Surg Pathol* 2014; 38: 333-338.
- [49] Koloff ZB, Hamstra DA, Wei JT, Montgomery JS, Tomlins SA, Wu AJ, Morgan TM, Siddiqui J, Paich K, Chinnaiyan AM, Feng FY, Weizer AZ, Kunju LP, Hollenbeck BK, Miller DC, Palapattu GS and Mehra R. Impact of tertiary Gleason pattern 5 on prostate cancer aggressiveness: lessons from a contemporary single institution radical prostatectomy series. *Asian J Urol* 2015; 2: 53-58.
- [50] Kweldam CF, Kümmerlin IP, Nieboer D, Verhoef EI, Steyerberg EW, Incrocci L, Bangma CH, van der Kwast TH, Roobol MJ and van Leenders GJ. Prostate cancer outcomes of men with biopsy Gleason score 6 and 7 without cribriform or intraductal carcinoma. *Eur J Cancer* 2016; 66: 26-33.
- [51] Cinar I and Cinar E. Assessment of interobserver variability in Gleason grading for prostate carcinoma. *North Clin Istanbul* 2025; 12: 337-343.
- [52] Campanella G, Hanna MG, Geneslaw L, Mirafior A, Werneck Krauss Silva V, Busam KJ, Brogi E, Reuter VE, Klimstra DS and Fuchs TJ. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat Med* 2019; 25: 1301-1309.
- [53] Ström P, Kartasalo K, Olsson H, Solorzano L, Delahunt B, Berney DM, Bostwick DG, Evans AJ, Grignon DJ, Humphrey PA, Iczkowski KA, Kench JG, Kristiansen G, van der Kwast TH, Leite KRM, McKenney JK, Oxley J, Pan CC, Samarasinghe H, Srigley JR, Takahashi H, Tsuzuki T, Varma M, Zhou M, Lindberg J, Lindskog C, Ruusuvaari P, Wählby C, Grönberg H, Rantalainen M, Egevad L and Eklund M. Artificial intelligence for diagnosis and grading of prostate

- cancer in biopsies: a population-based, diagnostic study. *Lancet Oncol* 2020; 21: 222-232.
- [54] Nagpal K, Foote D, Liu Y, Chen PC, Wulczyn E, Tan F, Olson N, Smith JL, Mohtashamian A, Wren JH, Corrado GS, MacDonald R, Peng LH, Amin MB, Evans AJ, Sangoi AR, Mermel CH, Hipp JD and Stumpe MC. Erratum: publisher correction: development and validation of a deep learning algorithm for improving Gleason scoring of prostate cancer. *NPJ Digit Med* 2019; 2: 113.
- [55] Bulten W, Pinckaers H, van Boven H, Vink R, de Bel T, van Ginneken B, van der Laak J, Hulsbergen-van de Kaa C and Litjens G. Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study. *Lancet Oncol* 2020; 21: 233-241.
- [56] Castellani D, Cecchini S, Mazzucchelli R, Soraci L, Di Rosa M, Fabbietti P, Palagonia E, Puccio F, Carnevali F, Paci E, Montironi R and Galosi AB. A magnetic resonance imaging-based nomogram for predicting clinically significant prostate cancer at radical prostatectomy. *Urol Oncol* 2022; 40: 379.e1-379.e8.
- [57] Wilkinson AN, Mainprize JG, Yaffe MJ, Robinson J, Cordeiro E, Look Hong NJ, Williams P, Moideen N, Renaud J, Seely JM and Rushton M. Cost-effectiveness of breast cancer screening using digital mammography in Canada. *JAMA Netw Open* 2025; 8: e2452821.
- [58] Howard FM, Dolezal J, Kochanny S, Schulte J, Chen H, Heij L, Huo D, Nanda R, Olopade OI, Kather JN, Cipriani N, Grossman RL and Pearson AT. The impact of site-specific digital histology signatures on deep learning model accuracy and bias. *Nat Commun* 2021; 12: 4423.
- [59] Tellez D, Litjens G, Bándi P, Bulten W, Bokhorst JM, Ciompi F and van der Laak J. Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. *Med Image Anal* 2019; 58: 101544.
- [60] Mennella C, Maniscalco U, De Pietro G and Esposito M. Ethical and regulatory challenges of AI technologies in healthcare: a narrative review. *Heliyon* 2024; 10: e26297.
- [61] Hopson S, Mildon C, Kubalek C, Ebbert J, Vance R, Laverty L, Urie P and Della Corte D. PANDA-PLUS: improved dataset of prostate whole slide images from PANDA challenge with pixel-level expert annotations. *J Pathol Inform* 2025; 20: 100540.
- [62] Chen RJ, Ding T, Lu MY, Williamson DFK, Jaume G, Song AH, Chen B, Zhang A, Shao D, Shaban M, Williams M, Oldenburg L, Weishaupt LL, Wang JJ, Vaidya A, Le LP, Gerber G, Sahai S, Williams W and Mahmood F. Towards a general-purpose foundation model for computational pathology. *Nat Med* 2024; 30: 850-862.
- [63] Lu MY, Chen B, Williamson DFK, Chen RJ, Liang I, Ding T, Jaume G, Odintsov I, Le LP, Gerber G, Parwani AV, Zhang A and Mahmood F. A visual-language foundation model for computational pathology. *Nat Med* 2024; 30: 863-874.
- [64] Vorontsov E, Bozkurt A, Casson A, Shaikovski G, Zelechowski M, Severson K, Zimmermann E, Hall J, Tenenholtz N, Fusi N, Yang E, Mathieu P, van Eck A, Lee D, Viret J, Robert E, Wang YK, Kunz JD, Lee MCH, Bernhard JH, Godrich RA, Oakley G, Millar E, Hanna M, Wen H, Retamero JA, Moye WA, Yousfi R, Kanan C, Klimstra DS, Rothrock B, Liu S and Fuchs TJ. A foundation model for clinical-grade computational pathology and rare cancers detection. *Nat Med* 2024; 30: 2924-2935.
- [65] Xu H, Usuyama N, Bagga J, Zhang S, Rao R, Naumann T, Wong C, Gero Z, González J, Gu Y, Xu Y, Wei M, Wang W, Ma S, Wei F, Yang J, Li C, Gao J, Rosemon J, Bower T, Lee S, Weerasinghe R, Wright BJ, Robicsek A, Piening B, Bifulco C, Wang S and Poon H. A whole-slide foundation model for digital pathology from real-world data. *Nature* 2024; 630: 181-188.