Journal of Pathology Informatics 19 (2025) 100524

Contents lists available at ScienceDirect

Journal of Pathology Informatics

journal homepage: www.elsevier.com/locate/jpi

Review Article

Predicability of PD-L1 expression in cancer cells based solely on H&E-stained sections

Gavino Faa [a,b], Matteo Fraschini [c,*], Pina Ziranu [d], Andrea Pretta [d], Giuseppe Porcu [e], Luca Saba [a], Mario Scartozzi [d], Nazar Shokun [f,g], Massimo Rugge [h]

[a] Department of Medical Sciences and Public Health, Università degli Studi di Cagliari, 09123 Cagliari, Italy
[b] Department of Biology, College of Science and Technology, Temple University, Philadelphia, PA 19122, USA
[c] Department of Electrical and Electronic Engineering, Università degli Studi di Cagliari, 09123 Cagliari, Italy
[d] Medical Oncology Unit, University Hospital of Cagliari, Università degli Studi di Cagliari, 09123 Cagliari, Italy
[e] Department of Pathology, Ospedale Oncologico A. Businco, ARNAS G. Brotzu, Cagliari, Italy
[f] National Cancer Institute, Kyiv, Ukraine
[g] Associazione "Angela Serra" per la ricerca sul cancro, Modena, Italy
[h] Department of Medicine – DIMED; General Anatomic Pathology and Cytopathology Unit, Università degli Studi di Padova, 35121 Padova, Italy

ARTICLE INFO

Keywords:
Artificial intelligence
PD-L1
Digital pathology

ABSTRACT

PD-L1 expression is an important biomarker for selecting patients who are eligible for immune checkpoint inhibitor (ICI) therapy. However, evaluating PD-L1 through immunohistochemistry often faces significant interobserver variability and requires considerable time and resources. Recent advancements in artificial intelligence (AI) have transformed the field of pathology, leading to more standardized and reproducible methods for biomarker quantification. In this study, we examine the application of AI-driven models, particularly deep learning algorithms, to predict PD-L1 expression directly from hematoxylin and eosin-stained histological slides. Several AI-based approaches have been studied, demonstrating high accuracy in estimating PD-L1 expression and predicting responses to ICIs across various cancer types. AI-driven assessments of PD-L1 have been shown to reduce the subjectivity associated with manual scoring methods, such as the Tumor Proportion Score and the Combined Positive Score. Moreover, integrating AI with multimodal data, including genomics, radiomics, and real-world clinical data, can further enhance predictive accuracy and improve patient stratification for immunotherapy. Finally, AI-driven computational pathology offers a transformative approach to biomarker evaluation, providing a faster, more objective, and cost-effective alternative to traditional methods, with significant implications for personalized oncology and precision medicine. Despite these promising results, several challenges remain to be addressed, such as the need for large-scale validation, standardization of AI models, and regulatory approvals for clinical implementation. Tackling these issues will be crucial for incorporating AI-based PD-L1 assessments into routine pathology workflows.

Introduction

Programmed cell death 1 (PD-1) and programmed cell death 2 (PD-2) are two key immune checkpoint receptors expressed on the surface of activated T lymphocytes, B lymphocytes, natural killer cells, and monocytes.[1] The interaction of PD-1 and PD-2 with their ligands, PD-L1 and PD-L2, leads to the downregulation of cytotoxic T-cell activity, inducing immunotolerance to cancer cells and ultimately promoting tumor immune escape.[2]

The PD-1/PD-L1 signaling pathway represents a crucial immunological checkpoint that plays a fundamental role in immune evasion across various cancer types,[3] including non-small cell lung cancer (NSCLC),[4] colorectal cancer (CRC),[5] gastric cancer,[6] hepatocellular carcinoma (HCC),[7] and diffuse large B-cell lymphoma (DLBCL).[8] PD-L1 expression on tumor cells is sufficient to inhibit CD8+ T-cell cytotoxicity, thus enabling immune evasion.[9] Additionally, PD-L1 can be secreted by cancer cells via exosomes, further contributing to immunosuppression.[10]

Notably, PD-L1 expression is not limited to tumor cells but is also observed in the tumor microenvironment (TME), highlighting the crucial role of the TME in PD-1/PD-L1-mediated immune escape and immunotolerance.[11,12] Consequently, the assessment of PD-L1 expression for effective immunotherapy should not be restricted to cancer cells but should also encompass the cells within the TME.[13,14] In summary, when PD-1 receptors on activated T cells interact with PD-L1 expressed by cancer

* Corresponding author.
E-mail address: fraschin@unica.it (M. Fraschini).

http://dx.doi.org/10.1016/j.jpi.2025.100524
Received 10 April 2025; Received in revised form 8 September 2025; Accepted 22 October 2025
Available online 30 October 2025

cells, T-lymphocyte activity is suppressed, allowing tumor cells to evade immune surveillance.[15]

Collectively, these findings underscore PD-L1 as a critical predictive biomarker of response to immunotherapy and the gold-standard for predicting responsiveness to immune checkpoint inhibitors (ICIs).[16–18] Tumor immunotherapy is one of the most innovative approaches in cancer research,[19,20] offering significant benefits to patients across various cancer types, even in advanced stages, by improving overall survival.[21] This work is presented as a literature review, rather than a systematic review, with the aim of summarizing current knowledge and highlighting key developments in the field.

Scoring systems for the evaluation of PD-L1 expression

In clinical practice, the assessment of PD-L1 expression is conducted through immunohistochemistry (IHC) using anti-PD-L1 antibodies (commonly referred to as CD 274). This evaluation typically employs two primary scoring systems. The first system is the Tumor Proportion Score (TPS), which quantifies the percentage of PD-L1-positive tumor cells relative to the total number of viable tumor cells. This is calculated by dividing the number of PD-L1-positive tumor cells by the total number of viable tumor cells and then multiplying by 100. The TPS can range from 0 % to 100 %, and it is essential to ensure that a minimum of 100 tumor cells are present in the sample for the TPS assessment to be considered valid.[22]

The second system is known as the Combined Positive Score (CPS). This score evaluates the total number of PD-L1 immunoreactive cells, including both tumor cells and immune cells within the TME. The CPS is derived by dividing the total number of PD-L1 immunoreactive cells by the total number of viable tumor cells, then multiplying by 100. Similar to the TPS, the CPS range extends from 0 % to 100 %.[23] Scores for the CPS are categorized numerically from 1 to 5, with specific thresholds: score 1 for less than 1 %, score 2 for more than 1 % but less than or equal to 5 %, score 3 for more than 5 % but less than or equal to 10 %, score 4 for more than 10 % but less than or equal to 50 %, and score 5 for more than 50 %.[5] Both the TPS and CPS scoring systems for PD-L1 expression evaluation are complex and may require considerable time for accurate analysis. Furthermore, the proficiency and experience of pathologists play a significant role in the reliable quantitative assessment of PD-L1 immunostaining scores, leading to notable variability in results among different practitioners.[24–28]

Interindividual variability in manual evaluation of PD-L1 expression can occur in both cancer cells and in the TME. Significant interpathologist variability has been reported in the initial studies evaluating PD-L1 immunoreactivity in NSCLC, leading to considerable disagreement among pathologists.[24,25] Notably, in the Ratcliffe study, the agreement between two pathologists improved from a low level of 75 % to a very high level of 96 % as the PD-L1 positivity increased from 1 % to 50 %. The complexity of the previously described scoring systems results in considerable interindividual variability among pathologists when applying the CPS for PD-L1 across multiple cancer types.[26] This variability is particularly pronounced in NSCLC[27] and, more recently, in gastric and esophageal cancers.[28]

There are several reasons for the high variability among pathologists in evaluating PD-L1 expression. One key reason is the uneven distribution of immunoreactive tumor cells. For instance, in gastric cancer, PD-L1 expression often appears more widespread in superficial biopsies, whereas deeper biopsies tend to show a more irregular distribution of PD-L1 immunostaining.[29] In this study, tissue microarray cores taken from the invasive front of gastric carcinoma did not accurately represent the PD-L1 expression observed in the entire tumor mass. Furthermore, due to the patchy distribution of PD-L1-expressing cancer cells, it is essential to perform at least four surface biopsies to achieve a PD-L1 expression evaluation that closely resembles that of the resection specimen.[29]

A recently proposed method for evaluating PD-L1 expression in gastric cancer addresses the challenge of heterogeneous PD-L1 distribution is the Tumor Area Proportion (TAP) Score. Unlike the TPS and the CPS, which count individual cells, the TAP score quantifies the percentage of the tumor area that shows PD-L1 positivity in relation to the total tumor area. This method may provide a more accurate assessment, especially in cases where PD-L1 expression is patchy or localized in specific regions of the tumor.[30,31] Both TPS and CPS can be complex and time-consuming, with accurately heavily reliance on the expertise of the pathologist. This reliance can lead to significant variability in PD-L1 immunostaining assessments between observers. The newly introduced TAP score aims to reduce these challenges by offering an area-based quantification of PD-L1 expression, which could be especially beneficial for tumors with heterogeneous expression patterns, such as gastric cancer. However, its clinical implementation is currently limited by a lack of standardization, interobserver variability, and the need for larger tissue samples. Additionally, compared to TPS and CPS (Table 1), which are widely validated, the TAP score has only recently been introduced and has been used in a limited number of prospective studies. Therefore, further validation is required before it can be established as a standard method for PD-L1 assessment and its predictive value for immunotherapy response.[30–32]

Given these challenges with traditional quantitative evaluations of PD-L1, the development of artificial intelligence (AI)-driven models to predict PD-L1 expression across various tumor types is becoming increasingly important. Recent advancements in AI are improving pathology by providing standardized methods for biomarker quantification. This study focuses on the use of AI-driven models, particularly deep learning (DL) algorithms, to predict PD-L1 expression from hematoxylin and eosin (H&E)-stained histological slides (Table 2). These advancements aim to enhance diagnostic accuracy and improve patient outcomes in cancer therapy.

Artificial intelligence-assisted score analysis of PD-L1

A recent systematic review focused on the use of AI for predictive biomarker discovery in oncology analyzed 90 studies, with 80 % of them published between 2021 and 2022. This trend highlights the increasing interest of the scientific community in this area in recent years.[33] Multiple studies have assessed the application of DL-driven models in the analysis of multiple biomarkers in IHC, including HER2 status in whole slide images (WSIs),[34] Ki67,[35] EGFR, and KRAS mutation status.[36] The use of a well-trained DL model for the quantitative evaluation of IHC markers could help overcome the drawbacks associated with traditional methods, such as time consumption and interobserver variability in evaluating certain IHC markers.[37,38]

AI-assisted systems have been employed to evaluate PD-L1 expression in NSCLC, demonstrating the capability of AI-driven models to predict

Table 1
Scoring systems for the evaluation of PD-L1 expression.

Scoring system	Calculation method	Pros	Cons
Tumor Proportion Score (TPS)	$\frac{\text{PDL1}^+\text{ tumor cells}}{\text{Total viable tumor cells}} * 100$	✓ Widely validated ✓ Correlates with ICI response ✓ Standardized cutoffs	✗ Ignores immune cells ✗ May not capture full PD-L1 role in tumor immune evasion
Combined Positive Score (CPS)	$\frac{\text{PDL1}^+\text{ tumor cells+immne cells}}{\text{Total viable tumor cells}} * 100$	✓ Includes tumor & immune cells ✓ Broadly used in clinical trials	✗ More complex & time-consuming ✗ Higher interobserver variability
Tumor Area Proportion (TAP)	$\frac{\text{PDL1}^+\text{ tumor area}}{\text{Total tumor area}} * 100$	✓ Considers PD-L1 spatial distribution ✓ Reflects PD-L1 heterogeneity better than TPS/CPS ✓ Used in recent phase 3 trials	✗ Less validated than TPS/CPS ✗ Requires standardization ✗ Limited prospective studies.

Table 2
Summary of studies on AI-based PD-L1 evaluation.

Study	Cancer type	AI method	PD-L1 evaluation method	Main findings	Scoring	Reference
Wiesweg et al. (2020)	NSCLC	ML	Gene expression analysis	AI predicted response to ICIs independent of PD-L1 levels	H&E/Global	39
Wu et al. (2022)	NSCLC	DL using WSIs from 22C3 to assess TPS	TPS quantification in IHC	AI improved precision diagnosis of PD-L1 expression	IHC/ROI	40
Baxi et al. (2022)	Various cancers	CNN	PD-L1 IHC expression in tumor and immune cells	AI outperformed manual scoring in ICI response prediction	IHC/ROI	42
Jin et al. (2024)	Various cancers	MILTS model	PD-L1 prediction from H&E-stained slides	AI identified histopathological patterns linked to PD-L1 expression	H&E/Slide	43
Kim et al. (2020)	Gastric cancer	Digital image analysis	PD-L1 quantification in IHC	Digital PD-L1 scoring correlated well with manual pathologist assessment	IHC/Slide	44
Sha et al. (2019)	NSCLC	DL using WSIs	PD-L1 status from H&E-stained slides	AI model predicted PD-L1 expression with AUC = 0.80. PD-L1 expression links to tumor pattern and TME	H&E/ROI	45
Shamai et al. (2022)	Breast cancer	DL techniques from H&E-stained images	PD-L1 status from IHC and H&E-stained slides	AI model predicted PD-L1 status with AUC = 0.91–0.93	H&E and IHC/Global	46
Van Eekelen et al. (2024)	NSCLC	DL algorithm	Digital vs. manual quantification	AI achieved agreement with pathologists in TPS scoring	IHC/ROI	50
Yan et al. (2024)	DLBCL	CNN + SAM	PD-L1 IHC assessment	AI-based scoring had high agreement with expert hematopathologists	IHC/ROI	52
Liu et al. (2024)	Gastric cancer	ICIsNet model (DL-driven tool based on histopathological images)	H&E-stained slides	AI predicted response to PD-1 blockade with AUC = 0.92–0.96	H&E/ROI	54
Wang et al. (2021)	Breast cancer	DL-based AI-assisted for PD-L1 IHC quantification	PD-L1 assessment in IHC	AI-assisted quantification improved accuracy and efficiency	IHC/ROI	60

ML = Machine Learning; PD-L1 = programmed death-ligand 1; AI = Artificial Intelligence; ICIs = Immune Checkpoints Inhibitors; DL = Deep Learning; CNN = Convolutional Neural Network; SAM = Segment Anything Model; WSIs = Whole-Slide Images; TPS = Tumor Proportion Score; MILTS = Teacher-Student collaborated Multiple Instance Learning framework; H&E = Hematoxylin and Eosin; IHC = immunohistochemistry; TME = Tumor Microenvironment; TNBC = Triple Negative Breast Cancer. As for the scoring system, Global refers to methods that produce a global score at the patient level, ROI (Region of Interest) refers to methods that produce a score at the level of a region, and Slide refers to methods that produce a score at the level of a single slide.

responses to immunotherapy.[24,39,40] Research shows that these models can more accurately predict responses to ICIs compared to traditional IHC alone across various tumor types.[41] A strong correlation has been observed between digital and manual quantification of PD-L1 expression in tumor and immune cells across multiple cancer types, including bladder urothelial carcinoma, CRC, gastric carcinoma, cervical squamous cell carcinoma, endocervical carcinoma, renal papillary cell carcinoma, triple-negative breast cancer, and melanoma.[42,43] In these studies, digital scoring of PD-L1 expression has been closely linked to predicting responses to ICI therapy in clinical trials, often outperforming manual pathological quantification. Additionally, when predicting responses to the checkpoint inhibitor pembrolizumab in patients with gastric cancer, the results of PD-L1 expression evaluated through digital image analysis were not significantly different from those provided by expert gastrointestinal pathologists conducting manual evaluations.[44]

PD-L1 status extraction from H&E-stained sections

In 2019, Sha and colleagues[45] contributed significantly to the ongoing discussion about PD-L1 and AI models. Whereas previous studies focused primarily on how well AI-driven models could interpret IHC images of PD-L1 expression, Sha and their team took a different approach. They evaluated the ability of algorithms to predict PD-L1 expression using H&E-stained images.[45] The authors found that the average computational time to generate a PD-L1 probability was just 40 s, highlighting the potential advantages of this method for reducing diagnostic time in clinical practice. Despite its significance, the study had some limitations, as indicated by an AUC of 0.80 for predicting PD-L1 status. The DL model was trained in 48 cases and tested on 82 cases of lung cancer, with a particular focus on adenocarcinomas. The relatively low number of training cases contributed to some performance issues, as the model sometimes misclassified PD-L1 negative tumors as PD-L1 positive.

The approach proposed by Sha and colleagues was set aside for several years until 2022, when Shamai and his team revisited the project aimed at extracting PD-L1 status from whole-slide H&E-stained images.[46] The authors emphasized the importance of quantifying PD-L1 to predict the efficacy of immunotherapy and discussed the challenges related to manual quantification, which often suffers from significant interobserver variability.[47] To overcome this challenge, Shamai and his colleagues developed a dataset to evaluate PD-L1 prediction from H&E-stained sections of a large cohort of breast cancer patients. Their system was validated using two external datasets, demonstrating robustness and consistent predictive performance for both PD-L1 quantification and immunotherapy response. The ability to extract IHC data on PD-L1 expression solely from H&E sections represents a significant advancement in pathology laboratories. This AI-driven model can predict PD-L1 status with a high area under the curve (AUC) of 0.91–0.93, all without increasing costs and within a remarkably short time frame.[46]

Different scoring systems for PD-L1 immunoreactivity

As often happens in pathology research, the anticipated breakthroughs in AI-driven PD-L1 evaluation within breast pathology were not immediately embraced in other areas of histopathology. One study examined the predictive value of various scoring systems for PD-L1 assessment in lung cancer, emphasizing the manual evaluation of this biomarker in lung biopsies. It concluded that the CPS serves as a better predictive biomarker compared to the TPS.[48] Another study examining the prognostic significance of PD-L1 in gastric cancer patients with peritoneal metastases found that high manual expression of PD-L1 (CPS > 10) should be regarded as an independent favorable prognostic factor for this type of tumor.[49] More recently, research comparing DL and pathologist quantification of PD-L1 expression in NSCLC demonstrated that the DL algorithm produced favorable agreement with pathologists when quantifying the TPS of PD-L1.[50] The newly developed DL system for the automatic quantification of PD-L1 expression in NSCLC enabled efficient TPS quantification.[50] Moreover, the study of Butter and colleagues discussed how a pathologist's personality can influence interobserver variability and diagnostic accuracy in quantifying PD-L1 expression,[51] highlighting the challenges associated with manual quantification of PD-L1.

PD-L1 expression in lymphomas

AI-based assessment of PDL1 expression has been proposed in DLBCL, an aggressive form of blood cancer characterized by rapid progression.[8] A recent study reported that an AI-driven algorithm, designed to automatically evaluate IHC for PD-L1 in lymphomas, demonstrated a high level of agreement with expert hematopathologists.[52] These findings underscore the potential of AI-driven models to improve the objectivity and interpretability of PD-L1 quantification in DLBCL, contributing to the development of new treatment strategies for this aggressive blood cancer type.[53]

The extraction of the PD-L1 status from H&E-stained frozen sections

A significant advancement in the study of PD-L1 expression levels extracted from histopathology slides is represented by a pivotal work from Darui and colleagues, published online on April 9, 2024.[43] In this study, the authors introduced a weakly supervised learning-based method called MILTS (Teacher–Student Collaborated Multiple Instance Learning Framework). This method extracts tile information from whole slide H&E images to generate slide-level embeddings for final predictions. The results of this study highlight a notable morpho-transcriptomic connection across various cancer types linked to PD-L1 gene expression. The weighted average AUC for frozen sections was found to be 0.83, whereas the AUC for paraffin-section WSIs was 0.74.

When applying MILTS to CRC WSIs, the algorithm identified specific morphotypes associated with PD-L1 “hot” tumor regions. These regions typically exhibited a mixed inflammatory stroma, characterized by high levels of eosinophil infiltration, and a cribriform growth pattern of cancer cells with hyperchromatic nuclei. In contrast, low PD-L1 expression was associated with tumor necrosis, coagulation necrosis, and mucinous differentiation of cancer cells. The AI model's predictions were validated by IHC data, which revealed high PD-L1 immunoreactivity in these specific tumor areas.

The findings clearly show that the molecular basis of cancer cells can be analyzed through cell morphology and features detectable in the TME using DL models. This research demonstrates that predicting PD-L1 expression based solely on morphology (using H&E WSIs) could serve as a revolutionary tool, potentially reducing the need for additional costly tests in the near future.

One interesting observation from this study is that the performance of MILTS is higher with frozen sections compared to paraffin sections. This difference may be because frozen tissues are generally better at preserving the cancer cells' structural and molecular integrity, a trend also noted in AI-driven models applied to CRC biopsies.[50] These data indicate that an AI-driven algorithm can identify unique histopathological patterns associated with PD-L1 expression, which may be difficult to discern under a microscope due to their uneven distribution throughout the tumor mass. Overall, the Jin study[43] provides valuable insights into the complex interactions between histology and molecular biology.

Very recent progress in PD-1 blockade in advanced gastric cancer

A recent study conducted by Liu and colleagues aimed to develop a DL model capable of extracting information about the response to first-line PD-1 blockade from pre-treatment H&E-stained WSIs of patients with advanced gastric cancer.[54] The researchers utilized a new DL-driven tool called the ICIsNet model, which demonstrated impressive predictive capabilities in both internal and external cohorts of gastric cancer patients, assessing the immunotherapy response of each of the 313 patients analyzed. The AUC values highlighted the effectiveness of this approach in distinguishing between good responders and poor responders to immunotherapy. In the Liu study, the AUC ranged from 0.92 to 0.96, showcasing the remarkable predictive power of the model. The ICIsNet model also facilitated the creation of the Immune Checkpoint Inhibitors Response Score (ICIsRS), a novel biomarker derived from H&E-stained WSIs.

Additionally, the study addressed the typical black-box nature of neural networks and the challenges associated with explaining their self-learning characteristics. To tackle the important issue of DL model explainability, Liu and colleagues generated heatmaps for the entire WSI based on the output of each tile. These heatmaps visually represented histological features linked to a high ICIsRS and a more sensitive response to immunotherapy, including abundant lymphocytic infiltration, lower cellular heterogeneity, and stronger cell adhesion. At the other extreme of the spectrum, lower levels of lymphocytic infiltration, diffuse and signet ring cell architecture, and mucinous differentiation are associated with a poor response to immunotherapy. These findings may represent a novel aspect in understanding the complexities of explainable DL models. This study suggests that higher lymphocytic infiltration correlates with better outcomes from immunotherapy. The substantial presence of immune cells is crucial for the immune system's effectiveness in attacking cancer cells, a process enhanced by immunotherapies targeting checkpoint inhibitors such as PD-1 and PD-L1. Conversely, low levels of lymphocytic infiltration may diminish the effectiveness of immunotherapy.[54]

Collectively, the study by Liu and colleagues confirms the value of H&E-stained WSI and highlights the numerous molecular and IHC features that can be extracted by a well-trained DL model, such as the ICIsNet model used in this study. Another significant conclusion from this research is the utility of generating heat maps for the entire WSI to interpret the outputs of neural networks. Finally, the authors contribute importantly to the explainability of DL-driven models, demonstrating that well-trained algorithms can identify critical features for predicting the efficacy of immunotherapy, which may not be easily recognized by the human eye, even by experienced pathologists. Recognizing these features represents a key challenge for pathologists and oncologists involved in advancing computational pathology.

Advantages of computational pathology

AI represents a revolutionary tool for discovering predictive biomarkers in human tumors. AI-based methods are transforming the landscape for researchers involved in biomarker discovery, as well as for pathologists and oncologists. These methods demonstrate the effectiveness of well-trained algorithms in extracting relevant molecular data from routinely stained H&E sections.[55]

When applied in clinical settings, the advantages of this pioneering approach are numerous. For example, the average computational time to generate a PD-L1 probability map from standard H&E-stained sections is approximately 40 s, with a range from 7.9 to 66 s.[45] This rapid processing means that using a robust, well-trained DL model, pathologists can predict PD-L1 scores and assess the potential efficacy of immunotherapy within a matter of seconds. This quick analysis provides practicing oncologists with crucial information to inform their therapeutic strategies for any single patient.[43,52]

The ability of AI approaches to identify significant biomarkers from H&E sections could lead to substantial cost savings. In summary, the traditional role of IHC, which relies on complex machinery operated by expert technicians, may be completely re-evaluated. Identifying genes related to responses to ICIs, such as PD-L1, without the use of antibodies, presents significant challenges but also exciting opportunities for developing molecular tests, even in peripheral pathology labs in low-resource settings.

Another significant advantage of this innovative approach to “artificial” histology, which relies on AI-driven models, is the capability of analyzing digitalized histological slides to extract valuable information beyond just PD-L1 expression. This information can enhance clinical decision-making in oncology. Immune pathology, which underpins therapies based on ICIs, remains a relatively underexplored area in human pathology. AI-driven methods have the potential to identify novel meta-biomarkers that can more accurately predict the effectiveness of ICI therapies.[33] This hypothesis is supported by recent studies demonstrating that AI models can predict immune and inflammatory gene signatures in hepatocellular carcinoma directly from histological images.[56]

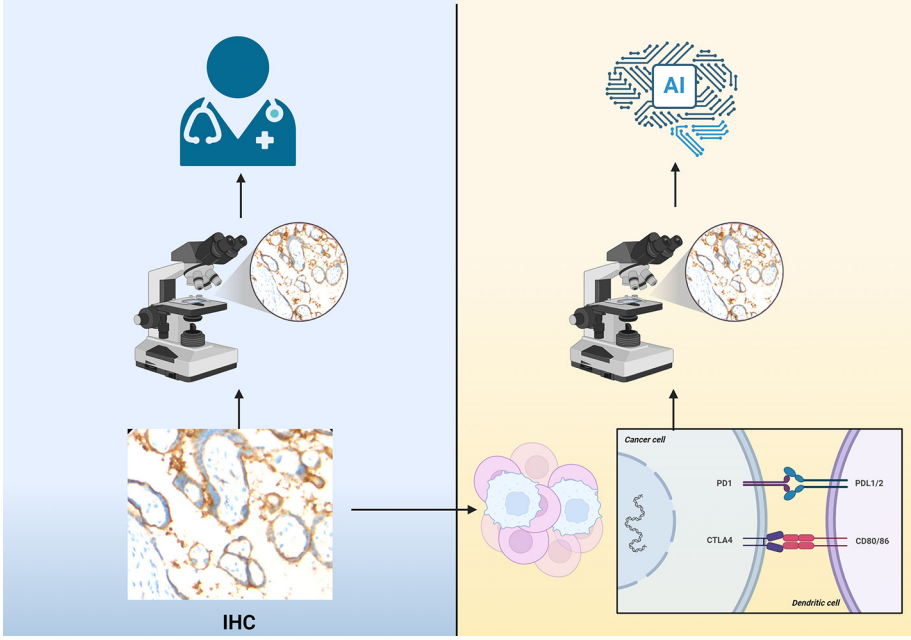

Fig. 1. Comparison between conventional and AI-assisted assessment of PD-L1 expression in cancer tissue.

In summary, these findings highlight the substantial potential of AI, particularly DL models, to extract a diverse range of biomarkers from histological images, thereby facilitating the discovery of new predictive biomarkers (Fig. 1). As a result, computational pathology (or pathomics) presents an exciting opportunity to become a significant diagnostic tool for medical professionals, including physicians, oncologists, and surgeons. By providing histological, genetic, and molecular information efficiently and cost-effectively, we have the opportunity to advance the field of cancer diagnosis and treatment.[57]

Another advantage of computational pathology is the acquisition of a large dataset of digitized histological images (pathomics), which could be integrated with various types of clinical data, including real-world data, genomic, epigenomic, microbiological, radiological (Radiomics), and laboratory data. Integrating multimodal data could allow the creation of metabiomarkers that enhance the performance of AI-driven models, resulting in improved AUC metrics when compared to unimodal models using the same dataset.[58]

Another important advantage of computational pathology is the development of new algorithms that reduce interobserver variability among both expert and general pathologists. This variability is particularly pronounced in the evaluation of PD-L1 expression.[47,59,60] DL models demonstrate a superior ability to assess PD-L1, making them valuable auxiliary tools in clinical practice. However, when these models are compared to their ability to directly and quantitatively extract molecular biomarker expression from histology, the achievement of mitigating interobserver variability appears to be a modest accomplishment in this exciting field.

Challenges to be solved

Despite all the advantages of computational pathology as described here, many challenges remain to be solved before AI models can be integrated into clinical practice, particularly in immuno-oncology. The major limitations hindering the introduction of these algorithms into the clinical workflow include the following:

1. Data quality and quantity: effective algorithm performance relies on large volumes of high-quality data. Unfortunately, oncological datasets are often incomplete, heterogeneous, biased, and inherently complex.
2. Model selection challenges: the existence of various ML and DL models can make it difficult for researchers and clinicians to determine which methodology best suits their needs. It is suggested that classical ML models may be more effective, especially when working with low-volume data.[64]
3. Certification requirements: certification is a crucial step to ensure that AI models meet the necessary standards for clinical practice. Therefore, a standardized process for the certification of AI-driven models should be promoted.
4. Trust issues among medical professionals: a significant challenge to the adoption of algorithms in medicine is the lack of trust from healthcare professionals in these new computational tools and in digital pathology, although this remains a matter of ongoing debate within the scientific community.[61–63] Research focusing on the interpretability of these models is essential for building trust in ML and DL systems. Using eXplainable Artificial Intelligence methods can enhance the transparency of algorithms, making them easier for clinicians and pathologists to understand.[64] This approach could help address the “black-box” issue associated with these models.[65]
5. Robustness and validation: ML models that perform well on internal datasets must also be tested with diverse datasets from multiple centers to ensure their robustness, which is often lacking and may not meet the standards required in clinical practice. Comprehensive validation of AI models using various heterogeneous and relevant datasets is essential to assess their reliability.[66] Deeper insights in this direction can be enhanced by comparing each AI model against multiple validation datasets through a method known as divergent validation, which may also improve model interpretability and generalizability[67] or cross-validation and grand challenges.
6. Algorithm biases: algorithm biases may arise from differences in staining intensity, data structure, and patient demographics. It is important to avoid bias to enhance the performance of AI models.[68–70]

Conclusions

In conclusion, the studies reviewed demonstrate the capability of new algorithms to extract molecular data from H&E-stained sections. This ability represents an intriguing and revolutionary aspect of recent research on the application of AI in histopathology. Nevertheless, it is important to highlight that AI-based approaches depend on digital WSI, which is not yet widely accessible, particularly in resource-limited settings. Large-scale prospective validation studies are essential to develop new interpretable

and responsible AI tools that meet the necessary standards for their integration into clinical workflows. Creating and adopting generalizable and robust models is the correct way to address the challenges currently preventing the adoption of AI-driven models in clinical settings, ultimately enhancing the confidence of pathologists and clinicians in computational pathology. In perspective, significant progress in this field will require the development of more standardized algorithm evaluation protocols, using diverse and representative datasets that integrate robust follow-up clinical information, and using computational pathology tools in prospective clinical trials. Patient response to therapy remains the most clinically relevant benchmark, and designing studies that align algorithmic predictions with treatment outcomes will be essential, despite the open challenges of collecting and sharing such data.[71] Furthermore, without coordinated regulatory frameworks and collaborative settings, AI algorithms risk spreading rapidly without interoperability, undermining clinical trust and adoption. AI progress in pathology will require overcoming challenges in data quality, validation, clinical alignment, workflow integration, and regulatory approval. Addressing these correlated issues collectively is crucial to unlock AI's full potential in clinical practice.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Sharpe AH, Wherry EJ, Ahmed R, Freeman GJ. The function of programmed cell death 1 and its ligands in regulating autoimmunity and infection. Nat Immunol 2007;8:239–245. doi:10.1038/ni1443.
2. Cui J-W, Li Y, Yang Y, et al. Tumor immunotherapy resistance: revealing the mechanism of PD-1/PD-L1-mediated tumor immune escape. Biomed Pharmacother 2024;171, 116203. doi:10.1016/j.biopha.2024.116203.
3. Wang Y, Wang H, Yao H, Li C, Fang J-Y, Xu J. Regulation of PD-L1: emerging routes for targeting tumor immune evasion. Front Pharmacol 2018:9. doi:10.3389/fphar.2018.00536.
4. Vallejo J, Singh H, Larkins E, et al. Impact of increasing PD-L1 levels on outcomes to PD-1/PD-L1 inhibition in patients with NSCLC: a pooled analysis of 11 prospective clinical trials. Oncologist 2024;29:422–430. doi:10.1093/oncolo/oyae006.
5. Frančina M, Mikuš M, Mamić M, et al. Evaluation of PD-L1 expression in colorectal carcinomas by comparing scoring methods and their significance in relation to clinicopathologic parameters. Diagnostics 2024;14:1007. doi:10.3390/diagnostics14101007.
6. Hirano H, Yamada Y, Nagashima K, et al. Impact of PD-L1 expression on survival in patients with unresectable/recurrent gastric cancer receiving first-line chemotherapy without immune checkpoint inhibitors. JCO 2024;42:391. doi:10.1200/JCO.2024.42.3_suppl.391.
7. Kang JG, Han K, Chung T, Rhee H. Prediction of PD-L1 expression in unresectable hepatocellular carcinoma with gadoxetic acid-enhanced MRI. Eur J Radiol 2024;181, 111772. doi:10.1016/j.ejrad.2024.111772.
8. Song M-K, Park B-B, Uhm J. Understanding immune evasion and therapeutic targeting associated with PD-1/PD-L1 pathway in diffuse large B-cell lymphoma. Int J Mol Sci 2019;20:1326. doi:10.3390/ijms20061326.
9. Juneja VR, McGuire KA, Manguso RT, et al. PD-L1 on tumor cells is sufficient for immune evasion in immunogenic tumors and inhibits CD8 T cell cytotoxicity. J Exp Med 2017;214:895–904. doi:10.1084/jem.20160801.
10. Chen G, Huang AC, Zhang W, et al. Exosomal PD-L1 contributes to immunosuppression and is associated with anti-PD-1 response. Nature 2018;560:382–386. doi:10.1038/s41586-018-0392-8.
11. Jiang X, Wang J, Deng X, et al. The role of microenvironment in tumor angiogenesis. J Exp Clin Cancer Res 2020;39:204. doi:10.1186/s13046-020-01709-5.
12. Donisi C, Pretta A, Pusceddu V, et al. Immunotherapy and cancer: the multi-omics perspective. Int J Mol Sci 2024;25:3563. doi:10.3390/ijms25063563.
13. Sadeghi Rad H, Monkman J, Warkiani ME, et al. Understanding the tumor microenvironment for effective immunotherapy. Med Res Rev 2021;41:1474–1498. doi:10.1002/med.21765.
14. Baghban R, Roshangar L, Jahanban-Esfahlan R, et al. Tumor microenvironment complexity and therapeutic implications at a glance. Cell Commun Signal 2020;18:59. doi:10.1186/s12964-020-0530-4.
15. Han Y, Liu D, Li L. PD-1/PD-L1 pathway: current researches in cancer. Am J Cancer Res 2020;10:727–742.
16. Incorvaia L, Fanale D, Badalamenti G, et al. Programmed death ligand 1 (PD-L1) as a predictive biomarker for pembrolizumab therapy in patients with advanced non-small-cell lung cancer (NSCLC). Adv Ther 2019;36:2600–2617. doi:10.1007/s12325-019-01057-7.
17. Diaz LA, Shiu K-K, Kim T-W, et al. Pembrolizumab versus chemotherapy for microsatellite instability-high or mismatch repair-deficient metastatic colorectal cancer (KEYNOTE-177): final analysis of a randomised, open-label, phase 3 study. Lancet Oncol 2022;23:659–670. doi:10.1016/S1470-2045(22)00197-8.
18. Dubois M, Liscia N, Brunetti O, et al. The role of immune checkpoint inhibitors in the treatment sequence of advanced gastric or gastro-esophageal junction cancer: a systematic review and meta-analysis of randomized trials. Crit Rev Oncol Hematol 2022;173, 103674. doi:10.1016/j.critrevonc.2022.103674.
19. Grigg C, Rizvi NA. PD-L1 biomarker testing for non-small cell lung cancer: truth or fiction? J Immunother Cancer 2016;4:48. doi:10.1186/s40425-016-0153-x.
20. Sengupta R, Honey K. AACR cancer progress report 2019: transforming lives through innovative cancer science. Clin Cancer Res 2019;25:5431. doi:10.1158/1078-0432.CCR-19-2655.
21. Llovet JM, Castet F, Heikenwalder M, et al. Immunotherapies for hepatocellular carcinoma. Nat Rev Clin Oncol 2022;19:151–172. doi:10.1038/s41571-021-00573-2.
22. Lea D, Zaharia C, Søreide K. Programmed death ligand-1 (PD-L1) clone 22C3 expression in resected colorectal cancer as companion diagnostics for immune checkpoint inhibitor therapy: a comparison study and inter-rater agreement evaluation across proposed cut-offs and predictive (TPS, CPS and IC) scores. Cancer Treat Res Commun 2024;38, 100788. doi:10.1016/j.ctarc.2023.100788.
23. Mercier A, Conan-Charlet V, Quintin-Roué I, Doucet L, Marcorelles P, Uguen A. Reproducibility in PD-L1 immunohistochemistry quantification through the tumor proportion score and the combined positive score: could dual immunostaining help pathologists? Cancers (Basel) 2023;15:2768. doi:10.3390/cancers15102768.
24. Büttner R, Gosney JR, Skov BG, et al. Programmed death-ligand 1 immunohistochemistry testing: a review of analytical assays and clinical implementation in non-small-cell lung cancer. J Clin Oncol 2017;35:3867–3876. doi:10.1200/JCO.2017.74.7642.
25. Ratcliffe MJ, Sharpe A, Midha A, et al. Agreement between programmed cell death ligand-1 diagnostic assays across multiple protein expression cutoffs in non-small cell lung cancer. Clin Cancer Res 2017;23:3585–3591. doi:10.1158/1078-0432.CCR-16-2375.
26. Nuti S, Zhang Y, Zerrouki N, et al. High interobserver and intraobserver reproducibility among pathologists assessing PD-L1 CPS across multiple indications. Histopathology 2022;81:732–741. doi:10.1111/his.14775.
27. Chang HY, Jung CK, Woo JI, et al. Artificial intelligence in pathology. J Pathol Transl Med 2019;53:1-12. doi:10.4132/jptm.2018.12.16.
28. Robert ME, Rüschoff J, Jasani B, et al. High interobserver variability among pathologists using combined positive score to evaluate PD-L1 expression in gastric, gastroesophageal junction, and esophageal adenocarcinoma. Mod Pathol 2023;36, 100154. doi:10.1016/j.modpat.2023.100154.
29. Schoemig-Markiefka B, Eschbach J, Scheel AH, et al. Optimized PD-L1 scoring of gastric cancer. Gastric Cancer 2021;24:1115–1122. doi:10.1007/s10120-021-01195-4.
30. Moehler M, Oh D-Y, Kato K, et al. 397MO tislelizumab (TIS) plus chemotherapy (CT) vs placebo (PBO) plus CT in HER2-negative advanced or metastatic gastric or gastro-esophageal junction adenocarcinoma (GC/GEJC): PD-L1 biomarker analysis from RATIONALE-305. Ann Oncol 2024;35:S160–S161. doi:10.1016/j.annonc.2024.05.312.
31. Qiu M-Z, Oh D-Y, Kato K, et al. Tislelizumab plus chemotherapy versus placebo plus chemotherapy as first line treatment for advanced gastric or gastro-oesophageal junction adenocarcinoma: RATIONALE-305 randomised, double blind, phase 3 trial. BMJ 2024;385, e078876. doi:10.1136/bmj-2023-078876.
32. Xu J, Kato K, Raymond E, et al. Tislelizumab plus chemotherapy versus placebo plus chemotherapy as first-line treatment for advanced or metastatic oesophageal squamous cell carcinoma (RATIONALE-306): a global, randomised, placebo-controlled, phase 3 study. Lancet Oncol 2023;24:483–495. doi:10.1016/S1470-2045(23)00108-0.
33. Prelaj A, Miskovic V, Zanitti M, et al. Artificial intelligence for predictive biomarker discovery in immuno-oncology: a systematic review. Ann Oncol 2024;35:29–65. doi:10.1016/j.annonc.2023.10.125.
34. Khameneh FD, Razavi S, Kamasak M. Automated segmentation of cell membranes to evaluate HER2 status in whole slide images using a modified deep learning network. Comput Biol Med 2019;110:164–174. doi:10.1016/j.compbiomed.2019.05.020.
35. Saha M, Chakraborty C, Arun I, Ahmed R, Chatterjee S. An advanced deep learning approach for Ki-67 stained hotspot detection and proliferation rate scoring for prognostic evaluation of breast cancer. Sci Rep 2017;7:3213. doi:10.1038/s41598-017-03405-5.
36. Dong Y, Hou L, Yang W, et al. Multi-channel multi-task deep learning for predicting EGFR and KRAS mutations of non-small cell lung cancer on CT images. Quant Imag Med Surg 2021;11:2354–2375. doi:10.21037/qims-20-600.
37. Ahuja AS. The impact of artificial intelligence in medicine on the future role of the physician. PeerJ 2019;7, e7702. doi:10.7717/peerj.7702.
38. Ahmad Z, Rahim S, Zubair M, Abdul-Ghafar J. Artificial intelligence (AI) in medicine, current applications and future role with special emphasis on its potential and promise in pathology: present and future impact, obstacles including costs and acceptance among pathologists, practical and philosophical considerations. A comprehensive review. Diagn Pathol 2021;16:24. doi:10.1186/s13000-021-01085-4.
39. Wiesweg M, Mairinger F, Reis H, et al. Machine learning reveals a PD-L1-independent prediction of response to immunotherapy of non-small cell lung cancer by gene expression context. Eur J Cancer 2020;140:76–85. doi:10.1016/j.ejca.2020.09.015.
40. Wu J, Liu C, Liu X, et al. Artificial intelligence-assisted system for precision diagnosis of PD-L1 expression in non-small cell lung cancer. Mod Pathol 2022;35:403–411. doi:10.1038/s41379-021-00904-9.
41. Lu S, Stein JE, Rimm DL, et al. Comparison of biomarker modalities for predicting response to PD-1/PD-L1 checkpoint blockade: a systematic review and meta-analysis. JAMA Oncol 2019;5:1195–1204. doi:10.1001/jamaoncol.2019.1549.
42. Baxi V, Lee G, Duan C, et al. Association of artificial intelligence-powered and manual quantification of programmed death-ligand 1 (PD-L1) expression with outcomes in patients treated with nivolumab ± ipilimumab. Mod Pathol 2022;35:1529–1539. doi:10.1038/s41379-022-01119-2.

43. Jin D, Liang S, Shmatko A, et al. Teacher-student collaborated multiple instance learning for pan-cancer PDL1 expression prediction from histopathology slides. Nat Commun 2024;15:3063. doi:10.1038/s41467-024-46764-0.
44. Kim H-N, Jang J, Heo YJ, et al. PD-L1 expression in gastric cancer determined by digital image analyses: pitfalls and correlation with pathologist interpretation. Virchows Arch 2020;476:243–250. doi:10.1007/s00428-019-02653-2.
45. Sha L, Osinski BL, Ho IY, et al. Multi-field-of-view deep learning model predicts nonsmall cell lung cancer programmed death-ligand 1 status from whole-slide hematoxylin and eosin images. J Pathol Inform 2019;10:24. doi:10.4103/jpi.jpi_24_19.
46. Shamai G, Livne A, Polónia A, et al. Deep learning-based image analysis predicts PD-L1 status from H&E-stained histopathology images in breast cancer. Nat Commun 2022;13:6753. doi:10.1038/s41467-022-34275-9.
47. Hoda RS, Brogi E, D'Alfonso TM, et al. Interobserver variation of PD-L1 SP142 immunohistochemistry interpretation in breast carcinoma: a study of 79 cases using whole slide imaging. Arch Pathol Lab Med 2021;145:1132–1137. doi:10.5858/arpa.2020-0451-OA.
48. Ulas EB, Hashemi SMS, Houda I, et al. Predictive value of combined positive score and tumor proportion score for immunotherapy response in advanced NSCLC. JTO Clin Res Rep 2023;4, 100532. doi:10.1016/j.jtocrr.2023.100532.
49. Chen X-J, Wei C-Z, Lin J, et al. Prognostic significance of PD-L1 expression in gastric cancer patients with peritoneal metastasis. Biomedicines 2023;11:2003. doi:10.3390/biomedicines11072003.
50. van Eekelen L, Spronck J, Looijen-Salamon M, et al. Comparing deep learning and pathologist quantification of cell-level PD-L1 expression in non-small cell lung cancer whole-slide images. Sci Rep 2024;14:7136. doi:10.1038/s41598-024-57067-1.
51. Butter R, Hondelink LM, van Elswijk L, et al. The impact of a pathologist's personality on the interobserver variability and diagnostic accuracy of predictive PD-L1 immunohistochemistry in lung cancer. Lung Cancer 2022;166:143–149. doi:10.1016/j.lungcan.2022.03.002.
52. Yan F, Da Q, Yi H, et al. Artificial intelligence-based assessment of PD-L1 expression in diffuse large B cell lymphoma. Npj Precis Onc 2024;8:1-12. doi:10.1038/s41698-024-00577-y.
53. Casulo C, Santoro A, Ando K, et al. Durvalumab (anti PD-L1) as monotherapy or in combination therapy for relapsed/refractory (r/r) diffuse large B-cell lymphoma (DLBCL) and follicular lymphoma (FL): a subgroup analysis from the phase 1/2 fusion NHL-001 global multicenter trial. Blood 2019;134:5320. doi:10.1182/blood-2019-124102.
54. Liu Y, Chen W, Ruan R, et al. Deep learning based digital pathology for predicting treatment response to first-line PD-1 blockade in advanced gastric cancer. J Transl Med 2024;22:438. doi:10.1186/s12967-024-05262-z.
55. Faa G, Fraschini M, Didaci L, et al. "Artificial histology" in colonic neoplasia: a critical approach. Dig Liver Dis 2024. doi:10.1016/j.dld.2024.11.001.
56. Zeng Q, Klein C, Caruso S, et al. Artificial intelligence predicts immune and inflammatory gene signatures directly from hepatocellular carcinoma histology. J Hepatol 2022;77:116–127. doi:10.1016/j.jhep.2022.01.018.
57. Bülow RD, Hölscher DL, Costa IG, Boor P. Extending the landscape of omics technologies by pathomics. Npj Syst Biol Appl 2023;9:38. doi:10.1038/s41540-023-00301-9.
58. Vanguri RS, Luo J, Aukerman AT, et al. Multimodal integration of radiology, pathology and genomics for prediction of response to PD-(L)1 blockade in patients with non-small cell lung cancer. Nat Cancer 2022;3:1151–1164. doi:10.1038/s43018-022-00416-8.
59. Reisenbichler ES, Han G, Bellizzi A, et al. Prospective multi-institutional evaluation of pathologist assessment of PD-L1 assays for patient selection in triple negative breast cancer. Mod Pathol 2020;33:1746–1752. doi:10.1038/s41379-020-0544-x.
60. Wang X, Wang L, Bu H, et al. How can artificial intelligence models assist PD-L1 expression scoring in breast cancer: results of multi-institutional ring studies. Npj Breast Cancer 2021;7:1-10. doi:10.1038/s41523-021-00268-y.
61. Catapan S de C, Sazon H, Zheng S, et al. A systematic review of consumers' and healthcare professionals' trust in digital healthcare. Npj Digit Med 2025;8:115. doi:10.1038/s41746-025-01510-8.
62. Sagona M, Dai T, Macis M, Darden M. Trust in AI-assisted health systems and AI's trust in humans. Npj Health Syst 2025;2:10. doi:10.1038/s44401-025-00016-5.
63. Afroogh S, Akbari A, Malone E, Kargar M, Alambeigi H. Trust in AI: progress, challenges, and future directions. Humanit Soc Sci Commun 2024;11:1568. doi:10.1057/s41599-024-04044-8.
64. Ali S, Abuhmed T, El-Sappagh S, et al. Explainable artificial intelligence (XAI): what we know and what is left to attain trustworthy artificial intelligence. Inform Fusion 2023;99, 101805. doi:10.1016/j.inffus.2023.101805.
65. Jarrahi MH, Davoudi V, Haeri M. The key to an effective AI-powered digital pathology: establishing a symbiotic workflow between pathologists and machine. J Pathol Inform 2022;13, 100156. doi:10.1016/j.jpi.2022.100156.
66. Goh WWB, Kabir MN, Yoo S, Wong L. Ten quick tips for ensuring machine learning model validity. PLoS Comput Biol 2024;20, e1012402. doi:10.1371/journal.pcbi.1012402.
67. Ho SY, Phua K, Wong L, Bin Goh WW. Extensions of the external validation for checking learned model interpretability and generalizability. Patterns (N Y) 2020;1, 100129. doi:10.1016/j.patter.2020.100129.
68. Cross JL, Choma MA, Onofrey JA. Bias in medical AI: implications for clinical decision-making. PLOS Digit Health 2024;3, e0000651. doi:10.1371/journal.pdig.0000651.
69. Mittermaier M, Raza MM, Kvedar JC. Bias in AI-based models for medical applications: challenges and mitigation strategies. Npj Digit Med 2023;6:113. doi:10.1038/s41746-023-00858-z.
70. Koçak B, Ponsiglione A, Stanzione A, et al. Bias in artificial intelligence for medical imaging: fundamentals, detection, avoidance, mitigation, challenges, ethics, and prospects. Diagn Interv Radiol n.d.;31:75–88. https://doi.org/10.4274/dir.2024.242854.
71. Varoquaux G, Cheplygina V. Machine learning for medical imaging: methodological failures and recommendations for the future. Npj Digit Med 2022;5:48. doi:10.1038/s41746-022-00592-y.