

Subject: Artificial Intelligence-Based Software for Prostate Cancer Detection**Document #:** LAB.00049**Status:** Reviewed**Publish Date:** 01/03/2024**Last Review Date:** 11/09/2023

Description/Scope

This document addresses the use of artificial intelligence-based software that analyzes prostate biopsy slides to assist accurate cytopathologic diagnosis.

Note: Please see the following related document for additional information:

- [LAB.00026 Systems Pathology and Multimodal Artificial Intelligence Testing for Prostate Cancer](#)

Position Statement

Investigational and Not Medically Necessary:

Use of artificial intelligence-based software for prostate cancer detection is considered **investigational and not medically necessary** for all indications.

Rationale

Artificial intelligence-based software and devices are being developed to assist in cytopathologic diagnosis. Artificial intelligence (AI) is a field of computer science in which computers “learn” to perform tasks that are typically done by humans. The gold standard for prostate cancer diagnosis is to have a biopsy (diagnostic surgical pathology) read by a human pathologist. A biopsy is a procedure where small samples of tissue are removed and then examined under a microscope by a pathologist to see if they contain cancer cells. Hematoxylin and eosin (H&E) are two dyes commonly used in biopsy analysis. H&E is useful in distinguishing nuclear and cytoplasmic structures of cells. PIN4 is an immunohistochemistry staining technique that uses alpha-methylacyl coenzyme A racemase, tumor protein p63, and high-molecular-weight cytokeratin antibodies to identify cells with biomarkers for prostate adenocarcinoma.

One proposed application for AI is increasing the accuracy and efficiency of prostate biopsies. An example of this application is software that reviews scanned whole-slide images (WSIs) from prostate needle biopsies. The premise is the software detects the biopsied tissue suspicious for cancer and provides coordinates on the image for further review by a pathologist.

A 2020 industry-sponsored study by Raciti and colleagues reported on whether AI systems could accurately detect prostate cancer from digital WSIs of H&E-stained core needle biopsies. In the first phase of the study, three pathologists reviewed 304 anonymous H&E-stained WSI prostate needle biopsies. They classified each slide as either benign or cancerous. In the second phase of the study (approximately 4 weeks later), the same pathologists re-reviewed the slides with the aid of an AI software system. As a standalone tool, the AI software showed a sensitivity of 96% to detect cancer with a specificity of 98%. For the pathologists, the average sensitivity in detecting cancer without the use of AI was 73.8% with a specificity of 96.6%. With the aid of AI, the pathologist average sensitivity was 90.0% with specificity of 95.2%. The authors concluded that use of the AI system may improve sensitivity of diagnosing prostate cancer; however, none of the pathologists in the study had genitourinary pathology experience and the results for these three pathologists may not be generalizable to the broad practice of community-based pathology. Also, the included dataset was limited as it did not contain more benign mimickers of malignancy or a greater variety in Gleason grade. The pathologists only analyzed the slides alone, without ancillary studies or consultative second readings which may not be applicable to real world practice.

Also in 2020, Steiner and colleagues sought to validate an AI-based software tool to assist with prostate biopsy interpretation. This retrospective review reported findings by 20 general pathologists who reviewed 240 prostate needle biopsies from 240 individuals. No clinical information was provided from any of the study individuals. A cohort of 3 urologic pathologists reached consensus on Gleason grade scoring for each specimen. For each specimen, the specialist pathologists reviewed 3 slides stained with H&E and 1 slide stained with PIN4. Thirty generalist pathologists were then randomized to 1 of 2 study cohorts. One cohort used AI in their examinations and the other did not. The 2 cohorts reviewed every case in their assigned modality (with or without AI). Neither the published study nor its supplemental information clearly state whether the generalist pathologists reviewed the PIN4-stained slides. The modality switched after every 10 cases. The generalist pathologists reviewed the cases again after a 4-week washout period using the modality opposite to what they had previously used. The biopsies were interpreted based on International Society of Urological Pathology grading guidelines which gave information about Gleason grade groups and tumor and Gleason pattern quantitation. Grade group for each biopsy reported by the pathologist was compared to the majority opinion of urologic pathology subspecialists. Agreement with subspecialists for reviews unassisted by AI was 69.7%, and was 75.3% for assisted reviews. In terms of Gleason grade classification, unassisted general pathologist interpretations were incorrect 45.1% of the time. Assisted by AI, general pathologist interpretations were correct 78.1% of the time. Of the 61 biopsies with incorrect interpretation with AI, the agreement between pathologists and majority opinion of subspecialists was 45.1% for unassisted reviews and 38.0% for assisted reviews. For tumor detection, accuracy was 92.7% for general pathologist reviews unassisted by AI, 94.2% for general pathologist reviews assisted by AI, and 95.8% for the AI algorithm alone. For AI assisted general pathologist reviews, specificity for tumor detection was 96.1%. For general pathologist reviews unassisted by AI, specificity for tumor detection was 93.5%. Sensitivity for tumor detection for general pathologist reviews assisted by AI was 93.9% and 92.6% by general pathologist reviews not assisted by AI. In this study, the biopsies contained only one core biopsy per case so the impact of AI on multiple cores was not addressed. No demographic data was available for the individuals from whom biopsies were obtained. As this study was retrospective in design in a nonclinical setting, generalizability may not be possible and there is risk of population bias. The authors note the need for further prospective study of diverse populations to better understand the diagnostic benefits of this AI tool.

Nagpal and colleagues (2020) conducted a study in which they developed a deep learning system (DLS) to look at 752 digitized prostate biopsy specimens. The study compared performance of the DLS first to a panel of expert subspecialists and then to a panel of general pathologists. Each selected specimen was randomized to either the development set or the validation set. There was one specimen per case. The four participating laboratories used different staining protocols (H&E ± PIN4). The work began with development of a validation set. Pairs were selected from a group of six urologic subspecialists with an average of 25 years of

experience. Each specimen from a set of 752 validation set cases was independently assigned a Gleason grade by a pair of subspecialists. A third pathologist reviewed the specimen if there was disagreement between the initial reviewers. Subspecialist reviewers had access to four thin slice layers to aid in their diagnosis. Only one thin slice level was available to the DLS and general pathologists in later stages of the study. Development of the DLS began with characterization of 114 million slide regions from 1339 cases into Gleason patterns. The DLS then assigned a Gleason grade group to the entire biopsy specimen. The second stage of DLS development entailed training by using 580 biopsy specimen reviews. Iterative refinement of the DLS neural network algorithms led to development of tools to assist interpretation of digitized prostate biopsy images. In determining which biopsy specimens contained tumors and which did not in the 752 validation set cases, the rate of agreement with the expert subspecialists for the deep learning system was 94.3%. The rate of agreement between the specialist pathologists and general pathologists was 94.7%. There were 498 specimens determined to contain tumor tissue. The rate of agreement between the DLS and subspecialists was 71.7%. For determination of the Gleason grade for specimens with tumors, the rate of agreement of the general pathologists to expert subspecialists was 58.0%. In this study, only 1 biopsy specimen was used per case when typically a clinical case involves 12-18 specimens. There was also no evaluation of the Gleason grading with clinical outcomes. The authors conclude "Future work will need to assess the diagnostic and clinical effect of the use of a DLS for increasing the accuracy and consistency of Gleason grading to improve patient care."

Another industry-sponsored study by da Silva and colleagues (2021) reported on the diagnostic performance of an AI software system in the evaluation of WSIs taken from transrectal ultrasound-guided prostate biopsies. Using 600 previously diagnosed specimens from 100 individuals, two pathologists re-reviewed the slides and classified them as either benign, malignant, or suspicious. The slides were then run through the AI software. The software categorized the slides into benign or suspicious for cancer. In total, results were generated for 579 biopsies. The AI software classified 34.60% (200/579) slides as suspicious for cancer and 65.46% (379/579) as benign. Of the 579 specimens analyzed, there were 42 discordant results observed between the AI software and the pathologists (7.3%). In 3 of the specimens, AI rendered a diagnosis of benign whereas the original pathologist diagnosis was suspicious or malignant. After re-review by immunohistochemical staining, 2 of the specimens were prostate cancer and 1 was benign. In 6.7% (39/579) of specimens, the AI software rendered a diagnosis of suspicious where the original diagnosis by the pathologist was benign. Of these, 11 specimens contained cancer or lesions which would require further clinical intervention and 27 were noted to be benign. One of the specimens was an error in the database and was not included in the analysis. With a total of 41 discordant readings, at the individual level, the AI software rendered a sensitivity of 1.0 (confidence interval [CI], 0.93-1.0), negative predictive value (NPV) 1.0 (CI, 0.91-1.0), and a specificity of 0.78 (CI, 0.54-0.89). At the specimen level, the AI software rendered a sensitivity of 0.99 (CI, 0.96-1.0), NPV 1.0 (CI, 0.98-1.0), and specificity 0.93 (CI, 0.90-0.96). The 41 discordant specimens were also then read by another pathologist. Of the 39 specimens classified as suspicious by AI, there were 7 (17.9%) considered malignant by the pathologist, 18 (46.2%) were considered benign, and 14 (36%) were deferred due to small sample size or suboptimal image resolution. The 2 specimens classified as benign by AI were re-classified as malignant and suspicious by the pathologist. In reviewing previously reviewed specimens, not all part-specimens were analyzed for some of the individuals (due to technical issues with scanning or image transfer). In this study, there were approximately 12 cores per individual, however some protocols require 18 cores so the results of this study may not show an accurate estimate of sensitivity and NPV. The AI software only classified slides as benign or suspicious. There was no further classification of specific descriptive diagnoses or grading of prostate cancers that were detected. This study did not prospectively evaluate the ability of AI-assisted cytopathologic analysis to affect individual-level health outcomes. Further studies are necessary to examine the potential benefits of this technology.

Another 2021 industry-sponsored study by Perincheri and colleagues reported the ability of an AI system to categorize WSIs of prostate core biopsies obtained at a tertiary academic center different from the center that developed the AI algorithms. There were 1876 prostate core biopsy diagnoses which had been previously established for 118 individuals. AI categorizations were compared to these previously obtained pathology diagnoses. Standard practice at the study's institution was to perform 20 or more core samples during prostate biopsy. The tissue diagnosis was made in each case by a board certified pathologist. As in the Nagpal study cited above, the pathologists had three levels of H&E-stained slides and possibly one or two PIN4-stained slides to review, whereas the AI system only examined one H&E-stained WSI. The original pathology diagnosed 86 of the 118 individuals with prostate cancer in at least 1 core. No prostate cancer had been pathologically diagnosed in any core sample for the remaining 32 individuals. In the 86 individuals found to have prostate cancer, AI analysis categorized at least 1 core as suspicious in 84. In the 32 individuals reported as not having cancer by original pathology, AI did not categorize suspicious cores in 26 of the 32 individuals. Of the 1876 core biopsies, there was a discrepancy between the pathology diagnosis and AI in 80 cores. Of these 80 discrepant cores, 46 were further analyzed and categorized as not suspicious by AI but final diagnosis was cancer. The remaining 34 discordant cores were categorized as suspicious by AI but final diagnosis was benign. The discordant 46 cores deemed not suspicious by AI yielded a NPV of 96.7% and specificity of 97.6%. Further review revealed that 35 discordant images were unable to be interpreted manually due to being out of focus or due to slide or tissue issues. The authors assert that removing discordant cores with technical issues improved the positive predictive value (PPV), but this was not an a priori declared analysis. This study had two stated aims 1) to assess whether the AI system could identify core samples that do not require manual review and 2) to identify whether the AI system can serve as a reliable "second read" to identify suspicious areas that may not have been detected on manual review. Neither aim was conclusively demonstrated. Because this study was also conducted in a single tertiary referral center with a high prevalence of malignancy, these results may not be generalizable to other practices and cohorts. Further study is needed.

In a 2023 study by Eloy and colleagues, they present the results of an AI system comparing the diagnostic performance of four pathologists diagnosing core needle biopsy specimens by usual means (phase 1), then in the second phase assisted by an AI system. There were 4 pathologists who read 105 core needle biopsies, then after a 2 week washout period, the same pathologists re-evaluated the same cases assisted by AI. During the initial reading in phase 1, the pathologists had a global diagnostic accuracy of 95% for the diagnosis of prostate cancer. The average sensitivity was 0.968, specificity was 0.939, PPV was 0.909, and NPV was 0.982. In phase 2, with the assistance of AI, the pathologists had a similar global diagnostic accuracy of 93.81%. The phase 2 results were also similar with a sensitivity of 0.955, specificity of 0.928, PPV 0.892, and NPV of 0.974. Of the 105 core needle biopsies, 66 were benign and 39 were positive for prostate cancer (all were acinar adenocarcinoma). There were fewer IHC studies requested in phase 2 (36.43%) compared to phase 1 (45.95%) and fewer second opinions in phase 2 (7.38%) vs. 12.14% in phase 1. Median turnaround time required for reading and reporting each slide was 139.00s in phase 1 and 108.50s in phase 2. The four pathologists who participated in this study had 2 years' experience working on the digital platform. These results may not be generalizable and translate to diverse populations.

Raciti and colleagues (2023) reported on the diagnostic accuracy of pathologists who read WSIs of prostate biopsies with and without AI assistance. Included were 610 prostate needle biopsy WSIs stained with H&E. There were 16 pathologists who read the WSIs initially unassisted, then immediately read the slides again with the assistance of AI. For the WSIs which were read unassisted, the average sensitivity in the unassisted group was 88.7% with an average specificity of 97.3%. The average sensitivity in the AI assisted group was 96.6% with an average specificity of 98.0%. Sensitivity for non-genitourinary (GU) pathologists for unassisted reads was an 8.5% gain and 3.9% gain for GU pathologists. There were also gains in specificity among non-GU and GU pathologists, 0.7% and 0.3% respectively. In this study, pathologists reviewed one H&E-stained WSI without clinical or radiologic context. The authors note that future studies should include complete cases.

The 2023 National Comprehensive Cancer Network® Clinical Practice Guidelines in Oncology (NCCN Guidelines®) for prostate cancer early detection does not address the use of AI to assist in cytopathologic diagnosis.

In 2023 the American Urologic Association published a guideline for early detection of prostate cancer (Wei, 2023). AI is not included in their recommendations and notes that the use of AI requires further study.

Current literature addressing use of AI to assist prostate cancer detection is limited to industry-sponsored studies that are mostly retrospective in design and mostly conducted in single academic centers. Prospective, multicenter trials involving diverse populations and a variety of treatment settings are needed to permit reasonable conclusions about the possible health benefits of this technology outside of a research setting.

Background/Overview

According to the American Cancer Society, in the year 2023 there were 288,300 new cases of prostate cancer diagnosed and approximately 34,700 deaths from prostate cancer. Approximately 1 in 8 individuals will be diagnosed with prostate cancer in their lifetime and about 1 in 41 will die from the disease. While prostate cancer is a serious disease, most individuals don't die from it. There are approximately 3.1 million individuals in the United States with a diagnosis of prostate cancer who are still alive.

The gold standard for diagnosis of prostate cancer is a prostate biopsy. According to the National Cancer Institute (NCI) (2023):

Needle biopsy is the most common method used to diagnose prostate cancer. Most urologists perform a transrectal biopsy using a bioptic gun with ultrasound guidance. Less frequently, a transperineal ultrasound-guided approach can be used for patients who may be at increased risk of complications from a transrectal approach. Over the years, there has been a trend toward taking eight to ten or more biopsy samples from several areas of the prostate with a consequent increased yield of cancer detection after an elevated PSA blood test, with a 12-core biopsy now standard practice.

In 2021 the United States Food and Drug Administration granted de novo request approval for the Paige Prostate (Paige.AI, Inc. New York, NY), a software algorithm device which assists users in digital pathology. The software is intended to evaluate whole slide images already acquired and provide information to the user about the presence, location and characteristics of areas on the image with potential clinical implications. There are other AI software devices also being developed.

Definitions

Artificial Intelligence (AI): A science of computer simulated thinking processes and human behaviors, which involves computer science, psychology, philosophy and linguistics.

Biopsy: The removal of a sample of tissue for examination under a microscope for diagnostic purposes.

Deep Learning: A branch of AI that uses multiple layers of algorithms to derive information from unstructured information.

Prostate: A walnut-shaped gland that extends around the urethra at the neck of the urinary bladder and supplies fluid that goes into semen.

Coding

The following codes for treatments and procedures applicable to this document are included below for informational purposes. Inclusion or exclusion of a procedure, diagnosis or device code(s) does not constitute or imply member coverage or provider reimbursement policy. Please refer to the member's contract benefits in effect at the time of service to determine coverage or non-coverage of these services as it applies to an individual member.

When services are Investigational and Not Medically Necessary:

For the following procedure code or when the code describes a procedure indicated in the Position Statement section as investigational and not medically necessary.

CPT

88399	Unlisted surgical pathology procedure [when specified as use of an AI-based software product for cytopathologic prostate cancer detection]
-------	--

ICD-10 Diagnosis

All diagnoses

References

Peer Reviewed Publications:

1. Bulten W, Pinckaers H, van Boven H, et al. Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study. *Lancet Oncol.* 2020; 21:233–241.
2. da Silva LM, Pereira EM, Salles PG, et al. Independent real-world application of a clinical-grade automated prostate cancer detection system. *J Pathol.* 2021; 254(2):147-158.
3. 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(3):595-604.
4. Nagpal K, Foote D, Tan F, et al. Development and validation of a deep learning algorithm for Gleason grading of prostate cancer from biopsy specimens. *JAMA Oncol.* 2020; 6:1372-1380.
5. 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(8):1588-1595.
6. 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(10):2058-2066.
7. Raciti P, Sue J, Retamero JA, et al. Clinical validation of artificial intelligence-augmented pathology diagnosis demonstrates significant gains in diagnostic accuracy in prostate cancer detection. *Arch Pathol Lab Med.* 2023; 147(10):1178-1185.
8. Singhal N, Soni S, Bonthu S, et al. A deep learning system for prostate cancer diagnosis and grading in whole slide images of core needle biopsies. *Sci Rep.* 2022; 12(1):3383.
9. Steiner DF, Nagpal K, Sayres R, et al. Evaluation of the use of combined artificial intelligence and pathologist assessment to review and grade prostate biopsies. *JAMA Netw Open.* 2020; 3(11): e2023267.
10. Ström P, Kartasalo K, Olsson H, et al. Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a

Government Agency, Medical Society, and Other Authoritative Publications:

1. NCCN Clinical Practice Guidelines in Oncology®. 2023 National Comprehensive Cancer Network, Inc. For additional information visit the NCCN website: <http://www.nccn.org/index.asp>. Accessed on September 20, 2023.
 - Prostate Cancer Early Detection (V1.2023). Revised January 9, 2023.
2. U.S. Food and Drug Administration De Novo Database. Paige Prostate. DEN200080. Rockville, MD: FDA. September 21, 2021. Available at: https://www.accessdata.fda.gov/cdrh_docs/pdf20/DEN200080.pdf. Accessed on September 20, 2023.
3. Wei JT, Barocas D, Carlsson S, et al. Early detection of prostate cancer: AUA/SUO guideline part I: prostate cancer screening. J Urol. 2023; 210(1):45-53.

Websites for Additional Information

1. American Cancer Society. Prostate cancer. Available at: <https://www.cancer.org/cancer/prostate-cancer.html>. Accessed on September 20, 2023.
2. National Cancer Institute (NCI). Physician Data Query (PDQ®) Prostate Cancer Screening. Last updated May 22, 2023. Available at: <https://www.cancer.gov/publications/pdq>. Accessed on September 20, 2023.

Index

Artificial Intelligence
Paige Prostate

The use of specific product names is illustrative only. It is not intended to be a recommendation of one product over another, and is not intended to represent a complete listing of all products available.

Document History

Status	Date	Action
Reviewed	11/09/2023	Medical Policy & Technology Assessment Committee (MPTAC) review. Updated Description/Scope, Rationale, Background/Overview, References, and Websites for Additional Information sections.
Reviewed	11/10/2022	MPTAC review. Updated Rationale and References sections.
New	08/11/2022	MPTAC review. Initial document development.

Applicable to Commercial HMO members in California: When a medical policy states a procedure or treatment is investigational, PMGs should not approve or deny the request. Instead, please fax the request to Anthem Blue Cross Grievance and Appeals at fax # 818-234-2767 or 818-234-3824. For questions, call G&A at 1-800-365-0609 and ask to speak with the Investigational Review Nurse.

Federal and State law, as well as contract language, including definitions and specific contract provisions/exclusions, take precedence over Medical Policy and must be considered first in determining eligibility for coverage. The member's contract benefits in effect on the date that services are rendered must be used. Medical Policy, which addresses medical efficacy, should be considered before utilizing medical opinion in adjudication. Medical technology is constantly evolving, and we reserve the right to review and update Medical Policy periodically.

No part of this publication may be reproduced, stored in a retrieval system or transmitted, in any form or by any means, electronic, mechanical, photocopying, or otherwise, without permission from the health plan.

© CPT Only – American Medical Association