Am J Clin Exp Urol 2024;12(4):200-215
www.ajceu.us /ISSN:2330-1910/AJCEU0159769

Review Article

Artificial intelligence in pathologic diagnosis, prognosis and prediction of prostate cancer

Min Zhu[1], Rasoul Sali[1,2], Firas Baba[1], Hamdi Khasawneh[3], Michelle Ryndin[4], Raymond J Leveillee[5], Mark D Hurwitz[6], Kin Lui[7], Christopher Dixon[8], David Y Zhang[1,9]

[1]Department of Computational Pathology, NovinoAI, 1443 NE 4th Ave, Fort Lauderdale, FL 33304, USA; [2]Department of Radiation Oncology, Stanford University School of Medicine, Stanford, CA 94305, USA; [3]King Hussein School of Computing Sciences, Princess Sumaya University for Technology, Amman 11855, Jordan; [4]College of Agriculture and Life Sciences, Cornell University, 616 Thurston Ave, Ithaca, NY 14853, USA; [5]Department of Surgery, Florida Atlantic University, Division of Urology, Bethesda Hospital East, Baptist Health South Florida, 2800 S. Seacrest Drive, Boynton Beach, FL 33435, USA; [6]Department of Radiation Medicine, New York Medical College and Westchester Medical Center, Valhalla, NY 10595, USA; [7]Department of Urology, Mount Sinai Hospital, New York, NY 10029, USA; [8]Department of Urology, Good Samaritan Hospital, Westchester Medical Center Health Network, Suffern, NY 10901, USA; [9]Pathology and Laboratory Services, Department of Veterans Affairs New York Harbor Healthcare System, New York, NY 10010, USA

Received August 12, 2024; Accepted August 19, 2024; Epub August 25, 2024; Published August 30, 2024

Abstract: Histopathology, which is the gold-standard for prostate cancer diagnosis, faces significant challenges. With prostate cancer ranking among the most common cancers in the United States and worldwide, pathologists experience an increased number for prostate biopsies. At the same time, precise pathological assessment and classification are necessary for risk stratification and treatment decisions in prostate cancer care, adding to the challenge to pathologists. Recent advancement in digital pathology makes artificial intelligence and learning tools adopted in histopathology feasible. In this review, we introduce the concept of AI and its various techniques in the field of histopathology. We summarize the clinical applications of AI pathology for prostate cancer, including pathological diagnosis, grading, prognosis evaluation, and treatment options. We also discuss how AI applications can be integrated into the routine pathology workflow. With these rapid advancements, it is evident that AI applications in prostate cancer go beyond the initial goal of being tools for diagnosis and grading. Instead, pathologists can provide additional information to improve long-term patient outcomes by assessing detailed histopathologic features at pixel level using digital pathology and AI. Our review not only provides a comprehensive summary of the existing research but also offers insights for future advancements.

Keywords: Artificial intelligence, machine learning, prostate cancer, pathology, diagnosis, grading, prognosis, prediction, treatment

Overview of prostate cancer

According to GLOBOCAN, prostate cancer was the 4th most frequently diagnosed cancer in 2022 (1,466,680 new cases, 7.3% of all cancer globally). The highest incidence rates are seen in Northern Europe, Australia/New Zealand, the Caribbean, and North America, and it is the most frequently diagnosed cancer among men in almost two thirds (118 of 185) of the world's countries [1]. In the United States, it is estimated that prostate cancer, lung and bronchus cancer, and colorectal cancer account for almost one half (48%) of all incident cases in men, with prostate cancer alone accounting for 29% of diagnoses in 2024. It is the 2nd leading cause of cancer death in men [2].

The occurrence of prostate cancer is the result of the combined effects of genetic and environmental factors. Some of the risk factors are nonmodifiable, such as age, race, family history, and gene mutations [3]. The varying incidence of prostate cancer among different families and racial groups suggests the significant role of inherited or genetic factors. Genome-

https://doi.org/10.62347/JSAE9732

wide association studies have identified more than 290 single nucleotide variants associated with prostate cancer [4], with the most common aberration TMPRSS2-ERG fusions, SPOP loss-of-function mutations, FOXA1 gain-of-function mutations, PTEN-P53-RB1 mutations, as well as DNA damage response gene BRCA1, BRCA2, MLH1, MSH2, PSM2, ATM which are also important in prostate cancer [5]. On the other hand, there are some of the risk factors which are modifiable, such as cigarette smoking, ultraviolet rays' exposure, obesity, and metabolic syndrome [6].

The vast majority of prostatic cancers are acinar adenocarcinomas [7]. The non-acinar types account for about 5-10% of prostate-origin carcinomas [7]. The progression and prognosis of each histological type of prostate cancer vary significantly, and some rare types, such as treatment-emergent neuroendocrine prostate cancer, progressed from castration-resistant prostate cancer, are characterized by an aggressive clinical course and a poor overall prognosis [8, 9].

Artificial intelligence (AI) is the ability of machines to recognize patterns from training data, and then apply the learned representations on new "unseen" data in order to make decisions such as classification and prediction [10]. AI has been applied in pathology due to the machine enabled extraction of quantitative features from digital pathology slides using high dimensional and predefined computational operations [10]. Prostate cancer serves as a significant example for highlighting the potential advantages of machine learning (ML). In this review, we will look back on the historical progress and latest developments of AI in the diagnosis, prognosis, and treatment of prostate cancer. In addition, we will also present some examples of how AI can help in routine prostate biopsy practices. Our review aims to provide a comprehensive perspective on the development of AI in the field of prostate cancer pathology and offer insights for future advancements.

AI principles

In this section, we provide a very concise overview of some key concepts in AI that are referenced throughout this paper. Our aim is to offer readers a brief familiarity with these terms and ideas, ensuring they can follow the discussions and analyses in the subsequent sections. This overview is not intended to be comprehensive but rather to highlight the essential concepts necessary for understanding the content of this paper.

Comparative analysis of AI, ML, and deep learning

AI is a broad field where machines perform tasks adeptly, such as learning, problem-solving, and decision-making. It involves mimicking cognitive functions associated with human intelligence. AI can be classified into two types: narrow AI [11] and general AI [12]. Narrow AI, or weak AI, is designed to excel at a specific task, like diagnosing diseases from medical images or creating personalized treatment plans from patient data. It operates within well-defined constraints, making it highly specialized in its field. On the other hand, general AI, or strong AI, aims to apply intelligence across multiple domains, similar to human intelligence. However, achieving this level of AI remains a distant goal.

ML is a subset of AI that trains algorithms to learn from data and make predictions or decisions without being explicitly programmed for specific tasks [13]. In healthcare, ML has numerous applications, such as predicting patient outcomes, personalizing treatment plans, and identifying patterns in medical data. ML algorithms learn by recognizing patterns in data and making decisions based on the information they process. For example, ML can predict the presence of a disease in a patient using diagnostic tests and historical patient records, or it can group patients with similar symptoms to target treatments more effectively. This approach allows for the continuous optimization of treatment plans by learning from individual patient responses to various interventions and their outcomes over time.

Deep learning (DL) refers to a niche of ML, where multiple-layered neural networks are employed to analyze varied forms of information [14]. Every layer in a neural network works on part of the input data and this is what enables the system to learn complicated patterns and representations. It is remarkable how DL has altered various domains, such as healthcare, through its excellent performance

in major operations including identifying images, processing natural languages, and making predictions. As a result, neural networks which form the basis for DL have nodes (neurons) that connect to each other within layers. Within those nodes, information is processed and then forwarded to other layers. This hierarchy in data representation allows DL models to tackle difficult tasks such as predicting disease outcomes based on medical imaging analysis or genomic models for sequencing DNA strings with ease.

Healthcare relies on DL for several critical applications [14]. For example, DL algorithms can analyze medical images (e.g., X-rays or MRIs) to detect anomalies (e.g., tumors or fractures) with high precision. Convolutional neural networks (CNNs) are particularly effective for image recognition tasks. In addition, DL models can predict patient outcomes by processing electronic health records (EHRs), which helps in proactive patient management through anticipating readmission likelihoods or disease progression. Furthermore, DL models can also process and comprehend human language that helps to extract important insights from clinical notes, research articles, and patient interactions. Diagnosis of condition, treatment recommendations, and patient data management all benefit from this feature.

In conclusion, AI mainly refers to the ability of machines to exhibit intelligent behavior. Under AI is ML, which is a technique of getting algorithms to learn from data. A more specific form of ML is DL that takes care of complex tasks using neural networks with different layers. These technologies can enhance diagnostic accuracy, develop personalized treatment plans, and improve patient outcomes through advanced data analysis and pattern recognition within the healthcare sector. It is essential to recognize these differences to better leverage AI capabilities in clinical medicine.

Types of ML

ML encompasses various techniques that enable computers to learn from data, each suited to different types of problems. Among these techniques are Fully-Supervised Learning, Unsupervised Learning, Weakly-Supervised Learning, Multiple Instance Learning (MIL), and Reinforcement Learning. Each type of ML algorithm offers distinct advantages and functionalities that groups can utilize for various assignments.

Fully-supervised learning involves training a model on a labeled dataset, where the outcome variable is already known. In this approach, each training sample consists of an input-output pair, with a feature vector as the input and the target label as the output. This method is commonly used for tasks such as fraud detection, image recognition, risk assessment, and predictive analytics. The effectiveness and success of fully-supervised learning largely depends on the quantity and quality of the labeled data, as well as the choice of the model and training algorithm [15].

This type of ML employs various techniques, such as the linear regression algorithm, which is used to predict the value of the dependent variable based on new, unseen data. It models the relationship between the input features and the target variable to estimate or forecast numerical values. Some common subtypes of regression algorithms include gradient boosting, random forest, and linear regression [16].

Another commonly used type in fully-supervised learning is the classification algorithm, which categorizes data into predefined classes or labels. These algorithms learn from labeled training data, where each example is associated with a specific class label. The goal of a classification algorithm is to find a decision boundary or rule that can accurately assign new, unseen instances to one of the classes, predicting the correct label for given input data. The model is fully trained on the training data and then evaluated on test data before it is used to make predictions on unseen data [17].

Unsupervised learning involves training models on data without output labels. The goal is to look for patterns or hidden structures in the data. Common methods include clustering, which puts related data points together, and dimensionality reduction, which lowers the amount of features while keeping important information. For tasks like feature extraction, anomaly detection, and customer segmentation, unsupervised learning is employed. Since it doesn't require labeled data, it is especially helpful when labeling data is difficult or expensive. Unsupervised ML models, in contrast to

supervised learning, learn from data without human supervision and independently identify patterns and insights. Cluster analysis is the most widely used unsupervised learning method, utilizing clustering algorithms to group data points according to how similar their values are [18].

Weakly-Supervised Learning deals with situations where the training data is not fully or accurately labeled. This can happen when only a small part of the data has labels, the labels are not very accurate, or they are not detailed enough. The main challenge in weakly-supervised learning is to use these limited or imperfect labels to create a reliable model. Common techniques include semi-supervised learning, which combines a small set of labeled data with a large amount of unlabeled data, and learning with noisy labels, where the model is designed to handle incorrect labels. This approach is useful in real-world situations where getting high-quality labeled data is difficult [19].

An additional technique for ML is MIL, it is a type of weakly supervised learning algorithm where training data is arranged in the form of bags where each bag contains multiple instances. In contrast to standard supervised learning, instead of providing labels at the instance level, they do it at the bag level. A bag is classified as positive in MIL if it contains at least one positive occurrence; otherwise, it is classified as negative. The model implicitly infers the labels of individual occurrences while learning to make predictions at the bag level. MIL is helpful in applications like picture classification, where the object of interest may only be present in a portion of the image, and drug activity prediction, where a medicine's effectiveness may depend on at least one active ingredient in a combination [20].

Reinforcement Learning, also called reinforcement learning from human feedback (RLHF), is a kind of dynamic programming that uses a reward-punishment mechanism to train algorithms. An agent acts in a particular environment to accomplish a predefined goal in order to implement reinforcement learning. Based on a predetermined criterion (usually points), the agent is rewarded or penalized for its activities, which incentivizes it to stick to good practices and abandon bad ones. The agent picks up the most effective techniques through practice. In summary, reinforcement learning enables an agent to learn in an interactive environment by trial and error using feedback from its own actions and experiences. It is most applicable in domains such as video game development and is frequently used to teach robots how to replicate human tasks [21].

In essence, ML includes a range of methods designed for certain data kinds and problem specifications. While unsupervised learning finds patterns in unlabeled data, fully-supervised learning uses labeled datasets to train models for precise predictions. Weakly-supervised learning makes use of limited information to create strong models while handling imperfect labeling. MIL focuses on scenarios where only high-level labels are available, allowing models to infer details from grouped data. Behavioral psychology serves as the inspiration for reinforcement learning, which teaches agents to make decisions through interactions with their surroundings and optimizes behavior based on rewards and penalties. When combined, these methods provide flexible answers for a broad variety of practical uses.

Neural networks

A class of ML models called neural networks is modeled after the composition and operations of the human brain. They are made up of linked layers of artificial neurons, each of which can carry out basic calculations. An input layer, one or more hidden layers, and an output layer are the standard components of a neural network. After processing information from the neurons in the layer above, each neuron in a layer transfers the outcome to the neurons in the layer below. Using techniques like backpropagation, the network learns by modifying the weights of these connections to reduce the difference between the expected and actual outputs. Since they can recognize intricate patterns in data, neural networks are extensively utilized in a variety of applications, including speech recognition, image identification, and natural language processing. Neural Networks are the foundation of many advanced ML techniques, including Convolution neural network (CNNs) for image-related tasks and Recurrent Neural Networks (RNNs) for sequential data [22].

CNN frequently outperforms other similar algorithms for image classification. Sample fea-

tures are extracted by the convolution layer and sub-sampling layer, and the feature of sharing weights reduces the training parameters of the network. CNNs have shown great promise in prostate cancer detection and diagnosis, according to recent research. These networks perform exceptionally well in automated image analysis because they can accurately identify malignant cells by processing digital histopathology images. CNNs can assist in the early diagnosis of anomalies in the prostate gland by analyzing MRI and ultrasound data. CNNs identify cancer types, grades, and extract complicated traits, which are vital information for customized therapy planning. Their impartial and consistent analysis improves diagnosis reliability by lowering human error [23].

RNN is designed to handle sequential input, such as text, audio, and time series. RNNs feature directed cycles formed by their connections, in contrast to standard neural networks that process input data independently. Because of their cyclical structure, RNNs are effectively endowed with a memory that enables them to store information from prior inputs. RNNs can identify patterns and relationships over time in a data sequence thanks to this memory.

When analyzing a sentence, for instance, an RNN can recall the words that came before and utilize this context to help comprehend and anticipate the words that will follow. RNNs are very effective for tasks involving sequential data because of this feature. However, because of problems like vanishing gradients, ordinary RNNs may have trouble with long-term dependencies. The model finds it challenging to learn long-range relationships in the data due to disappearing gradients, which happen during training when the gradients used to update the model's weights get extremely small.

Advanced RNNs, like Gated Recurrent Unit (GRU) networks and Long Short-Term Memory (LSTM) networks, have been created in order to overcome this restriction. By incorporating techniques to preserve and control information flow over extended sequences, these sophisticated RNNs improve their ability to learn and retain long-term dependencies [24]. A study published by Azizi S et al. [25] used RNNs for prostate cancer detection using analysis of Temporal Enhanced Ultrasound (TeUS) on a study of 255 prostate biopsy cores of 157 patients. The findings imply that temporal modeling of TeUS with RNN can greatly increase the accuracy of cancer detection compared to earlier studies.

RNNs can be very helpful in the detection and diagnosis of prostate cancer when processing sequential medical data, such as test time-series data and patient health records. RNNs can offer insights into the course of the disease by identifying temporal patterns and changes over time. This allows medical professionals to diagnose patients more precisely and customize treatment regimens for each of their patients. RNNs are an effective tool in medical research and treatment because of their capacity to manage sequential data; by making accurate and timely interventions, which may enhance patient outcomes [25].

Multimodal learning

In the ever-evolving landscape of precision medicine, multimodal learning is redefining diagnostics and patient care by seamlessly fusing the heterogeneous, yet complementary data streams - ranging from radiological images, histopathology images, genomic profile and molecular biomarkers to clinical information, and patient demographics - into a holistic, singular, and insightful narrative. This approach harnesses the unique strengths of each modality: high-resolution microscopy images reveal tissue morphology, while radiology images provide macro-level and spatial perspective of the same organ and genomic data offer insights into genetic mutations and expression profiles. By harmonizing these diverse datasets, multimodal learning aims to enhance the detection and classification of diseases, improve prognostic predictions, and identify potential therapeutic targets [26]. Through advanced DL and ML algorithms, multimodal learning unveils hidden patterns and subtle correlations that might be missed by human experts alone. This integration not only can enhance diagnostic precision but also facilitates personalized treatment plans, ultimately leading to more effective and targeted patient care.

Clinical applications in prostate cancer

Diagnosis, grading and quantification

The original Gleason grading system was established more than half a century ago [27]. It

grades prostate cancer based on histological patterns viewed at relatively low magnification (40 to 100×). It uses five growth patterns (Gleason patterns 1-5) to determine a score from 2-10 by summing primary and secondary patterns. It has been one of the most powerful predictors of prognosis in prostate cancer over the past 50 years [28]. Overtime, while pathologists have continued to use this system, they have also been addressing its shortcomings, such as the confusion caused by starting the scoring at 6 instead of 1 and the issue of mixing 3+4=7 and 4+3=7 - two categories having distinctly different prognosis [29, 30]. Based on the modified Gleason scores, the 2014 International Society of Urological Pathology (ISUP) consensus conference adopted a new five-tier grading system (Grade group 1 to Grade group 5), offering better stratification and prognosis prediction, with distinct 5-year biochemical recurrence-free progression rates [31]. This ISUP grade group system, now recommended for pathology reports [29, 32], has been further updated [33] and used in the WHO classification of prostate tumors [34, 35].

Traditionally, the diagnosis, grading and quantification of prostate cancer have been performed by pathologists examining glass slides under a microscope, visually identifying tumor-specific structures, and providing an overall score for the tissue sample, estimating tumor volume using the percentage of core involvement and the linear extent of involvement detected on glass slides. However, as the sample volume continues to increase because of MRI guided fusion biopsies and saturation biopsies, as well as an ongoing shortage of experienced pathologists, this traditional workflow has shown limitations. In fact, the demand for pathologists is quickly outstripping the supply, which is limited by the number of trainees and has remained steady over the past 2 decades [36]. Furthermore, the prognosis and risk stratification of prostate cancer are based on accurate grading and quantification, further standardization is needed. Here are some major concerns and limitations:

1. The traditional pathology workflow has been struggling to cope with the enormous workload. Prostate cancer can be multifocal and have a heterogeneous Gleason pattern distribution within the same patient [37]. There is considerable time spent by the pathologist to review each of the 12+ biopsy samples and 50-100 slides per case [10, 38, 39] and report an individual diagnosis Gleason score (grade group) for each part/container of prostate biopsy specimens, while most slides typically do not contain cancer, histopathological analysis could be streamlined significantly if these normal slides could automatically be excluded without missing any slides containing cancer.

2. There can be substantial interobserver variability in the Gleason grading of a biopsy specimen, particularly for pathologists with less experience interpreting prostate biopsies. In a study in which the interpretations from 29 pathologists were compared with that of an expert in prostate cancer pathology on an average of 278 samples, only 68 percent of samples were correctly classified as Gleason score <7, 7, or >7 [40]. Therefore, additional training may be necessary for those pathologists who are unfamiliar with the Gleason grading system. Manual estimation of tumor volume is subjective and also accounts for interobserver variability.

DL was introduced into the diagnosis of whole slide image (WSI) in 2016, when Litjens et al. initially trained a CNN using manually delineated and annotated digitized H&E-stained slides and yielded a receiver operator characteristic area under the curve (ROC-AUC) of 0.99 for distinguishing malignant from benign in an independent set of 270 biopsy slides [41]. In 2017, Kwak et al. [42] generated a nuclear seed map and trained CNN to detect cancer by recognizing nuclear architecture, which yielded an AUC of 0.974 in 491 tissue microarrays.

Soon after, the manual annotations at the pixel level by expert pathologists were found to be exhaustive and time-consuming, error prone, and not sufficient for generalized use in clinical-grade, real-world data. In 2019, Campanella et al. [43] used a MIL-based DL system, which used slide-level diagnosis instead of the extensive time-consuming pixel-wise manual annotations as training methods. The system detected malignancy in 24,859 pathology slides derived from 7159 prostate cancer patients, achieved an AUC of 0.994 for prostate cancer diagnosis, and it provided a potential of accelerating the clinical workflow by automatically excluding 65-75% of the slides that a pathologist usually

reviews, while keeping a sensitivity of 100%. This study demonstrated that the weakly supervised MIL system had a clear advantage over conventional fully supervised learning, as it enables training on massive, diverse datasets that cannot be annotated manually at the pixel level.

DL system was applied in Gleason grading by Nagpal et al. [44] in 2019 on radical prostatectomy specimens, demonstrating greater accuracy in Gleason, more precise quantitation of Gleason patterns, finer-grained discretization of the differentiation spectrum, and better risk stratification. Later, Ryu et al. [45] developed a deep neural networks system for automated Gleason scoring of core needle biopsy samples. They used 1133 cases of biopsy samples to train the system and validated it on 700 cases. The results showed a substantial diagnostic concordance between the system-grade group classification and reference standard.

Since 2020, several AI algorithms for diagnosis and Gleason grading have been developed and validated using more and more cases. Strom et al. [46] trained two CNNs ensembles on a total of 6953 digitized slides from needle core biopsies obtained from 1069 patients, Bulten et al. [47] trained a DL system with 5759 biopsies from 1243 patients to grade biopsies following the Gleason grading standard, and Negpal et al., built on previous research [44], refined a DL system showing a significantly higher rate of agreement with expert urologic subspecialists than general pathologists [48].

AI assistance improves pathologists' performance at Gleason grading of prostate biopsies. Pathologists assisted by the AI system not only improved compared with unassisted reads but also achieved higher median performance than the standalone AI. AI assistance decreased variance of performance and reduced observer variability between different pathologists, leading to more consistent Gleason grading scores. These results indicate that there is a potential benefit of pathologists using AI assistance as a supportive tool during diagnosis, especially in geographic regions where the number of pathologists is limited or subspecialized pathologists are not available. In these instances, AI systems can support pathologists in achieving higher grading accuracy and consistency [49].

The Paige Prostate AI-based digital diagnostic is one such tool that categorizes a prostate biopsy WSI as either "Suspicious" or "Not Suspicious" for prostatic adenocarcinoma [50]. To evaluate the performance of this program on prostate biopsies that were collected, processed, and independently diagnosed at an unrelated institution, it was used to review 1,876 prostate core biopsy WSI. The results were compared to the original pathology diagnosis made from the glass slides, showing a sensitivity of 97.7%, a positive predictive value of 97.9%, a specificity of 99.3%, and a negative predictive value of 99.2% in identifying core biopsies with cancer in this independent dataset [51]. Marketing of Paige Prostate was authorized by the US Food and Drug Administration (FDA) in 2021 as the first-to-market AI-based software designed to identify an area of interest on a scanned prostate biopsy image for it to be further reviewed and signed off by a pathologist. Since then, Paige Prostate has undergone further studies [52, 53] and numerous evaluations. One recent study showed that the Paige Prostate model demonstrated a high stand-alone diagnostic accuracy with a sensitivity of 97.4% and a specificity of 94.8% at the WSI level [54]. Moreover, Paige Prostate can assist pathologists by increasing their reading sensitivity and specificity, both on-site and remotely, bringing the performance of non-GU pathologists closer to that of GU specialists. Importantly, the accuracy gains attributed to Paige Prostate were observed across all histologic grades and tumor sizes, regardless of patient age, race, and ethnicity [54].

In 2020, Pantanowitz and colleagues from IBEX developed Galen Prostate [55], a commercial diagnosis and Gleason grading algorithm, using 1357480 labeled image patches extracted from manual annotations on 549 slides. The algorithm achieves an AUC of 0.991 for cancer detection, 0.941 for distinguishing between Gleason $\leq$6 and Gleason $\geq$7, and 0.957 for perineural invasion.

In 2021, Huang et al. [56] developed a deep convolutional neural network-based AI-powered platform to clearly distinguish prostate cancer epithelium from benign glands or epithelium and stroma, segment and label each cancer gland or epithelial patch with a corresponding Gleason pattern on the tissue, then sum

those values to generate the Gleason pattern volumes, Gleason score, and cancer volume with high accuracy at the patch-pixel level (AUC=0.92). The pixel method was used to measure prostate cancer volume as prostate cancer pixels relative to total tissue pixels, and the percentages of Gleason patterns as pixels of each Gleason pattern relative to total prostate cancer pixels. The results demonstrated that with AI assistance, pathologists achieved significantly higher concordance in grading and quantification compared to traditional manual methods, while reducing the time needed and improving efficiency in generating a diagnosis.

Bulten et al. published in 2022 the result of the Prostate cANcer graDe Assessment (PANDA) challenge [57], the largest AI-based histopathology challenge during which international teams from 1290 developers competed to develop AI algorithms for Gleason grading using 10,616 digitized prostate biopsies. This competition proved that multiple AI models could effectively generalize across diverse patient populations from different centers for Gleason grading [58]. The representative algorithm had higher sensitivity for tumor (98.2%, 95% CI of all algorithms 97.4-100.0) than the representative pathologist (96.5%, 95% CI of all pathologists 95.4-100.0) and higher specificity (100.0%, 95% CI of all algorithms 90.6-100.0, versus 92.3%, 95% CI of all pathologists 77.8-97.8). On average, the algorithms missed 1.0% of cancers, whereas the pathologists missed 1.8% [57].

Risk evaluation and prognosis

With AI now capable of diagnosis and grading, it is demonstrating an even more significant role and that is advancing further to predict long-term outcomes. This refinement goes beyond the initial goal of using AI to assist pathologists in managing workflow challenges. From here, pathology, supported by AI, can truly realize its critical potential in bridging and integrating various aspects of medicine [59].

Before AI, in 1998, D'Amico et al., based on the clinical TNM stage, PSA level, biopsy Gleason score, developed a combined modality staging system stratifying patients into groups with a low, intermediate, or high-risk of biochemical recurrence after radical prostatectomy or radiotherapy [60]. This model, while broadly accepted and used in both practice and clinical trial designs, showed a major shift in patients' distribution among risk groups over time [61] and its prognostic accuracy remained to be improved [10]. Based on the D'Amico model, National Comprehensive Cancer Network (NCCN) risk categories [62] have been used in guidelines and is the most commonly used risk-stratification tool. However, NCCN may lead to over- and under treatment due to a spectrum of outcomes still existing within each of the current 6 categories (very low risk, low risk, favorable intermediate, unfavorable intermediate, high and very high risk) [63]. Another widely used surrogate risk rating model, the Cancer of the Prostate Risk Assessment (CAPRA) score [64-66], is also based on risk factors quantification, similar to the D'Amico model.

In 2021, Wulczyn and colleagues from Google Health [67] developed a Gleason grading AI system by introducing a retrospective cohort of 2807 prostatectomy patients with 5-25 years to follow-up, to predict prostate cancer-specific mortality and evaluate its risk-stratification. This system produced continuous AI risk scores with a concordance index (C-index) significantly greater than that of the Grade Group from the original pathology report, as well as the pathologist's Grade Group, demonstrating better risk stratification by the AI risk score system [67].

Biomedical recurrence is traditionally defined as PSA rise again after treatment, indicating the regrowth of prostate cancer cells. The risk of biochemical recurrence of prostate cancer had been assessed in clinical practice through a combination of the Gleason grade groups, the PSA value at diagnosis, and the TNM staging criteria. In 2018, Ren et al. used deep neural networks to identify a set of computational biomarkers from WSI and genomic data of histopathology specimens. The computational biomarkers showed a recurrence hazard ratio of 5.73 among Gleason score 7 patients [68]. In 2022, Pinckaers et al. developed a DL based prognostic biomarker by training the DL system with H&E-stained histopathological tissue microarrays [69]. The prognostic biomarker provides a strong correlation (odds ratio =3.32) with early biochemical recurrence, suggesting that there is more morphological information in the tissue besides the Gleason grade groups.

In 2022, Esteva et al. demonstrated a multimodal DL AI (MMAI) system and trained six distinct models on a dataset of 16,204 histopathology slides and clinical data from 5,654 patients with a median follow-up of 11.4 years [70]. The models were used to predict long-term, clinically relevant outcomes including 5- and 10-year distant metastasis, 5- and 10-year biochemical failure, 10-year prostate cancer-specific survival, and 10-year overall survival. The AUC of sensitivity and specificity of these models were measured compared to the NCCN risk group. The MMAI model consistently outperformed the NCCN risk groups across all tested outcomes, with a substantial improvement in AUC varying from 9.2% to 14.6% [70]. Since accurate prediction of distant metastasis at 5 and 10 years is particularly important for identifying patients who may have more aggressive disease and require additional treatment, the MMAI system shows its promising potential for personalized prostate cancer therapy by predicting long-term, clinically relevant outcomes.

In 2024, Fernandes-Mateos et al. employed DL techniques to evaluate morphological heterogeneity [71]. This morphological heterogeneity, quantified by the Gleason Morisita index, was not just a surrogate of some aggressive subpathology, but reflected the dispersion and intermixing of different Gleason patterns within the tissue. They also sequenced DNAs in the matched tissues to assess genomic intratumor heterogeneity. The assessment of morphological and genomic heterogeneity was incorporated into clinical trials with a median follow-up of 12 years. They demonstrated that both genomic and morphological heterogeneity were independent and substantial predictors of recurrence (hazard ratio =3.12 and 2.45, respectively). With the help of DL to calculate "continuous Gleason", they showed that morphological heterogeneity and genomic heterogeneity were partly related. Furthermore, the combined "joint diversity" yielded a hazard ratio of 2.76, helping to identify the most genetically and morphologically diverse individuals with a much poorer prognosis [71].

Treatment options and responses

Current treatments of prostate cancer include active surveillance, radiation therapy (external beam source and/or brachytherapy), hormone treatment (androgen deprivation therapy), surgery (radical prostatectomy), immunotherapy, chemotherapy, and ablative techniques (cryotherapy, high-intensity focused ultrasound, photodynamic therapy with an interstitial laser, irreversible electroporation and thermal water vapor ablation). Based on clinical stage and risk stratification, different treatment strategies are selected either individually or in combination. Before the application of AI, the NCCN guidelines had been used as guidelines to categorize risk and select treatments. However, with the advent of AI in risk prediction, it has started to play a role in treatment selection as well.

The first attempt to use AI to predict treatment plans and outcomes began in 2022, when Nakata et al. [72] studied a relatively small group of patients with metastatic prostate cancer and predicted the time to castration-resistant prostate cancer (CRPC) by combined androgen blockade therapy. Digitized H&E slides from eligible patients were randomly cropped into 224*224 pixel-size jpeg images, creating 7,440 images from 18 CRPC>30 months patients and 5,210 images from 6 CRPC<6 months patients. These images were used to train a DL algorithm, with a CNN applied to predict the time to CRPC. Then, using 3,399 images from 8 patients with CRPC>24 months and 3,727 images from 8 patients with CRPC<6 months for validation, the prediction accuracy proved to be significant.

In 2023, Tsuneki et al. [73] developed a DL model to classify prostate cancer into indolent (applicable for active surveillance) and aggressive (necessary for definitive therapy). The model was trained using a combination of transfer, weakly supervised, and fully supervised learning approaches to 1,300 core needle biopsy WSIs, achieving AUC of 0.846 for indolent and 0.98 for aggressive. This study showed how AI can assist in identifying suitable patients for active surveillance.

Also in 2023, Spratt et al. [74], based on the MMAI system developed by Esteva et al. [70], built a predictive model to determine whether patients would benefit from short-term androgen deprivation therapy (ADT) over a 15-year outlook for distant metastasis and prostate cancer-specific mortality. The model was vali-

dated in a clinical trial which randomly assigned patients to radiotherapy plus or minus 4 months of ADT. The validation results showed that patients whose predictive models were positive for 15-year distant metastasis and 15-year prostate cancer-specific mortality were more likely to benefit from radiation plus short-term ADT regimen, whereas patients whose predictive models were negative for 15-year distant metastasis and 15-year prostate cancer-specific mortality were less likely to benefit from ADT [74]. This study showed that the MMAI predictive model was able to identify candidates from a broad range of patients from over 500 centers for personalized treatment options.

Practical uses of AI in routine pathology services

Although AI has demonstrated great potential in prostate pathology, the practical use of AI in routine pathology diagnosis has not been widely accepted [75]. There are many hurdles in adoption of AI based pathology systems [76]: 1) lack of necessary digital infrastructure to support AI based systems; 2) upfront high cost of digital equipment and AI software, especially in low-resources settings; 3) shortage of trained professionals who are skilled in both digital pathology and AI; 4) lack of robust IT and AI support and specialists which can manage and troubleshoot these advanced systems; 5) concerns of regulatory standards and ethics; 6) accessibility of clinically suitable Image Management System (IMS) and AI systems for routine clinical use; 7) lack of first-hand experience of AI-based pathology and misperception of AI technology by conventional pathologists, etc.

In this section, we present some potentials of AI-based pathology in routine pathology practice, focusing on the improvement of workflow of laboratories, enhancement of efficiency of pathologists, and providing additional information to clinicians based on histopathology.

When glass slides are digitized using a scanner, some defects in images may happen which are not encountered in transitional glass slides. Blurry images are commonly seen in digital images, especially when using high throughput WSI scanners. The rate of blurry images varies depending on different scanners. Campanella et al. [77] reported a blurry rate up to 17% and developed a blur detection method using CNNs. The accuracy of feature-based and deep-learning based approaches for sharpness classification was (99.74% accuracy) and regression (MSE 0.004). The author demonstrates superior performance over the state-of-the-art QC pipeline comprising commercial software and human expert inspection by reducing the error rate from 17% to 4.7%. Our own experience of blurry images in prostate core biopsies can be as high as 10%, which adds additional burden to laboratory staff for retrieval of the slides to rescan and causes delay of pathologist review and release of reports. In order to identify those blurry images, our AI model is able to detect and flag the blurry images in our IMS. This allows staff to quickly identify blurry images and assess the level of blurriness and rescan immediately, if necessary, before filing the slides. When pathologists are ready to review the cases, blurry images have already been corrected, therefore improving the efficiency of review and release reports by pathologists.

Another scanning defect is missing small tissue which may carry significant consequences. Many of the scanners are set to minimize scanning area to reduce unnecessary increase of image size. As a result, some small fragments away from main tissue may not be in scanning areas. Atallah et al. [78] reviewed 40,160 breast WSIs were examined and compared with their corresponding glass slides. The frequency of missing tissue ranged from 2% to 19%. The area size of the missed tissue ranged from 1-70%. In most cases (>75%), the missing tissue area size was <10% and peripherally located. AI can be very useful in identifying missing small fragments on slide based WSIs.

AI assisted digital pathology can improve workflow of prostate biopsy review and sign-out. After scanning of the slides, the images can be organized in ordinary slide tray format and assigned to a pathologist for review by preestablished laboratory criteria [76]. In our system, it is designed that AI prescreen all prostate biopsies and identify biopsies with cancerous tissue. The cancerous areas are highlighted and the cases with cancer are flagged so that pathologists can review positive cases first and get the reports out sooner, resulting in significant improvement of turn-around-time.

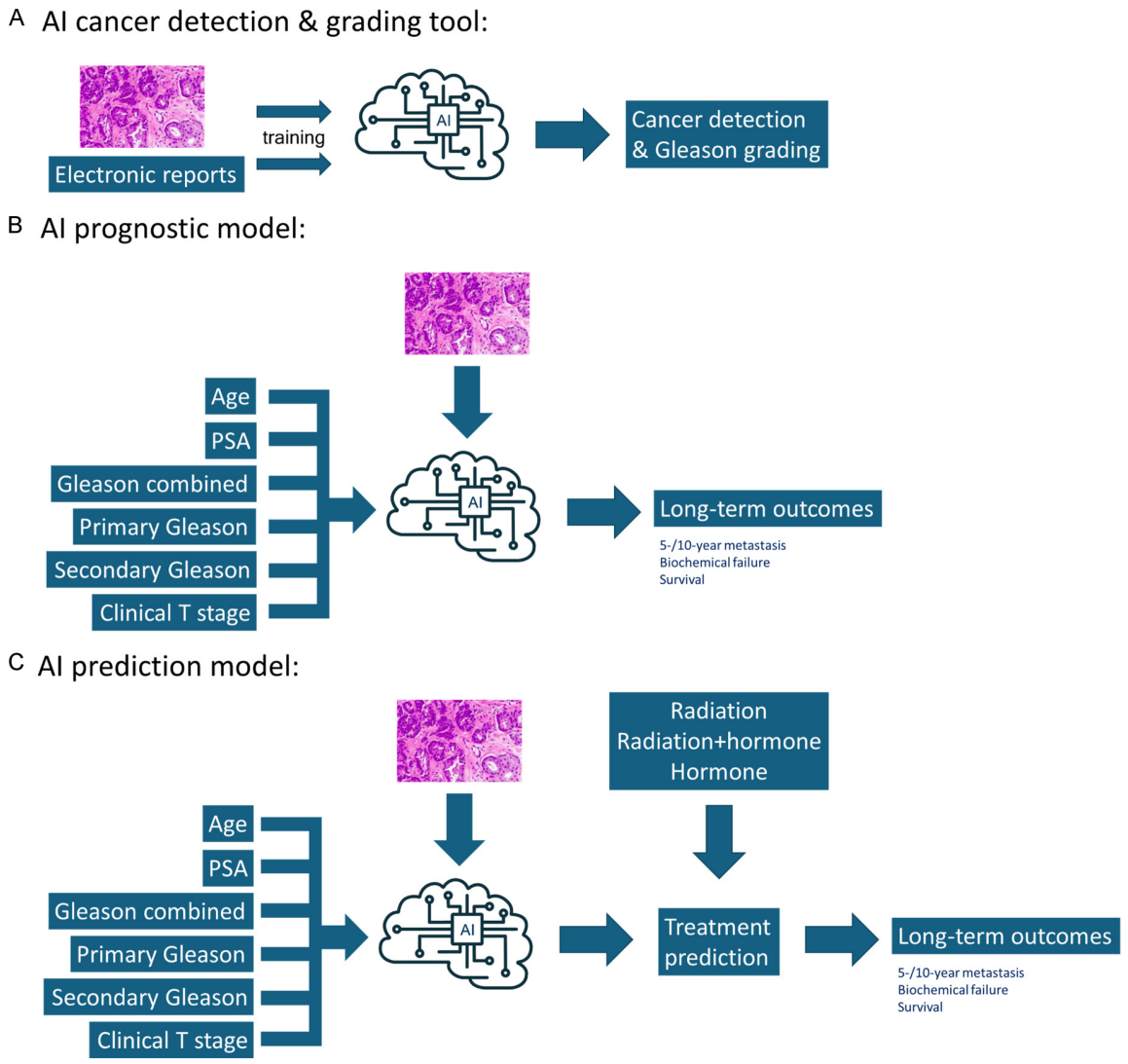

Figure 1. Significant breakthroughs and advancement so far. A. AI application as prostate cancer detection and grading tool. B. The application of AI prognostic model in the long-term outcomes of prostate cancer. C. The application of AI prediction model in treatment prediction based on long-term outcomes.

Furthermore, our AI is also trained to identify small focus of tumor or focus of atypical glands where PIN4 immunostain may be required. Those cases require PIN4 immunostains will be flagged and reviewed by a pathologist to determine if PIN4 is indeed necessary. This process further improved turn-around-time and pathologist's sign-out efficiency. In fact, our experience demonstrated a significant reduction of PIN4 immunostain requests since AI is able to add additional confidence for pathologists to render a correct diagnosis.

One of the time-consuming tasks in reviewing and generating pathology reports of prostate biopsy is accurate classification of Gleason scores, tissue and tumor length measurements, and adding additional information such as percentage of Gleason 4 pattern and perineural invasion. With the assistance of AI, creation of pathology reports becomes easier, faster and more accurate since all these parameters can be provided by AI models. Pathologists only need to transfer the information to reports or confirm the information that has already been pre-populated in the report. As studies showed, AI is more accurate in determining Gleason patterns and tumor volume [55, 58]. Finally, the correct information is critical in patients' management since clinical stag-

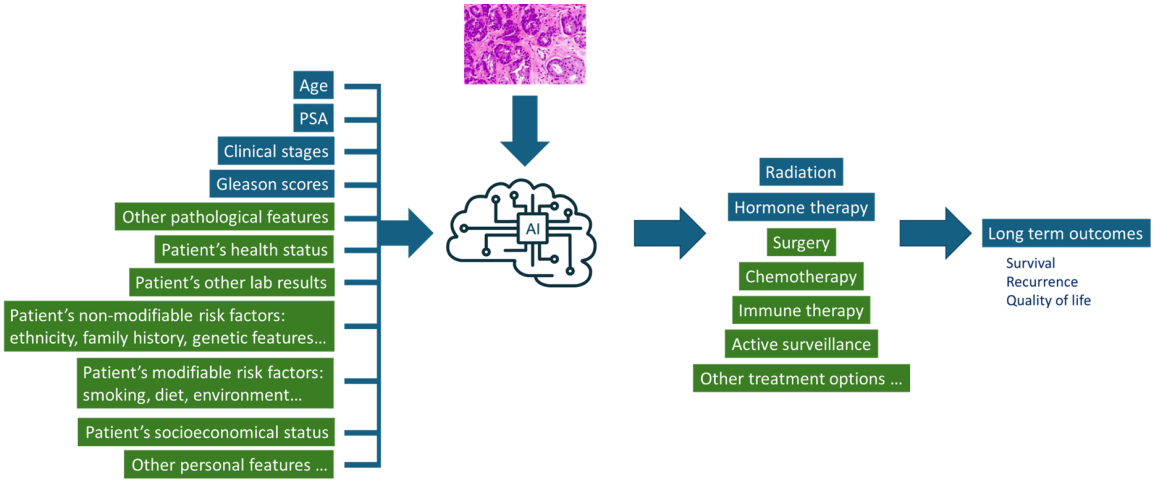

Figure 2. Future perspectives of how AI assists pathology in playing a critical role in various parts of medicine. Blue color indicates the patients' status and treatment options, which have already been involved in research. Green color indicates that there are more personalized patients' status and more abundant treatment options, which can be assessed with the application of AI assisted pathology.

ing and risk stratification (NCCN) are based on pathology information.

Summary and future perspectives

Through this review, we can observe the development of AI over the past decade. There have been at least four significant breakthroughs and advancements so far. The first breakthrough was the introduction of AI in pathology, where learning was fully supervised and relied on manually annotated slides. The second breakthrough occurred when the DL systems were trained to recognize WSI independently of human annotation in a weakly supervised mode to detect the presence of cancer and even grading. This AI detection and grading tool has become so advanced that one has received FDA approval. The third breakthrough was that AI was no longer limited to detection and grading but also focused on the long-term outcome of patients. This is how AI prognostic models work, using clinical features and all the image information from digitized H&E stained slides to provide detailed prognostic evaluations for patients. The fourth breakthrough most recently developed involves incorporating treatment options into the prognostic model, creating a prediction model for better treatment selection based on long-term outcomes (**Figure 1**).

The exciting advancements motivate us to forge ahead. The current prediction models indicate a promising direction where AI can be trained with a variety of treatment options, enabling personalized predictions for treatment outcomes. With DL algorithms, diverse inputs and outputs can be incorporated into the training system. In addition to clinical features outlined by current guidelines, one can now integrate patients' non-modifiable and modifiable risk factors, their overall health status, past medical history, disease progression, and even their socioeconomic status - areas often underrepresented in clinical practice. Moreover, with the support of AI's precise image interpretation and powerful comprehensive analysis, we can study types of prostate cancer that are difficult to grade using Gleason system, such as ductal cell carcinoma and treatment-emergent small cell neuroendocrine prostate cancer, allowing us to provide treatment options for these rare histological types, which currently have a poor prognosis. Moving towards more precise and personalized patient care, our ultimate goal is to achieve better long-term outcomes and improve patients' quality of life (**Figure 2**).

Disclosure of conflict of interest

None.

Address correspondence to: Dr. David Y Zhang, Veterans Affairs New York Harbor Healthcare System, New York, NY 10010, USA. Tel: 718-757-8534; E-mail: Dyzhang01@gmail.com

References

[1] Rudd RM. Malignant mesothelioma. Br Med Bull 2010; 93: 105-123.

[2] Siegel RL, Giaquinto AN and Jemal A. Cancer statistics, 2024. CA Cancer J Clin 2024; 74: 12-49.

[3] Bergengren O, Pekala KR, Matsoukas K, Fainberg J, Mungovan SF, Bratt O, Bray F, Brawley O, Luckenbaugh AN, Mucci L, Morgan TM and Carlsson SV. 2022 update on prostate cancer epidemiology and risk factors-a systematic review. Eur Urol 2023; 84: 191-206.

[4] Witte JS. Prostate cancer genomics: towards a new understanding. Nat Rev Genet 2009; 10: 77-82.

[5] Wong EY, Chu TN and Ladi-Seyedian SS. Genomics and artificial intelligence: prostate cancer. Urol Clin North Am 2024; 51: 27-33.

[6] Gandaglia G, Leni R, Bray F, Fleshner N, Freedland SJ, Kibel A, Stattin P, Van Poppel H and La Vecchia C. Epidemiology and prevention of prostate cancer. Eur Urol Oncol 2021; 4: 877-892.

[7] Humphrey PA. Histological variants of prostatic carcinoma and their significance. Histopathology 2012; 60: 59-74.

[8] Yamada Y and Beltran H. Clinical and biological features of neuroendocrine prostate cancer. Curr Oncol Rep 2021; 23: 15.

[9] de Kouchkovsky I, Chan E, Schloss C, Poehlein C and Aggarwal R. Diagnosis and management of neuroendocrine prostate cancer. Prostate 2024; 84: 426-440.

[10] Baydoun A, Jia AY, Zaorsky NG, Kashani R, Rao S, Shoag JE, Vince RA Jr, Bittencourt LK, Zuhour R, Price AT, Arsenault TH and Spratt DE. Artificial intelligence applications in prostate cancer. Prostate Cancer Prostatic Dis 2024; 27: 37-45.

[11] Shane J. You look like a thing and I love you: how artificial intelligence works and why it's making the world a weirder place. New York: Voracious/Little, Brown and Company; 2019.

[12] Fjelland R. Why general artificial intelligence will not be realized. Humanit Soc Sci Commun 2020; 7: 1-9.

[13] Mahesh B. Machine learning algorithms-a review. International Journal of Science and Research 2020; 9: 381-386.

[14] Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, Santamaría J, Fadhel MA, Al-Amidie M and Farhan L. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data 2021; 8: 53.

[15] Frank SM, Qi A, Ravasio D, Sasaki Y, Rosen EL and Watanabe T. Supervised learning occurs in visual perceptual learning of complex natural images. Curr Biol 2020; 30: 2995-3000, e2993.

[16] Huang S, Ailer E, Kilbertus N and Pfister N. Supervised learning and model analysis with compositional data. PLoS Comput Biol 2023; 19: e1011240.

[17] Chartrand G, Cheng PM, Vorontsov E, Drozdzal M, Turcotte S, Pal CJ, Kadoury S and Tang A. Deep learning: a primer for radiologists. Radiographics 2017; 37: 2113-2131.

[18] Ju W, Gu Y, Luo X, Wang Y, Yuan H, Zhong H and Zhang M. Unsupervised graph-level representation learning with hierarchical contrasts. Neural Netw 2023; 158: 359-368.

[19] Hassan H, Ren Z, Zhou C, Khan MA, Pan Y, Zhao J and Huang B. Supervised and weakly supervised deep learning models for COVID-19 CT diagnosis: a systematic review. Comput Methods Programs Biomed 2022; 218: 106731.

[20] Nguyen T and Raich R. Incomplete label multiple instance multiple label learning. IEEE Trans Pattern Anal Mach Intell 2022; 44: 1320-1337.

[21] Chen R and Goldberg JH. Actor-critic reinforcement learning in the songbird. Curr Opin Neurobiol 2020; 65: 1-9.

[22] Doya K, Friston K, Sugiyama M and Tenenbaum J. Neural networks special issue on artificial intelligence and brain science. Neural Netw 2022; 155: 328-329.

[23] Chui KT, Gupta BB, Chi HR, Arya V, Alhalabi W, Ruiz MT and Shen CW. Transfer learning-based multi-scale denoising convolutional neural network for prostate cancer detection. Cancers (Basel) 2022; 14: 3687.

[24] Kriegeskorte N and Golan T. Neural network models and deep learning. Curr Biol 2019; 29: R231-R236.

[25] Jose A, Roy R, Moreno-Andres D and Stegmaier J. Automatic detection of cell-cycle stages using recurrent neural networks. PLoS One 2024; 19: e0297356.

[26] Boehm KM, Khosravi P, Vanguri R, Gao J and Shah SP. Harnessing multimodal data integration to advance precision oncology. Nat Rev Cancer 2022; 22: 114-126.

[27] Gleason DF and Mellinger GT. Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging. J Urol 1974; 111: 58-64.

[28] Bostwick DG. Gleason grading of prostatic needle biopsies. Correlation with grade in 316 matched prostatectomies. Am J Surg Pathol 1994; 18: 796-803.

[29] Yang XJ. Interpretation of prostate biopsy. Wolters Kluwer; 2022.

[30] Epstein JI, Egevad L, Amin MB, Delahunt B, Srigley JR and Humphrey PA; Grading Commit-

tee. 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: 244-252.

[31] Epstein JI, Zelefsky MJ, Sjoberg DD, Nelson JB, Egevad L, Magi-Galluzzi C, Vickers AJ, Parwani AV, Reuter VE, Fine SW, Eastham JA, Wiklund P, Han M, Reddy CA, Ciezki JP, Nyberg T and Klein EA. A contemporary prostate cancer grading system: a validated alternative to the gleason score. Eur Urol 2016; 69: 428-435.

[32] Epstein JI, Amin MB, Fine SW, Algaba F, Aron M, Baydar DE, Beltran AL, Brimo F, Cheville JC, Colecchia M, Comperat E, da Cunha IW, Delprado W, DeMarzo AM, Giannico GA, Gordetsky JB, Guo CC, Hansel DE, Hirsch MS, Huang J, Humphrey PA, Jimenez RE, Khani F, Kong Q, Kryvenko ON, Kunju LP, Lal P, Latour M, Lotan T, Maclean F, Magi-Galluzzi C, Mehra R, Menon S, Miyamoto H, Montironi R, Netto GJ, Nguyen JK, Osunkoya AO, Parwani A, Robinson BD, Rubin MA, Shah RB, So JS, Takahashi H, Tavora F, Tretiakova MS, True L, Wobker SE, Yang XJ, Zhou M, Zynger DL and Trpkov K. The 2019 Genitourinary Pathology Society (GUPS) white paper on contemporary grading of prostate cancer. Arch Pathol Lab Med 2021; 145: 461-493.

[33] van Leenders GJLH, van der Kwast TH, Grignon DJ, Evans AJ, Kristiansen G, Kweldam CF, Litjens G, McKenney JK, Melamed J, Mottet N, Paner GP, Samaratunga H, Schoots IG, Simko JP, Tsuzuki T, Varma M, Warren AY, Wheeler TM, Williamson SR and Iczkowski KA; ISUP Grading Workshop Panel Members. The 2019 International Society of Urological Pathology (ISUP) consensus conference on grading of prostatic carcinoma. Am J Surg Pathol 2020; 44: e87-e99.

[34] Kench JG, Amin MB, Berney DM, Compérat EM, Cree IA, Gill AJ, Hartmann A, Menon S, Moch H, Netto GJ, Raspollini MR, Rubin MA, Tan PH, Tsuzuki T, Turjalic S, Van Der Kwast TH, Zhou M and Srigley JR. WHO classification of tumours fifth edition: evolving issues in the classification, diagnosis, and prognostication of prostate cancer. Histopathology 2022; 81: 447-458.

[35] Surintrspanont J and Zhou M. Prostate pathology: what is new in the 2022 WHO classification of urinary and male genital tumors? Pathologica 2022; 115: 41-56.

[36] Gross DJ, Robboy SJ, Cohen MB, Vernon L, Park JY, Crawford JM, Karcher DS, Wheeler TM and Black-Schaffer WS. Strong job market for pathologists: results from the 2021 college of American pathologists practice leader survey. Arch Pathol Lab Med 2023; 147: 434-441.

[37] Aihara M, Wheeler TM, Ohori M and Scardino PT. Heterogeneity of prostate cancer in radical prostatectomy specimens. Urology 1994; 43: 60-66; discussion 66-67.

[38] Bjurlin MA, Carter HB, Schellhammer P, Cookson MS, Gomella LG, Troyer D, Wheeler TM, Schlossberg S, Penson DF and Taneja SS. Optimization of initial prostate biopsy in clinical practice: sampling, labeling and specimen processing. J Urol 2013; 189: 2039-2046.

[39] Tolkach Y, Ovtcharov V, Pryalukhin A, Eich ML, Gaisa NT, Braun M, Radzhabov A, Quaas A, Hammerer P, Dellmann A, Hulla W, Haffner MC, Reis H, Fahoum I, Samarska I, Borbat A, Pham H, Heidenreich A, Klein S, Netto G, Caie P and Buettner R. An international multi-institutional validation study of the algorithm for prostate cancer detection and Gleason grading. NPJ Precis Oncol 2023; 7: 77.

[40] Burchardt M, Engers R, Muller M, Burchardt T, Willers R, Epstein JI, Ackermann R, Gabbert HE, de la Taille A and Rubin MA. Interobserver reproducibility of Gleason grading: evaluation using prostate cancer tissue microarrays. J Cancer Res Clin Oncol 2008; 134: 1071-1078.

[41] Litjens G, Sanchez CI, Timofeeva N, Hermsen M, Nagtegaal I, Kovacs I, Hulsbergen-van de Kaa C, Bult P, van Ginneken B and van der Laak J. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. Sci Rep 2016; 6: 26286.

[42] Kwak JT and Hewitt SM. Nuclear architecture analysis of prostate cancer via convolutional neural networks. IEEE Access 2017; 5: 18526-18533.

[43] Campanella G, Hanna MG, Geneslaw L, Miraflor 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.

[44] 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. Development and validation of a deep learning algorithm for improving Gleason scoring of prostate cancer. NPJ Digit Med 2019; 2: 48.

[45] Ryu HS, Jin MS, Park JH, Lee S, Cho J, Oh S, Kwak TY, Woo JI, Mun Y, Kim SW, Hwang S, Shin SJ and Chang H. Automated Gleason scoring and tumor quantification in prostate core needle biopsy images using deep neural networks and its comparison with pathologist-based assessment. Cancers (Basel) 2019; 11: 1860.

[46] Strom 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, Samaratunga H, Srigley JR, Takahashi H, Tsuzuki T, Varma M, Zhou M, Lindberg J, Lindskog C, Ruusuvuori P, Wahlby C, Gronberg 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.

[47] 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.

[48] Nagpal K, Foote D, Tan F, Liu Y, Chen PC, Steiner DF, Manoj N, Olson N, Smith JL, Mohtashamian 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, Gadepalli 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.

[49] Bulten W, Balkenhol M, Belinga JA, Brilhante A, Cakır A, Egevad L, Eklund M, Farre X, Geronatsiou K, Molinie V, Pereira G, Roy P, Saile G, Salles P, Schaafsma E, Tschui J and 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.

[50] Raciti P, Sue J, Ceballos R, Godrich R, Kunz JD, Kapur S, Reuter V, Grady L, Kanan C, Klimstra DS and Fuchs TJ. Novel artificial intelligence system increases the detection of prostate cancer in whole slide images of core needle biopsies. Mod Pathol 2020; 33: 2058-2066.

[51] Perincheri S, Levi AW, Celli R, Gershkovich P, Rimm D, Morrow JS, Rothrock B, Raciti P, Klimstra D and Sinard J. An independent assessment of an artificial intelligence system for prostate cancer detection shows strong diagnostic accuracy. Mod Pathol 2021; 34: 1588-1595.

[52] da Silva LM, Pereira EM, Salles PG, Godrich R, Ceballos R, Kunz JD, Casson A, Viret J, Chandarlapaty S, Ferreira CG, Ferrari B, Rothrock B, Raciti P, Reuter V, Dogdas B, DeMuth G, Sue J, Kanan C, Grady L, Fuchs TJ and Reis-Filho JS. Independent real-world application of a clinical-grade automated prostate cancer detection system. J Pathol 2021; 254: 147-158.

[53] Eloy C, Marques A, Pinto J, Pinheiro J, Campelos S, Curado M, Vale J and Polonia A. Artificial intelligence-assisted cancer diagnosis improves the efficiency of pathologists in prostatic biopsies. Virchows Arch 2023; 482: 595-604.

[54] Raciti P, Sue J, Retamero JA, Ceballos R, Godrich R, Kunz JD, Casson A, Thiagarajan D, Ebrahimzadeh Z, Viret J, Lee D, Schuffler PJ, DeMuth G, Gulturk E, Kanan C, Rothrock B, Reis-Filho J, Klimstra DS, Reuter V and Fuchs TJ. Clinical validation of artificial intelligence-augmented pathology diagnosis demonstrates significant gains in diagnostic accuracy in prostate cancer detection. Arch Pathol Lab Med 2023; 147: 1178-1185.

[55] Pantanowitz L, Quiroga-Garza GM, Bien L, Heled R, Laifenfeld D, Linhart C, Sandbank J, Albrecht Shach A, Shalev V, Vecsler M, Michelow P, Hazelhurst S and Dhir R. 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.

[56] 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.

[57] Bulten W, Kartasalo K, Chen PC, Strom 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, Gronberg H, Samaratunga H, Delahunt B, Tsuzuki T, Hakkinen T, Egevad L, Demkin M, Dane S, Tan F, Valkonen M, Corrado GS, Peng L, Mermel CH, Ruusuvuori 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.

[58] Azadi Moghadam P, Bashashati A and Goldenberg SL. Artificial intelligence and pathomics: prostate cancer. Urol Clin North Am 2024; 51: 15-26.

[59] Shmatko A, Ghaffari Laleh N, Gerstung M and Kather JN. Artificial intelligence in histopathology: enhancing cancer research and clinical oncology. Nat Cancer 2022; 3: 1026-1038.

[60] D'Amico AV, Whittington R, Malkowicz SB, Schultz D, Blank K, Broderick GA, Tomaszewski JE, Renshaw AA, Kaplan I, Beard CJ and Wein A. Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. JAMA 1998; 280: 969-974.

[61] Hernandez DJ, Nielsen ME, Han M and Partin AW. Contemporary evaluation of the D'amico risk classification of prostate cancer. Urology 2007; 70: 931-935.
[62] Eastham JA, Auffenberg GB, Barocas DA, Chou R, Crispino T, Davis JW, Eggener S, Horwitz EM, Kane CJ, Kirkby E, Lin DW, McBride SM, Morgans AK, Pierorazio PM, Rodrigues G, Wong WW and Boorjian SA. Clinically localized prostate cancer: AUA/ASTRO guideline, part I: introduction, risk assessment, staging, and risk-based management. J Urol 2022; 208: 10-18.
[63] Moses KA, Sprenkle PC, Bahler C, Box G, Carlsson SV, Catalona WJ, Dahl DM, Dall'Era M, Davis JW, Drake BF, Epstein JI, Etzioni RB, Farrington TA, Garraway IP, Jarrard D, Kauffman E, Kaye D, Kibel AS, LaGrange CA, Maroni P, Ponsky L, Reys B, Salami SS, Sanchez A, Seibert TM, Shaneyfelt TM, Smaldone MC, Sonn G, Tyson MD, Vapiwala N, Wake R, Washington S, Yu A, Yuh B, Berardi RA and Freedman-Cass DA. NCCN guidelines(R) insights: prostate cancer early detection, version 1.2023. J Natl Compr Canc Netw 2023; 21: 236-246.
[64] Cooperberg MR, Broering JM and Carroll PR. Risk assessment for prostate cancer metastasis and mortality at the time of diagnosis. J Natl Cancer Inst 2009; 101: 878-887.
[65] Cooperberg MR, Hilton JF and Carroll PR. The CAPRA-S score: a straightforward tool for improved prediction of outcomes after radical prostatectomy. Cancer 2011; 117: 5039-5046.
[66] Brajtbord JS, Leapman MS and Cooperberg MR. The CAPRA score at 10 years: contemporary perspectives and analysis of supporting studies. Eur Urol 2017; 71: 705-709.
[67] Wulczyn E, Nagpal K, Symonds M, Moran M, Plass M, Reihs R, Nader F, Tan F, Cai Y, Brown T, Flament-Auvigne I, Amin MB, Stumpe MC, Muller H, Regitnig P, Holzinger A, Corrado GS, Peng LH, Chen PC, Steiner DF, Zatloukal K, Liu Y and Mermel CH. Predicting prostate cancer specific-mortality with artificial intelligence-based Gleason grading. Commun Med (Lond) 2021; 1: 10.
[68] Ren J, Karagoz K, Gatza ML, Singer EA, Sadimin E, Foran DJ and Qi X. Recurrence analysis on prostate cancer patients with Gleason score 7 using integrated histopathology whole-slide images and genomic data through deep neural networks. J Med Imaging (Bellingham) 2018; 5: 047501.
[69] Pinckaers H, van Ipenburg J, Melamed J, De Marzo A, Platz EA, van Ginneken B, van der Laak J and Litjens G. Predicting biochemical recurrence of prostate cancer with artificial intelligence. Commun Med (Lond) 2022; 2: 64.
[70] Esteva A, Feng J, van der Wal D, Huang SC, Simko JP, DeVries S, Chen E, Schaeffer EM, Morgan TM, Sun Y, Ghorbani A, Naik N, Nathawani D, Socher R, Michalski JM, Roach M 3rd, Pisansky TM, Monson JM, Naz F, Wallace J, Ferguson MJ, Bahary JP, Zou J, Lungren M, Yeung S and Ross AE; NRG Prostate Cancer AI Consortium; Sandler HM, Tran PT, Spratt DE, Pugh S, Feng FY and Mohamad O. Prostate cancer therapy personalization via multi-modal deep learning on randomized phase III clinical trials. NPJ Digit Med 2022; 5: 71.
[71] Fernandez-Mateos J, Cresswell GD, Trahearn N, Webb K, Sakr C, Lampis A, Stuttle C, Corbishley CM, Stavrinides V, Zapata L, Spiteri I, Heide T, Gallagher L, James C, Ramazzotti D, Gao A, Kote-Jarai Z, Acar A, Truelove L, Proszek P, Murray J, Reid A, Wilkins A, Hubank M, Eeles R, Dearnaley D and Sottoriva A. Tumor evolution metrics predict recurrence beyond 10 years in locally advanced prostate cancer. Nat Cancer 2024; [Epub ahead of print].
[72] Nakata W, Mori H, Tsujimura G, Tsujimoto Y, Gotoh T and Tsujihata M. Pilot study of an artificial intelligence-based deep learning algorithm to predict time to castration-resistant prostate cancer for metastatic hormone-naive prostate cancer. Jpn J Clin Oncol 2022; 52: 1062-1066.
[73] Tsuneki M, Abe M, Ichihara S and Kanavati F. Inference of core needle biopsy whole slide images requiring definitive therapy for prostate cancer. BMC Cancer 2023; 23: 11.
[74] Spratt DE, Tang S, Sun Y, Huang HC, Chen E, Mohamad O, Armstrong AJ, Tward JD, Nguyen PL, Lang JM, Zhang J, Mitani A, Simko JP, DeVries S, van der Wal D, Pinckaers H, Monson JM, Campbell HA, Wallace J, Ferguson MJ, Bahary JP, Schaeffer EM, Sandler HM, Tran PT, Rodgers JP, Esteva A, Yamashita R and Feng FY. Artificial intelligence predictive model for hormone therapy use in prostate cancer. NEJM Evid 2023; 2: EVIDoa2300023.
[75] Reis-Filho JS and Kather JN. Overcoming the challenges to implementation of artificial intelligence in pathology. J Natl Cancer Inst 2023; 115: 608-612.
[76] Zhang DY, Venkat A, Khasawneh H, Sali R, Zhang V and Pei Z. Implementation of digital pathology and artificial intelligence in routine pathology practice. Lab Invest 2024; 104: 102111.
[77] Campanella G, Rajanna AR, Corsale L, Schuffler PJ, Yagi Y and Fuchs TJ. Towards machine learned quality control: a benchmark for sharpness quantification in digital pathology. Comput Med Imaging Graph 2018; 65: 142-151.
[78] Atallah NM, Toss MS, Verrill C, Salto-Tellez M, Snead D and Rakha EA. Potential quality pitfalls of digitalized whole slide image of breast pathology in routine practice. Mod Pathol 2022; 35: 903-910.