

REVIEW

Artificial Intelligence in equivocal prostatic needle biopsies: benign mimics and cancer detection

Ahmed Abdulwahab Bawahab

Department of Basic Medical Sciences, College of Medicine, University of Jeddah, Jeddah, Saudi Arabia.

Summary

Background: Equivocal prostatic needle biopsies are diagnostically challenging because minute atypical glandular foci may represent early acinar adenocarcinoma or benign mimics. This review critically examines the role of artificial intelligence (AI) in small-focus cancer detection, recognition of benign mimics, and immunohistochemistry (IHC) triage.

Methods: A structured narrative search of PubMed/MEDLINE, Scopus, Web of Science Core Collection, and Google Scholar was conducted from database inception through 15 May 2026. English-language, peer-reviewed diagnostic and validation studies, consensus recommendations, and focused reviews addressing prostate needle-biopsy interpretation, benign mimics, IHC, digital pathology, and clinically relevant AI methods were included.

Results: AI applied to digitized H&E slides can support cancer detection, grading, workflow prioritization, and quality assurance. In equivocal biopsies, its most credible near-term roles are localization of minute suspicious foci and safety-prioritized identification of cases in which IHC might be avoided or requested earlier. However, partial atrophy, adenosis, basal cell hyperplasia, seminal vesicle or ejaculatory duct epithelium, inflammation, treatment effects, and tissue or scanning artefacts may cause clinically important errors. Current evidence is mainly retrospective, and equivocal cases are often defined by previous IHC-ordering practice rather than uniform morphological criteria.

Conclusions: AI may improve the consistency and efficiency of equivocal prostate-biopsy assessment, but it should remain an assistive triage tool. Clinical use requires mimic-enriched external validation, conservative thresholds, interpretable localization, and pathologist-led correlation with morphology and IHC.

KEY WORDS: Artificial Intelligence (AI); Prostate needle biopsy; Prostate cancer; Atypical small acinar proliferation; Immunohistochemistry.

Submitted 18 May 2026; Accepted 5 July 2026

INTRODUCTION

Prostate needle biopsy remains the definitive method for the histopathological diagnosis of prostatic adenocarcinoma. Although many cases are straightforward, diagnostic uncertainty arises when only a minute focus of atypical glands is present, particularly in fragmented cores, tan-

genially sectioned tissue, crushed glands, or foci with limited cytological atypia. In such cases, the diagnosis depends on careful assessment of architectural and cytological features, including infiltrative small glands, nuclear enlargement, prominent nucleoli, and loss of basal cells, supported when necessary, by immunohistochemistry (1).

Equivocal prostatic needle biopsies occupy the clinically important gray zone between benign glands and unequivocal carcinoma. *Atypical small acinar proliferation* (ASAP) is the best-recognized example, describing a suspicious focus that is insufficient for a definitive diagnosis of malignancy. This category is important because a proportion of patients are subsequently diagnosed with prostate cancer on repeat biopsy, although the risk of clinically significant disease varies across cohorts (2). At the same time, overdiagnosis must be avoided, because several benign entities may closely simulate adenocarcinoma in limited biopsy material. These include partial atrophy, post-atrophic hyperplasia, adenosis/atypical adenomatous hyperplasia, basal cell hyperplasia, seminal vesicle or ejaculatory duct tissue, inflammatory and reactive atypia, and treatment-related changes (3, 4).

Immunohistochemistry (IHC) is central to the workup of suspicious small acinar proliferations. Basal-cell markers such as high-molecular-weight cytokeratin and p63, often combined with *alpha methylacyl CoA racemase* (AMACR), are recommended in small foci suspicious for adenocarcinoma when morphology alone is insufficient (5). However, immunohistochemistry is not infallible. It may be affected by tissue exhaustion, patchy or aberrant staining, loss of the suspicious focus on deeper sections, and the need to interpret staining results in the context of H&E morphology (1, 5, 6).

Artificial intelligence (AI) has emerged as a promising adjunct in digital prostate pathology. Deep-learning systems applied to whole-slide images have shown high performance in cancer detection, Gleason grading, tumour quantification, and workflow triage (7-9). International validation efforts have further shown that AI algorithms can approach pathologist-level performance across diverse prostate biopsy cohorts, although generalizability, dataset bias, and external validation remain critical considerations (10). The equivocal biopsy setting is therefore a particularly important test of AI performance: algorithms must not only detect small cancer foci but also

avoid false-positive interpretation of benign mimics. This narrative review focuses on the role of AI in equivocal prostatic needle biopsies, with emphasis on benign mimics, cancer detection, immunohistochemistry triage, limitations, and future directions.

METHODS

A structured narrative literature search was performed in PubMed/MEDLINE, Scopus, Web of Science Core Collection, and Google Scholar from database inception through 15 May 2026. Search strings combined prostate pathology terms ("prostate biopsy" OR "prostatic needle biopsy" OR "atypical small acinar proliferation" OR "ASAP" OR "benign mimic*" OR "adenosis" OR "partial atrophy" OR "basal cell hyperplasia") with diagnostic and ancillary-test terms ("adenocarcinoma" OR "immunohistochemistry" OR "AMACR" OR "p63" OR "high-molecular-weight cytokeratin") and computational terms ("artificial intelligence" OR "deep learning" OR "machine learning" OR "digital pathology" OR "whole-slide imaging" OR "multiple-instance learning" OR "foundation model" OR "explainability"). Boolean operators and database-specific syntax were used.

Eligible publications were English-language, peer-reviewed original diagnostic or validation studies, consensus recommendations, and focused reviews relevant to prostate needle-biopsy interpretation, equivocal or IHC-triggering cases, benign mimics, cancer detection, grading, workflow support, or clinically transferable computational pathology methods. Conference abstracts without adequate methodological or outcome data, duplicate reports, non-prostate studies without clear methodological relevance, and studies limited to radiology, cytology, prostatectomy, or molecular data without biopsy-pathology application were excluded. Titles and abstracts were screened for relevance, followed by full-text review of key studies and backward reference searching. Because the objective was interpretive synthesis rather than exhaustive effect estimation, formal risk-of-bias scoring and meta-analysis were not performed.

Benign mimics of prostatic adenocarcinoma

Benign mimics of prostatic adenocarcinoma represent one of the most important causes of diagnostic difficulty in prostatic needle biopsies, particularly when the suspicious focus is small, distorted, or present only on limited tissue levels. These mimics may arise from benign acinar proliferations, atrophic lesions, basal-cell proliferations, normal anatomic structures, inflammatory processes, or therapy-related changes (3, 4, 11). Their significance lies in the risk of false-positive cancer diagnosis, which may lead to unnecessary repeat procedures, anxiety, overtreatment, or inappropriate definitive therapy.

Atrophic lesions are among the most frequent benign mimics. Simple atrophy, partial atrophy, and post-atrophic hyperplasia may show crowded small glands with scant cytoplasm and an apparently infiltrative pattern, closely resembling small acinar adenocarcinoma. Post-atrophic hyperplasia has long been recognized as a histological mimic of prostatic adenocarcinoma, especially in needle biopsy material (12). Partial atrophy is particularly problematic because the glands may have

enlarged nuclei and only inconspicuous basal cells, creating overlap with low-grade carcinoma (3, 4).

Adenosis, also known as atypical adenomatous hyperplasia, is another important mimic. It is composed of crowded small to medium-sized glands and may resemble Gleason pattern 3 adenocarcinoma, especially when basal cells are difficult to identify on routine haematoxylin and eosin sections (13). Unlike carcinoma, adenosis usually shows a lobular or circumscribed architecture, variable basal-cell retention, and less infiltrative growth. Basal cell hyperplasia may also simulate malignancy, particularly when it forms small crowded glands, nests, or pseudoinfiltrative patterns with cytological atypia (11).

Normal anatomic structures may also be misinterpreted as carcinoma. Seminal vesicle and ejaculatory duct epithelium can be encountered in prostate biopsy specimens and may show complex architecture, small acini, prominent nuclear atypia, and pigment, all of which can be mistaken for malignant or premalignant prostatic lesions (14). Inflammatory and reactive atypia, granulomatous prostatitis, treatment-related changes after radiation or hormonal therapy, and rare lesions such as nephrogenic adenoma further broaden the differential diagnosis (3, 11).

Recognition of benign mimics requires correlation of low-power architecture, cytological detail, basal-cell presence, and immunohistochemical findings. Basal-cell markers and AMACR are helpful in difficult cases, but they must be interpreted cautiously because benign mimics may show patchy basal-cell staining or occasional AMACR positivity (5, 6). For artificial intelligence systems, these lesions are especially relevant because they represent potential sources of false-positive cancer detection. Therefore, AI models intended for equivocal prostate biopsies should be trained and validated on mimic-rich datasets rather than only on clear benign and malignant examples.

From an AI-error perspective, partial atrophy and adenosis are likely to be the most consequential benign mimics because both reproduce the small crowded acini, reduced luminal caliber, subtle nucleolar change, and apparent infiltrative arrangement learned by cancer-detection models. Basal cell hyperplasia may generate false-positive attention through glandular crowding and pseudoinfiltrative nests, whereas seminal vesicle/ejaculatory duct epithelium may be overcalled because of intrinsic nuclear atypia, complex architecture, and pigment.

Inflammation-related and treatment-related atypia can distort nuclei and gland contours, while crush, folds, tangential sectioning, and scant tissue can either create false alerts or conceal minute carcinoma. These patterns indicate that model evaluation should report mimic-specific errors rather than collapsing all non-neoplastic cores into a single "benign" category.

Current diagnostic approach:

H&E, immunohistochemistry, and pathologist judgment

The current diagnostic approach to equivocal prostatic needle biopsies remains pathologist-led and begins with careful assessment of haematoxylin and eosin (H&E)-stained sections. The diagnosis of prostatic adenocarcinoma is based on a combination of architectural and cytological findings rather than any single feature. At low

power, pathologists evaluate glandular crowding, infiltrative growth, loss of lobular architecture, relation to adjacent benign glands, and supportive findings such as perineural invasion or mucinous fibroplasia. At higher magnification, nuclear enlargement, prominent nucleoli, amphophilic cytoplasm, intraluminal crystalloids, blue mucin, and absence of a basal-cell layer may support malignancy (1, 15). However, these features may be incomplete in minute atypical foci or obscured by crush artefact, inflammation, tangential sectioning, or tissue fragmentation.

When H&E morphology is suspicious but insufficient for a definitive diagnosis, immunohistochemistry is used as an adjunct. The most common approach combines basal-cell markers, such as high-molecular-weight cytokeratin, CK5/6, or p63, with the cancer-associated marker α -methylacyl-CoA racemase (AMACR) (5, 16). Retained basal-cell staining supports a benign process, whereas complete basal-cell loss in a morphologically suspicious focus supports adenocarcinoma. AMACR positivity may provide additional evidence for carcinoma, but it should

not be interpreted alone because staining may be weak, focal, absent in some cancers, or present in some benign mimics (5, 17).

Triple-marker or cocktail stains are particularly useful in needle biopsies because they allow simultaneous assessment of basal-cell loss and AMACR expression on the same tissue level, helping preserve limited biopsy material (16). Nevertheless, immunohistochemistry has limitations. Basal-cell staining may be patchy in benign glands, AMACR may be nonspecific, and deeper sections may exhaust or lose the suspicious focus (5, 6, 17). Therefore, the final diagnosis depends on integration of morphology, immunophenotype, tissue context, clinical information, and pathologist judgment. Diagnoses may include benign, high-grade prostatic intraepithelial neoplasia, atypical small acinar proliferation, suspicious for adenocarcinoma, or definitive adenocarcinoma. This nuanced workflow explains why artificial intelligence should be regarded as a supportive tool for focus detection or IHC triage, not a replacement for expert pathological interpretation (Table 1).

Table 1.
Common immunohistochemical markers used in the diagnosis of prostatic adenocarcinoma (5, 16, 17).

Marker	Category	Typical staining	Main diagnostic use	Key limitations/pitfalls
p63	Basal-cell marker	Nuclear basal-cell staining	Confirms retained basal cells in benign glands; useful in small suspicious foci	May be patchy; rare aberrant p63-positive carcinoma reported
HMWCK/34 β E12/CK903	Basal-cell marker	Cytoplasmic basal-cell staining	Supports benign diagnosis when basal cells surround glands	Absence alone is not diagnostic; staining may be weak or discontinuous
CK5/6	Basal-cell marker	Cytoplasmic/membranous basal staining	Alternative or complementary basal-cell marker	Patchy staining; must correlate with H&E
AMACR / P504S	Cancer-associated marker	Cytoplasmic, often granular staining	Supports adenocarcinoma in a suspicious focus	Not fully specific; may stain HGPIN, adenosis, partial atrophy, and some benign glands
p63 + HMWCK + AMACR cocktail	Combined diagnostic panel	Basal cells plus AMACR on same section	Useful for limited atypical foci; preserves tissue level comparison	Requires H&E correlation; affected by tissue loss, nonspecific AMACR, and patchy basal staining
NKX3.1	Prostatic lineage marker	Nuclear staining	Highly sensitive marker for prostatic origin, especially metastatic/poorly differentiated carcinoma	Does not distinguish benign from malignant prostate tissue
PSA	Prostatic lineage marker	Cytoplasmic staining	Confirms prostatic differentiation	May be reduced or lost in high-grade, treated, or neuroendocrine tumors
PSAP	Prostatic lineage marker	Cytoplasmic staining	Supports prostatic origin, usually with PSA/NKX3.1	Less reliable than NKX3.1 in poorly differentiated tumors
Prostein/P501S	Prostatic lineage marker	Cytoplasmic, often Golgi-pattern	Supports prostatic origin in metastases	Does not distinguish benign from malignant glands
PSMA	Prostatic lineage/theranostic marker	Membranous/cytoplasmic staining	Supports prostatic origin; relevant in metastatic and theranostic settings	Not fully prostate-specific; may stain tumor neovasculature and some non-prostatic tumors
Androgen receptor	Lineage/predictive marker	Nuclear staining	Supports prostate lineage; relevant in advanced disease	Not prostate-specific; altered expression in advanced/neuroendocrine transformation
ERG	Molecular-associated marker	Nuclear tumor-cell staining	Supports carcinoma when positive; correlates with TMPRSS2-ERG fusion	Limited sensitivity; endothelial staining is an internal control/pitfall
PTEN	Tumor suppressor/prognostic marker	Cytoplasmic and nuclear staining; loss is abnormal	Prognostic relevance in selected cases	Not mainly used for small suspicious foci; requires internal controls
Ki-67	Proliferation marker	Nuclear staining	Assesses proliferation in selected aggressive or neuroendocrine tumors	Nonspecific; not routine for typical acinar adenocarcinoma
Chromogranin A	Neuroendocrine marker	Cytoplasmic staining	Used when neuroendocrine differentiation is suspected	May be focal; not useful for routine acinar carcinoma diagnosis
Synaptophysin	Neuroendocrine marker	Cytoplasmic staining	Helps identify neuroendocrine differentiation	Can be focal in conventional carcinoma
CD56	Neuroendocrine marker	Membranous staining	Adjunct in suspected neuroendocrine carcinoma	Less specific; should not be used alone
INSM1	Neuroendocrine marker	Nuclear staining	Sensitive marker for neuroendocrine carcinoma, including small-cell carcinoma	Use only when morphology suggests neuroendocrine carcinoma

Artificial Intelligence for cancer detection in needle biopsies

AI has become one of the most active areas of development in prostate biopsy pathology, particularly after the adoption of whole-slide imaging and deep-learning methods. In prostate needle biopsies, AI systems are typically trained to analyse digitized *haematoxylin and eosin* (H&E)-stained slides, detect cancer-containing regions, generate heatmaps of suspicious foci, assist Gleason grading, and support tumour quantification. These applications are especially relevant because prostate biopsy interpretation is high-volume, time-consuming, and vulnerable to diagnostic difficulty when cancer foci are small or morphologically subtle (7-10).

Several validation studies have shown that AI can detect prostatic adenocarcinoma in whole-slide images of core needle biopsies with high diagnostic performance. *Pantanowitz et al.* reported a blinded clinical validation and deployment study of an AI algorithm for prostate cancer diagnosis in digitized core needle biopsies, demonstrating the feasibility of integrating AI into a real diagnostic laboratory workflow (9). Similarly, *Raciti et al.* showed that an AI system increased the detection of prostate cancer in whole-slide images of needle core biopsies, supporting its potential role as a second-read or quality-assurance tool (18). Independent assessment has also demonstrated strong diagnostic accuracy of AI for prostate cancer detection, including use cases such as prescreening negative cases and identifying potentially missed carcinoma foci (19).

Beyond binary cancer detection, AI has been developed for grading and reporting-related tasks. *Ström et al.* demonstrated that AI could support diagnosis and Gleason grading of prostate cancer in biopsy material (8). The PANDA challenge further showed that AI systems could achieve performance comparable to pathologists for prostate cancer diagnosis and Gleason grading across large international datasets, highlighting the value of multicentre training and validation (10). These findings are important because clinically useful AI must not only identify cancer but also contribute to risk stratification through grade group estimation, tumour extent measurement, and recognition of clinically significant patterns.

In practical diagnostic workflows, AI may be used in several ways: as a screening tool to prioritize suspicious cases, as a second reader to reduce missed small foci, as a visual aid through heatmaps, or as an efficiency tool to reduce reporting time and unnecessary ancillary testing. AI-assisted diagnosis has been reported to improve diagnostic efficiency in prostatic biopsies, including reduced reading time and fewer requests for additional workup in some settings (20). More recently, AI models have been specifically investigated for equivocal prostate biopsy cases requiring immunohistochemistry, with the aim of reducing unnecessary stains while maintaining diagnostic safety (21).

Despite these advances, AI performance in routine cancer detection should not be assumed to translate automatically to equivocal or mimic-rich biopsies. Small atypical foci, partial atrophy, adenosis, inflammation, basal-cell hyperplasia, and treatment-related changes may generate false-positive or false-negative outputs if they are underrepresented in training datasets (3, 4, 11). Therefore, AI

should be regarded as an adjunctive diagnostic technology rather than an autonomous diagnostic authority. Its safest near-term role is likely to be focus detection, workflow triage, quality assurance, and support for pathologist decision-making, particularly when combined with expert review and appropriate immunohistochemical correlation (Table 2).

AI in the gray zone: Mimics, ASAP, and IHC triage

The most clinically meaningful application of AI in prostate needle biopsy pathology may not be the detection of obvious carcinoma, but decision support in the equivocal “gray zone,” where morphology is suspicious but insufficient for a definitive diagnosis. This gray zone includes ASAP, minute suspicious glandular foci, small atypical glands adjacent to high-grade prostatic intraepithelial neoplasia, atypical glands distorted by inflammation or crush artefact, and benign mimics such as partial atrophy, post-atrophic hyperplasia, adenosis, basal cell hyperplasia, seminal vesicle or ejaculatory duct epithelium, and treatment-related atypia (2-4, 11). These cases are diagnostically important because false-negative interpretation may delay cancer diagnosis, whereas false-positive diagnosis may expose patients to unnecessary treatment, anxiety, and downstream morbidity.

ASAP is central to this diagnostic space. It is not a specific biological entity but a descriptive diagnosis indicating that a focus is suspicious for carcinoma yet lacks sufficient qualitative or quantitative features for certainty. The problem is often compounded by the small size of the focus, loss of the suspicious glands on deeper levels, tangential sectioning, or incomplete immunohistochemical support (1, 2, 5). From an AI perspective, ASAP is challenging because the target abnormality may consist of only a few glands and may overlap visually with both low-grade acinar adenocarcinoma and benign mimics. Therefore, an AI system that performs well in routine cancer detection may still perform less reliably in ASAP-rich or mimic-rich cohorts unless these cases are deliberately represented during training and validation.

The relationship between AI and benign mimics is especially important. Many AI cancer-detection models learn visual features associated with glandular crowding, small acini, cytological atypia, altered luminal contours, and stromal infiltration (8-10, 18, 19). These same features may occur in benign mimics. Partial atrophy can show small crowded glands with pale cytoplasm and subtle nucleoli; adenosis can resemble Gleason pattern 3 carcinoma; basal cell hyperplasia may appear crowded and pseudoinfiltrative; and seminal vesicle epithelium may show nuclear atypia and complex architecture (3, 4, 11). If these mimics are underrepresented or labelled too broadly as “benign” without detailed annotation, AI may produce false-positive heatmaps that increase unnecessary pathologist concern or trigger avoidable immunohistochemistry. Conversely, overcorrection against benign mimics could reduce sensitivity for small-volume carcinoma.

In this setting, AI heatmaps should be understood as attention-guiding tools rather than diagnostic proof. A heatmap may successfully direct the pathologist to a small suspicious focus, but it cannot by itself establish whether

Table 2.
Selected AI studies relevant to prostate needle-biopsy diagnosis and equivocal/IHC-triage workflows.

Study	Cohort	Primary task	Sensitivity	Specificity	AUC	Clinical interpretation and limitations
<i>Ström et al. (8)</i>	Development study: 6,682 biopsy slides from 976 men; external validation: 271 slides from 93 men	Cancer detection, grading, and tumour-length estimation	NR	NR	0.997 independent test; 0.986 external validation	Large population-based development and independent external validation. Performance was reported mainly by AUC; the cohorts were not selected for ASAP, benign mimics, or IHC-triggering uncertainty.
<i>Pantanowitz et al. (9)</i>	External validation: 100 consecutive cases, 1,627 WSIs; cancer-versus-benign analysis: 355 biopsy parts	Cancer detection, grading, tumour extent, perineural invasion, and clinical second-read deployment	98.46%	97.33%	0.991	Strong blinded external validation and real-world deployment. Metrics were reported at biopsy-part level; ASAP parts were excluded from the cancer-versus-benign analysis, and atrophic or crushed glands generated many alerts.
<i>Raciti et al. (18)</i>	232 prostate core-biopsy WSIs reviewed by three pathologists, with and without AI assistance	Effect of AI assistance on pathologist cancer detection	74% without AI; 90% with AI	97% without AI; no significant change with AI	NR	AI significantly improved average pathologist sensitivity, especially for smaller and lower-grade tumours. This reader-assistance study did not independently validate AI performance in equivocal or IHC-selected cases.
<i>Perincheri et al. (19)</i>	1,876 prostate core-biopsy WSIs from an independent institution	Independent cancer detection and potential prescreening	97.7%	99.3%	NR	Strong independent-institution performance without local retraining. Poor-quality scans and very small, foamy-gland, treated, or perineural foci accounted for important misses; equivocal/IHC-triggering cases were not the primary endpoint.
<i>Chatrhan et al. (22)</i>	Development: 299 WSIs from 241 patients; retrospective validation: 91 cases (222 slides), with one representative WSI per case reviewed by three uropathologists	Prediction of whether IHC would be requested from H&E morphology	Not reported as a single clinically fixed value	Not reported as a single clinically fixed value	Mean 0.99 in threefold cross-validation	Proof-of-principle model predicting whether pathologists would request IHC rather than diagnosing cancer. Validation was single-centre, and interobserver agreement on the need for IHC was only 61–64%.
<i>Nateghi et al. (21)</i>	25,570 slides from 1,641 patients; prospectively collected cancer-detection validation subset included 1,737 slides from 95 patients	Screening IHC-ordered equivocal blocks and standalone cancer detection	95.0%*	97.8%*	0.985*	The screening models classified 55% of IHC-ordered equivocal blocks with a 1.4% error rate. The starred metrics relate to a separate standalone cancer-detection task and should not be interpreted as sensitivity or specificity for equivocal-block screening.
<i>Blilie et al. (23)</i>	510 IHC-selected WSIs across three sites: 234 internal-validation, 112 external-France, and 164 external-Switzerland	Cancer detection to reduce IHC use in difficult routine cases	100% at sensitivity-prioritized operating points	80.6%, 71.2%, and 52.3% at the reported safety-first thresholds	0.980, 0.993, and 0.951	At selected thresholds, projected IHC reductions were 44.4%, 42.0%, and 20.7%, with no false-negative predictions. The findings are retrospective, threshold- and site-dependent, and require prospective local validation before clinical use.

* Sensitivity, specificity, and AUC refer to Nateghi et al.'s standalone cancer-detection model, not to the 55% equivocal-block screening result. NR: not reported in a directly comparable form; WSI: whole-slide image.

the glands lack basal cells, whether AMACR staining is diagnostically meaningful, or whether the H&E architecture is truly infiltrative. The final diagnosis still requires pathologist integration of morphology, immunophenotype, tissue context, and clinical information (1, 5, 17). Thus, the safest role for AI in the gray zone is not autonomous classification but triage: identifying cases that need careful human review, flagging small foci that may otherwise be overlooked, and helping decide whether ancillary stains are likely to be useful.

AI-guided immunohistochemistry triage is a particularly relevant and rapidly developing application. In current practice, IHC is frequently requested when H&E findings are suspicious but not definitive, usually using basal-cell markers such as p63, high-molecular-weight cytokeratin or CK5/6, together with AMACR (5, 16, 17). Although this approach improves diagnostic confidence, it increases cost, delays reporting, adds technical workload, and may exhaust minute foci (5, 6). *Chatrhan et al.* provided an early proof-of-principle that AI could identify prostate biopsy cases likely to require IHC in advance, potentially allowing stains to be ordered before pathologist review and reducing duplicated diagnostic effort (22).

More recent work has moved beyond advance ordering

toward reducing unnecessary IHC in difficult cases. *Nateghi et al.* scanned 25,570 slides from 1,641 patients and defined the equivocal target population operationally as prostate biopsy blocks for which IHC had been ordered in routine practice, rather than by a uniform morphologic ASAP definition. Their complementary screening models correctly classified 55% of these challenging blocks with a 1.4% error rate, while a separate standalone cancer-detection model achieved an AUC of 98.5%, sensitivity of 95.0%, and specificity of 97.8% (21). The 55% result therefore means that slightly more than half of historically IHC-triggering blocks might have been resolved by the model under the study framework; it does not mean that the algorithm correctly diagnosed only 55% of all cancers. Important limitations are the institutionally developed data and workflow, the dependence of “equivocal” status on prior stain-ordering behavior, and the absence of broad prospective external deployment. The authors also found that a general-purpose foundation model screened more cases but produced higher error rates and required substantially greater computation, cautioning against assuming that larger pretrained models are automatically safer.

Blilie et al. evaluated 510 H&E whole-slide images from

three pathology sites (234, 112, and 164 slides), all selected because routine pathologists had required basal-cell IHC to finalize the diagnosis (23). Thus, “difficult” or “equivocal” again reflected real-world IHC use and varied suspicion thresholds, not a standardized single histologic category. The model achieved AUC values of 0.951–0.993. At a deliberately low cancer-probability threshold designed to maximize sensitivity, estimated total IHC use fell by 44.4%, 42.0%, and 20.7% across the sites without a false-negative prediction in those retrospective cohorts. Clinically, this is a rule-out/triage use case: the model may reassure the pathologist when both morphologic suspicion and AI probability are very low, but it should not discourage IHC when morphology remains concerning. The site-to-site variation, retrospective case enrichment, threshold dependence, and reported false-negative examples under non-sensitivity-prioritized settings indicate that prospective local calibration and monitoring remain necessary.

Nevertheless, several limitations should be emphasized. First, equivocal biopsy AI requires validation on the exact population in which it will be used: small atypical foci, ASAP, benign mimics, inflammation, treated tissue, and cases with discrepant or limited IHC. Second, performance should be evaluated at the block, slide, focus, and patient levels, because a clinically important carcinoma may occupy only a tiny portion of one core. Third, scanner type, staining variation, section thickness, tissue folds, and laboratory workflow may influence model performance. Fourth, AI thresholds should be conservative: in IHC triage, the acceptable false-negative rate must be extremely low because the consequence of not ordering IHC may be a missed small carcinoma. Finally, pathologist acceptance depends on explainability; models that provide interpretable heatmaps, uncertainty scores, and clear triage categories are more likely to be adopted than opaque binary outputs.

Therefore, AI in the gray zone should be framed as a safety-enhancing and efficiency-improving adjunct. Its strongest near-term role is likely to be second-read screening, detection of small suspicious foci, prioritization of difficult cases, and selective reduction or advance ordering of IHC. For clinical implementation, future studies should use multicentre mimic-enriched datasets, include external validation, compare AI decisions with expert genitourinary pathologists, and report false-negative outcomes specifically in ASAP and IHC-triggered cases. In equivocal prostatic needle biopsies, AI should not replace pathologist judgment; rather, it should help pathologists focus attention where the diagnostic risk is greatest.

Modern AI methodologies and explainability

Recent computational pathology has shifted from fully supervised patch classifiers toward weakly supervised *multiple-instance learning* (MIL), in which slide-level labels guide attention toward diagnostically informative image regions. This approach reduces the need for exhaustive gland annotation, but minute carcinoma foci may be diluted among thousands of benign patches; multi-resolution sampling, focus-enriched training, and hard-negative examples are therefore important (24). Pathology

foundation models such as UNI and Phikon use large-scale self-supervised pretraining to provide transferable image representations that may reduce task-specific labeling requirements (25, 26). Nevertheless, pretraining scale does not ensure robustness to scanner, stain, population, or task shifts. For equivocal biopsies, clinically useful explainability should include precise focus localization, uncertainty estimates, and quality-control warnings. Heatmaps may direct attention but cannot establish malignancy or replace morphologic and immunohistochemical interpretation.

Implementation, regulation, and economic considerations

Clinical implementation requires a clearly defined intended use, analytical and clinical validation across representative scanners and laboratories, data governance, cybersecurity, version control, user training, and monitoring for performance drift. In the United States, Paige Prostate received FDA De Novo authorization in 2021 as assistive software for digital prostate-biopsy review, illustrating that authorization is indication- and workflow-specific rather than approval of autonomous AI diagnosis (27). Integration must also address scanning capacity, storage, laboratory information-system interoperability, case routing, display of AI outputs, and management of AI-pathologist disagreement. Economic evaluation should balance potential reductions in IHC, tissue use, and turnaround time against the costs of scanners, storage, licenses, validation, maintenance, and quality assurance; prospective studies should report net cost per case together with false-negative outcomes and effects on patient management.

DECLARATIONS

Ethical approval and consent for participate: Not applicable because this narrative review did not involve human participants or animals.

Large Language Model Usage Statement: AI-assisted tools were used for language editing and editorial refinement of this submission. The author reviewed and approved the final content and remains responsible for the accuracy, integrity, and originality of the manuscript.

Competing interests: The author declares no conflicts of interest.

Funding: No specific funding was received for this work.

Authors' contributions: Ahmed Abdulwahab Bawahab was responsible for the conception, literature review, manuscript drafting, critical revision, and approval of the final manuscript.

Acknowledgments: None.

Availability of data and material: No previously published tables, figures, or other copyrighted material were reproduced in this manuscript.

CONCLUSIONS

AI can provide useful second-read support in equivocal prostate needle biopsies by localizing minute suspicious foci and assisting conservative IHC triage. Performance remains vulnerable in mimic-rich, artefact-affected, and very small-volume lesions, and the available evidence is predominantly retrospective. AI should therefore not independently classify ASAP or exclude carcinoma; clinical adoption requires external site-specific validation, safety-prioritized thresholds, transparent error reporting, and final interpretation by a pathologist integrating H&E morphology and IHC.

REFERENCES

1. Magi-Galluzzi C. Prostate cancer: diagnostic criteria and role of immunohistochemistry. *Mod Pathol.* 2018; 31(Suppl 1):S12-S21.
2. Leone A, Gershman B, Rotker K, et al. Atypical small acinar proliferation (ASAP): is a repeat biopsy necessary ASAP? A multi-institutional review. *Prostate Cancer Prostatic Dis.* 2016; 19:68-71.
3. Trpkov K. Benign mimics of prostatic adenocarcinoma. *Mod Pathol.* 2018; 31(Suppl 1):S22-S46.
4. Herawi M, Parwani AV, Irie J, Epstein JI. Small glandular proliferations on needle biopsies: most common benign mimickers of prostatic adenocarcinoma sent in for expert second opinion. *Am J Surg Pathol.* 2005; 29:874-880.
5. Epstein JI, Egevad L, Humphrey PA, Montironi R; Members of the ISUP Immunohistochemistry in Diagnostic Urologic Pathology Group. Best practices recommendations in the application of immunohistochemistry in the prostate: report from the International Society of Urologic Pathology Consensus Conference. *Am J Surg Pathol.* 2014; 38:e6-e19.
6. Mandel P, Wenzel M, Hoeh B, et al. Immunohistochemistry for prostate biopsy—impact on histological prostate cancer diagnoses and clinical decision making. *Curr Oncol.* 2021; 28:2123-2133.
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; 25:1301-1309.
8. Ström P, Kartasalo K, Olsson H, et al. Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study. *Lancet Oncol.* 2020; 21:222-232.
9. 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-e416.
10. Bulten W, Kartasalo K, Chen PHC, et al. Artificial intelligence for diagnosis and Gleason grading of prostate cancer: the PANDA challenge. *Nat Med.* 2022; 28:154-163.
11. Srigley JR. Benign mimickers of prostatic adenocarcinoma. *Mod Pathol.* 2004; 17:328-348.
12. Cheville JC, Bostwick DG. Postatrophic hyperplasia of the prostate. A histologic mimic of prostatic adenocarcinoma. *Am J Surg Pathol.* 1995; 19:1068-1076.
13. Epstein JI. Adenosis of the prostate. *Histopathology.* 1995; 27:379-380. doi not available.
14. Arista-Nasr J, Martinez-Benitez B, Viveros-Contreras C, et al. Seminal epithelium in prostate biopsy can mimic malignant and pre-malignant prostatic lesions. *Actas Urol Esp.* 2016; 40:17-22.
15. Humphrey PA. *Histopathology of prostate cancer.* Cold Spring Harb Perspect Med. 2017; 7:a030411.
16. Jiang Z, Woda BA, Rock KL, et al. Using an AMACR/P504S/34βE12/p63 cocktail for the detection of small focal prostate carcinoma in needle biopsy specimens. *Am J Clin Pathol.* 2005; 123:231-236.
17. Brimo F, Epstein JI. Immunohistochemical pitfalls in prostate pathology. *Hum Pathol.* 2012; 43:313-324.
18. Raciti P, Sue J, Ceballos R, et al. Novel artificial intelligence system increases the detection of prostate cancer in whole-slide images of core needle biopsies. *Mod Pathol.* 2020; 33:2058-2066.
19. Perincheri S, Levi AW, Celli R, et al. An independent assessment of an artificial intelligence system for prostate cancer detection shows strong diagnostic accuracy. *Mod Pathol.* 2021; 34:1588-1595.
20. Eloy C, Marques A, Pinto J, et al. Artificial intelligence-assisted cancer diagnosis improves the efficiency of pathologists in prostatic biopsies. *Virchows Arch.* 2023; 482:595-604.
21. Nateghi R, Zhou R, Saft M, et al. AI-driven screening of equivocal prostate IHC cases: development and validation of a screening and cancer detection framework. *Mod Pathol.* 2025; 38:100904.
22. Chatrian A, Colling RT, Browning L, et al. Artificial intelligence for advance requesting of immunohistochemistry in diagnostically uncertain prostate biopsies. *Mod Pathol.* 2021; 34:1780-1794.
23. Blilie A, Mulliqi N, Ji X, et al. Artificial intelligence-assisted prostate cancer diagnosis for reduced use of immunohistochemistry. *Commun Med (Lond).* 2025; 5:425.
24. Gadermayr M, Tschuchnig M. Multiple instance learning for digital pathology: a review of the state-of-the-art, limitations and future potential. *Comput Med Imaging Graph.* 2024; 112:102337.
25. Chen RJ, Ding T, Lu MY, et al. Towards a general-purpose foundation model for computational pathology. *Nat Med.* 2024; 30:850-862.
26. Filiot A, Ghermi R, Olivier A, et al. Scaling self-supervised learning for histopathology with masked image modeling. *Med Image Anal.* 2024; 91:103013.
27. U.S. Food and Drug Administration. De Novo classification request for Paige Prostate (DEN200080). Decision date: September 21, 2021. Accessed May 15, 2026.

Correspondence

Ahmed Abdulwahab Bawaha (Corresponding Author)
aabawahab@uj.edu.sa

Department of Basic Medical Sciences, College of Medicine,
University of Jeddah, Jeddah 23218, Saudi Arabia