

Review

Not peer-reviewed version

Advancements in Precision Oncology: AI and Machine Learning Applications in Digital Pathology

[Zamart Ramazanova](#) , [Assem Ashimova](#) , [Yeldar Baiken](#) ^{*} , [Bakhyt Matkarimov](#) , [Askhat Myngbay](#) ^{*} ,
Arshat Urazbaev

Posted Date: 7 June 2024

doi: 10.20944/preprints202406.0378.v1

Keywords: AI; cancer imaging; oncology; machine learning; IHC



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Review

Advancements in Precision Oncology: AI and Machine Learning Applications in Digital Pathology

Zamart Ramazanov^{1,2}, Assem Ashimova³, Yeldar Baiken^{1*}, Bakhyt Matkarimov¹, Askhat Myngbay^{1,4} and Arshat Urazbaev^{1*}

- ¹ PI National Laboratory Astana, Nazarbayev University, Astana 010000, Kazakhstan; bmatkarimov@nu.edu.kz (B.M.); arshat@gmail.com (A.U.)
² School of Engineering and Digital Sciences, Nazarbayev University, Astana 010000, Kazakhstan; zamart.ramazanov@nu.edu.kz (Z.R.)
³ School of Sciences and Humanities, Nazarbayev University, Astana 010000, Kazakhstan
⁴ Educational Institution "Alikhan Bokeikhan University", Abai region, Semey 071400, Kazakhstan
* Correspondence: yebaiken@nu.edu.kz (Y.B.); askhat.myngbay@nu.edu.kz (A.M.)

Abstract: Over the last ten years, precision oncology has led to a surge in the need for prediction tools that facilitate patient screening and stratification for customized treatment regimens. Artificial intelligence (AI) and machine learning algorithms in digital pathology have been made possible by the emergence of tissue slide image digitization, which has the potential to improve patient care greatly. We have assessed many AI-based computational models in this context intended for use in digital pathology applications. This paper covers the implications for enhancing patient outcomes, treatment planning, and diagnostic accuracy in the field of oncology. It also gives an overview of the most recent AI applications in oncology, emphasizing present difficulties and potential future developments.

Keywords: Artificial Intelligence; cancer imaging; machine learning; oncology; IHC

1. Introduction

Globally, 19.3 million new instances of cancer were detected in 2021, and the disease was responsible for 10 million deaths (1). Globally, colon, prostate, stomach, lung, and breast cancers were the most common types of cancer (2). Personalized cancer care is becoming increasingly popular; it emphasizes treatments specific to each patient's disease progression and is developed through cooperation between physicians and researchers (3). Conventional cancer diagnosis is based on manual techniques and eye observation, which leads to a laborious and error-prone histological diagnosis. The subjectivity and unpredictability in expert assessments in histopathology analysis highlight the necessity for early detection and precise diagnosis to improve treatment success. The use of artificial intelligence (AI) help systems to improve diagnostic accuracy is made possible by the global pathologist shortage. Digitization of histopathology slides and computer model analysis are two aspects of digital pathology. Prewitt and Mendelsohn created a technique in 1965 for scanning microscope images, turning optical data into a matrix of optical density values, and using that information to determine if different cell types are present (4). Introduced in 1956, artificial intelligence (AI) includes machine learning (ML) techniques, in which computers learn from data to recognize patterns and anticipate outcomes without the need for explicit programming (5). Deep learning (DL), a subset of machine learning (ML), was made possible by the creation of artificial neural networks in the 1980s (6). Due to advancements in computational processing capability, DL has attracted interest in recent decades for its applications in digital pathology and other domains (7–9).

Significance of Immunohistochemistry in Cancer Research

In cancer pathology and research, immunohistochemistry (IHC) is an essential technique for analyzing protein expression linked to treatment resistance, metastasis, and cancer development (10). Using IHC, tumor tissue is analyzed for particular protein markers that help with exact cancer categorization, prognosis assessment, and therapy planning. These indicators play a key role in locating cellular and molecular alterations connected to the onset of cancer. For instance, choosing the right course of treatment for breast cancer requires an IHC assessment of HER2/neu expression and hormone receptor status (11). Similarly, the use of immune checkpoint inhibitors in non-small cell lung cancer is dictated by assessment of programmed death-ligand 1 (PD-L1) expression (12).

PD-L1 is a biomarker that identifies patients more likely to respond to PD-1/PD-L1 inhibitors such as pembrolizumab. Validating its use implies that selecting patients based on their PD-L1 status is reliable for identifying those who may benefit most from this treatment. Additionally, IHC allows for the assessment of immune cell infiltration and the tumor microenvironment. In a recent study, it has been demonstrated that the presence and localization of CD8(+) T cells within the invasive tumor margin may serve as a valuable biomarker for assessing a patient's potential response to therapeutic interventions. The PD-1/PD-L1 immune inhibitory axis plays a crucial role in regulating the immune system's response to cancer cells. The fact that pre-existing CD8(+) T cells at the tumor margin are linked to this axis suggests a potential mechanism by which the immune response to cancer is modulated and predicts patient outcomes (13).

In colorectal cancer, the presence of the epidermal growth factor receptor (EGFR) on tumor cells, determined by IHC, helps predict responsiveness to anti-EGFR therapies (14). In the case of cervical cancer, p16INK4a immunohistochemistry is used to identify overexpression of this protein, which is indicative of high-grade squamous intraepithelial lesions (HSILs) and an increased risk of cervical cancer (15). In prostate cancer, the assessment of markers such as prostate-specific antigen (PSA), Ki-67, and p53 helps predict the risk of disease progression and guides treatment decisions (Shen et al., 2010). Moreover, immunohistochemistry is employed to detect specific molecular alterations that may increase cancer risk. For instance, the loss of DNA mismatch repair proteins, as identified through immunohistochemistry, is associated with Lynch syndrome and an elevated risk of colorectal and other cancers (16).

The synergy between machine learning and immunohistochemistry (IHC) in cancer risk stratification represents a powerful approach to enhancing understanding of cancer biology and improving patient outcomes. It enables the integration of complex data types, enhances predictive accuracy, and facilitates personalized risk assessment, ultimately leading to more effective cancer prevention and management strategies. Machine learning leverages the analysis of IHC data to identify complex patterns and relationships, ultimately aiding in more accurate cancer risk stratification. This helps to reduce data dimensionality, focuses on the most informative markers, and facilitates the classification of tissue samples into different risk categories. As new IHC data become available, models can be updated to incorporate the latest findings, ensuring that risk stratification remains up-to-date and accurate. Researchers continue to explore innovative applications and develop robust ML models to leverage the wealth of information embedded in IHC data. In this review, we provide some ML techniques that can be used in IHC for cancer-related applications.

Image Analysis: In order to quantify the expression levels of certain proteins or to detect patterns of protein expression linked to distinct cancer types or disease stages, machine learning algorithms can be trained to evaluate IHC images. Gurcan et al. [14] and Veta et al. [15] have investigated the emergence of feature engineering and learning techniques in image pathology, particularly for tasks like object identification, picture segmentation, and tissue categorization. Komura and Ishikawa [16] have investigated the application of machine learning models in digital pathology, which includes computer-assisted diagnosis, content-based picture retrieval, and clinicopathological correlations. Researchers have talked about the use of deep learning models, their limitations, and the prospects for the field of diagnostic breast pathology in this work [17]. Deep learning (DL) techniques have been successfully applied to numerous image analysis challenges [18–21].

Tumor Classification: IHC staining patterns can be used to train machine learning algorithms that categorize different cancer kinds or subtypes. This can help with cancer diagnosis and patient

outcome prediction. Numerous investigators have investigated an array of methods with the goal of attaining low-error and high-performance brain tumor identification and classification. Stacke et al. study's [22] demonstrated how to optimize for images that maximally activate the neurons in a convolutional neural network to assess the network's generalization in tumor classification using H&E stained images. The study involved analyzing the representations generated by the network and documenting the features to which the network responded. In order to improve the accuracy of brain tumor classification using MRI images, Tandel G. et al. [23] created five clinical multiclass datasets and used a Convolutional Neural Network (CNN) based on transfer learning. The usefulness of tumor classification techniques for classifying MR brain imaging characteristics into n/a, multifocal, multicentric, and gliomatosis was assessed in this work [24]. The statistical characteristics of the incoming photographs were analyzed as part of the classification process, and the data were methodically separated into several categories. These categorized data were then tested using ML techniques such as KNN (k closest neighbor), RF (random forest), SVM (support vector machines), and LDA (linear discriminant analysis).

Prognostic and Predictive Biomarker Discovery: ML can help identify new prognostic or predictive biomarkers by analyzing IHC data alongside clinical outcomes (17). These biomarkers can guide treatment decisions and predict patient responses. Biomarkers are recognized as molecular indicators that signify an increased likelihood of benefits or the potential for toxicity associated with a particular medicine (18). Additionally, they can be defined as measurement variables linked to the outcome of a disease (19). Prognostic biomarkers provide insight into the prognosis of cancer as well as the administration of therapy (20). Predictive biomarkers serve as indicators of the probability of a patient's response to a treatment plan. They enable the categorization of patients into groups with higher or lower chances of responding to a specific regimen, thereby enhancing therapeutic precision and treatment effectiveness (21). Machine learning methods utilize both prognostic and predictive biomarkers to assess performance, validate models, and present key results pertinent to the research. One such work is studied in this work (22), the authors endeavored to predict a biomarker panel for lung cancer based on autoantibodies. They employed recursive feature elimination with random forest modeling and utilized least absolute shrinkage and selection operation (LASSO) regression with repeated 10-fold cross-validation in their approach (23). Deep-learning techniques were employed to analyze scanned sections of H&E (hematoxylin-eosin)-stained tissue, aiming to develop a biomarker for predicting patient outcomes after primary colorectal cancer surgery (24). The approach involved the use of two convolutional neural networks: the first network delineated cancerous tissue, and the second categorized patients into distinct prognostic groups. Risk stratification was carried out using both invariable and multivariable analyses.

Spatial Analysis: ML methods can be used to analyze the spatial relationships between different cell types and protein expression patterns in IHC images. This can provide insights into the tumor microenvironment and its impact on cancer progression (25). This mini-review discusses recent progress in machine learning and artificial intelligence concerning the spatial analysis of the tumor immune microenvironment in pathology slides. The integration of machine learning (ML) and artificial intelligence (AI) algorithms with digital pathology is revolutionizing the histopathological analysis of the tumor immune microenvironment (TIME) in tumor samples (26). The spatial assays, such as immunohistochemistry (IHC), only permit targeted analyses involving a limited number of molecular markers.

Data Integration: Immunohistochemistry (IHC) data can be combined with other omics datasets, such as transcriptomics and genomes, using machine learning (ML) to gain a comprehensive understanding of cancer biology and identify possible therapeutic targets. Current artificial intelligence (AI) frameworks, which rely only on machine learning (ML) techniques, have been applied to the combination of omics and phenotypic data to identify novel biomarkers (27). Additionally, ML algorithms have been used extensively to classify cancer using a variety of data sources (28). Computed tomography (CT) data along with radiomics features were utilized in two different studies to classify cancer patients, enhancing the prediction accuracy for lung cancer and pulmonary lesions, respectively (29).

Quality Control: ML algorithms can help in quality control by detecting staining artifacts, tissue irregularities, or other issues in IHC images, ensuring the reliability of the results (30). In this article, computational pathology (CPATH) in diagnostic breast pathology and IHC were explored with the application of ML/AI-based tools, and the use of AI applications in diagnostic breast pathology (31).

In this section, we have offered ML techniques that could be applied to IHC data in cancer; however, to work involving IHC in cancer using ML methods, researchers would need access to a large dataset of IHC images, clinical information, and possibly other molecular data, as well as expertise in machine learning and image analysis.

Artificial Intelligence Examples in Oncology

Software developers and data analysts are leading the implementation of novel AI-based image analysis methods in pathology and oncology. They are creating and applying AI tools for several tasks, such as improving diagnostic accuracy and identifying new and more effective biomarker techniques for precision oncology. In order to accurately characterize whole-slide images of five different types of colorectal polyps, Korbar et al. created a deep-learning algorithm using a modified version of a residual network architecture (32). They generated a few deep-learning approaches using a set of data of 2074 crop images handwritten by several professional data pathologists as reference standards. Pathologists have studied hematoxylin-eosin (H&E) stained slides for more than a century. Bychkov et al. analyzed a collection of scanned tumor tissue microarray (TMA) samples stained with hematoxylin and eosin from 420 patients with colorectal cancer who had accessible clinicopathological characteristics and outcome information (33). Their findings demonstrate that, in the classification of individuals into low- and high-risk groups, deep learning-based prognosis prediction beats visual histological assessment carried out by human specialists on both TMA spot and whole-slide levels. Litjens et. al. made accessible a dataset for the CAMELYON16 and CAMELYON17 Grand Challenges that had 1,399 captioned whole-slide images (WSIs) of lymph nodes, both with and without metastases (34). In order to detect prostate cancer tissue in whole-slide photos, Tolkach et al. constructed DL-based models based on a sizable, high-quality training dataset and a cutting-edge convolutional network architecture (NASNetLarge) (35). In their study of deep learning approaches for prostate cancer, Wildeboer et al. contrasted the techniques used to provide the CAD outputs to the operator for additional medical decision-making (36). Jaroensri et al. created deep learning algorithms to conduct histologic assessment of all three elements using digitalized hematoxylin and eosin-stained slides carrying malignant breast cancer (37). Oncotype DX is one of the many genetic tests that have been developed to determine which early-stage patients might benefit from more intensive starting therapy (38).

Machine learning (ML) is a type of artificial intelligence characterized by computer algorithms that learn from data, specifically through the process of learning to map input data to output predictions. Machine learning (ML) tasks are commonly categorized as either supervised or unsupervised. In supervised learning, the model is trained using input data that is paired with corresponding known outcomes which is characterized as classification or regression. Unsupervised learning is advantageous in situations where a specific output is unknown, or when researchers aim to uncover novel patterns within the data. The basic principles of machine learning (ML) and its applications in cancer diagnosis, prognosis, and treatment are illustrated in Figure 1. The application of machine learning spans various tasks throughout the spectrum of oncology, including diagnosis, prognosis, and treatment.

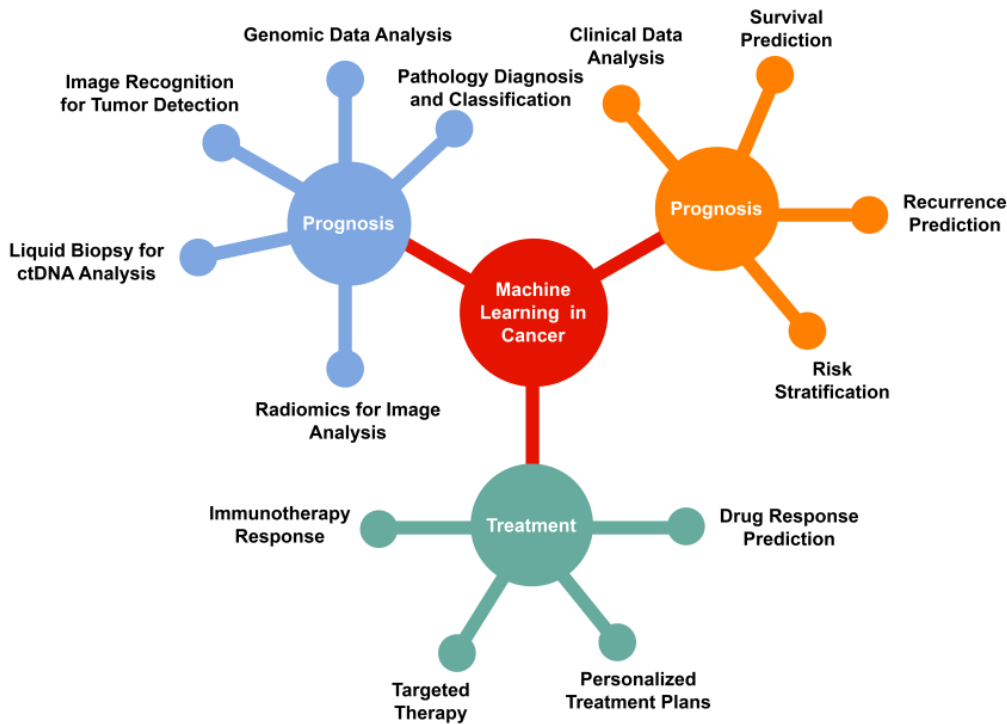


Figure 1. A graphic representation to effectively communicate the diverse applications of machine learning in cancer.

ML models analyze IHC data to aid in the early detection of cancerous lesions. Algorithms can learn patterns associated with malignancies, helping to improve accuracy and speed in diagnosing tumors.

We have compiled and summarized relevant published papers on machine learning algorithms in the domains of cancer diagnosis, prognosis, and treatment, incorporating studies focused on histopathology image analysis. Details are presented in Table 1.

Table 1. Overview papers on ML algorithms in diagnosis, prognosis, and treatment response in cancer.

Study	Study goal	Model	Type of cancer	Model Output
Ray A et al.	To evaluate the suitable classifier by comparing the performance of different ML algorithms for risk prediction and diagnosis of breast cancer	SVM (Support Vector Machine), K Nearest Neighbors (K-NN), Naive Bayes (NB), CART (Classification and Regression Tree)	Breast cancer	The SVM model achieved highest accuracy 98.14% with lower error rate.
Hollon T.C. et al.	Assessing brain tumor pathology slides in intra-surgery evaluation.	CNN	Brain cancer	Prospective clinical trials of stimulated Raman histology (SRH) images revealed higher diagnostic accuracy than conventional histologic images (94.6% vs 93.9%)
Che Y et al.	Automated predictive method for scoring whole-slide images	Deep learning network	Breast cancer	The automated scoring achieved an overall accuracy of 97.9% for slide-level classification. The proposed method

Zhou et al.	(WSI) of HER2 slides based on a deep-learning network. To identify of lymph node metastasis	CNN	Breast cancer	demonstrated excellent specificity, particularly for all IHC 0 and 3+ slides and the majority of 1+ and 2+ slides. Sensitivity/specificity of 85%/73% compared with trained radiologists (sensitivity/specificity 73%/63%)
Karla et al.	A specialized search engine in digital pathology symbolizing to histopathology images To predict resistance to hypomethylating agents in patients with myelodysplastic syndromes (MDS)	DenseNet	32 cancer types (TCGA)	The preliminary results of the TCGA archive demonstrate the feasibility of the technology and improve the accuracy and speed requirements of the Yottixel platform to make it more usable for diagnostics purposes.
Nazha et al.		Recommender system	Myelodysplastic syndrome	Molecular signatures were present in 30% of patients with three or more mutations/sample with an accuracy rate of 87% in the training cohort and 93% in the validation cohort.
Kharya S et al.	Cancer prognosis and prediction	SVM, artificial neural networks (ANN), Naive Bayes classifier, and AdaBoost tree	Breast cancer	Decision trees, regression trees, and so on, an artificial neural network (ANN) was found to be the most popular one.
Esteva A et al.	Classification of skin lesions	CNN	Skin cancer	The area under the curve (AUC) for each case is over 91% and d negligible changes in AUC (< 0.03).
Lou et al.	Identification of radiation sensitivity parameters to predict treatment failure	CNN	Lung cancer	Deep Profiler and clinical variables predicted treatment failures with a concordance index of 0.72

¹ Tables may have a footer.

The research employed machine learning techniques to classify cancer-related data and provide a diagnosis for breast cancer (49). In this study, various classification methods, including support vector machine classifiers, probabilistic neural networks, and K-nearest neighbors, were examined and applied to specific feature sets. Among these, the support vector machine classifier models showed the highest overall accuracy in diagnosing breast cancer. Hollon et al. (50). describe a concurrent workflow that combines stimulated Raman histology (SRH), a label-free optical imaging technique, with deep convolutional neural networks (CNNs). This integration facilitates the prediction of diagnoses at the bedside almost in real-time through an automated process. Another study (51) introduces an automated predictive approach for scoring whole-slide images (WSI) of the human epidermal growth factor receptor 2 (HER2) slides, using a deep-learning network. Zhou et al. detail the use of ultrasonography for the detection of biopsy-confirmed nodal metastasis in patients with breast tumors (52).

Their findings indicate sensitivity and specificity levels surpassing those achieved by human observers. In another study, established a diagnostic consensus across various cancer types by employing machine learning (ML) methodologies for the analysis of histopathology images (39). Initially, the authors utilized their previously developed image search engine, Yottixel, to index

approximately 30,000 whole-slide images (WSI) representing 32 cancer types from The Cancer Genome Atlas (TCGA) (40). In the context of next-generation sequencing, where conventional statistical methods may have limitations, a recommendation system inspired by those employed in online marketing has demonstrated the capability to forecast nonresponse to treatment in patients with myelodysplastic syndromes. In observational research, an assessment was conducted to compare the accuracy of the support vector machine, artificial neural networks (ANN), Naive Bayes classifier, and AdaBoost tree in order to identify an effective model for predicting breast cancer (41). In the study conducted by Esteva et al., machine learning techniques employed in skin cancer prognosis demonstrated comparable accuracy to that of a proficient dermatologist (42). Lou et al. studied the mainstay of treatment for patients with early-stage lung cancer and oligometastatic disease to the lung (43). AI has been investigated for enhancing the diagnosis of tuberculosis, meningitis, and sepsis, as well as predicting treatment challenges in patients with hepatitis C and B (44).

Machine learning models, particularly supervised algorithms such as support vector machines (SVMs) and deep neural networks, can recognize patterns in IHC data and classify tissue samples into different categories based on protein expression profiles (45).

Several case studies demonstrate the effectiveness of software tools in cancer risk stratification, helping clinicians make informed decisions about treatment options for cancer patients (46). For example, Oncotype DX is a genomic test used in breast cancer risk stratification. It analyzes the expression of 21 genes in tumor tissue to predict the likelihood of breast cancer recurrence and the potential benefit of chemotherapy (47). This study describes the effectiveness of Oncotype DX in guiding chemotherapy decisions for breast cancer patients. Cardoso et al. evaluated the clinical utility of the MammaPrint genomic test in guiding adjuvant chemotherapy decisions for early-stage breast cancer patients (48). It assesses the expression of 70 genes in tumor tissue to classify patients as either “low risk” or “high risk” for distant metastasis (49). Obtaining high-quality and sufficient quantities of data, including clinical, genetic, and lifestyle information, is essential for accurate risk stratification. Data for cancer risk stratification are collected through a combination of epidemiological studies, clinical trials, and genetic research. This includes information on demographics, lifestyle factors (e.g., smoking, diet, physical activity), and family history of cancer. Genetic studies involve the collection of DNA samples from individuals to identify genetic markers associated with cancer risk. The Cancer Genome Atlas (TCGA) project collected genomic and clinical data from various cancer types, facilitating the identification of cancer-related genetic alterations (50). Clinical trials collect data on cancer patients, including treatment outcomes, side effects, and genetic information. These data help in understanding how genetic factors influence treatment response. For instance, the TAILORx trial collected data on breast cancer patients to determine the optimal treatment strategy based on genetic testing of tumor tissue (47).

Challenges in AI Application

The effectiveness of any AI-based method is mainly determined by the volume and quality of the input information. In order to obtain the highest prediction performance, the data used to train an AI system must be clean, curated, have the highest signal-to-noise ratio, and be as accurate and thorough as feasible. For instance, the effectiveness of an AI system to segment a specific biological structure found in a WSI depends heavily on the accuracy of the reference annotations made by experienced pathologists in the learning set. The study of Doyle et al., who created an ML-based AI technique to automatically identify locations of prostate cancer in WSIs, highlights the significance of highly selected data (51). An AI approach might be used to locate specific spots on the slides for later imaging with superresolution techniques, enabling the scanning of important structures and locations in the image at a much greater resolution while also generating less digitally scanned data overall. Following this strategy, Kleppe et al. employed an ML-based algorithm to find that patients with chromatin characterized as homogenous had better survival outcomes than those with heterogeneous chromatin across several solid tumor types (52). Deep neural networks have been criticized for their lack of interpretability and for being less intuitive than hand-crafted networks,

despite their great accuracy and simplicity of use. This could be a barrier to their use in therapeutic settings. The goal of a few studies has been to give biological interpretability to DL tools using contemporary techniques, such as post hoc procedures or supervised ML models, to explain the results after the DL model has already made its prediction (53,54). Because pathologists and oncologists spend much time developing these methods, engineering hand-crafted features is frequently difficult and time-consuming. Fusion techniques that integrate DL and custom strategies have begun to gain popularity in recent years. These tactics might rely on handcrafted ML approaches for prediction after using DL algorithms for the initial detection of cells or elements, thereby utilizing domain expertise to guarantee the approach's biological interpretability. The generalizability of the methods must be ensured using multi-institutional data before AI- and ML-based solutions are adopted in clinical settings. Training and validation sets are frequently created using the data that are available for developing an AI strategy. Examples from the categories of interest are typically evenly represented in the initial dataset, which is also known as a training, learning, or discovery set. Once the model has been trained and finalized on the learning set, it is typically validated without further optimization on a test set. This test set is either derived from the original set of cases or obtained from a different institution. Digital pathology-based companion diagnostic tests for oncologists may offer extra helpful data for disease risk classification and patient selection for specialized treatments. With a turnaround time of roughly two weeks, genomic companion diagnostic tests require delivering tissue to a central site. AI-based technologies will mostly be required for pathologists to identify structures or particular regions of interest in digitized WSIs. To enable short turnaround times and the capacity to control the clinical process, the digital slide scanner has thus become a crucial component in the development of such systems. Despite these difficulties and limitations, the future of AI techniques for digital pathology is encouraging. Many organizations worldwide have made the decision to digitize their whole pathology practice in recent years.

Future Prospective

The advent of whole-slide digital scanning and the parallel expansion of deep learning (DL)-based neural networks for analyzing digital images of slides have propelled a surge in the popularity of digital pathology technologies. Despite the challenges and uncertainties related to regulatory measures, reimbursement, and implementation, there is growing interest within the pathology and cancer communities in the development and application of these technologies. Recent innovations include open-top light sheet microscopy, which generates 3D images of tissue samples without the need for destructive sectioning or slide preparation. To analyze and interpret these vast amounts of data, pathologists and oncologists will require the assistance of artificial intelligence (AI) methods.

Conclusion

In the foreseeable future, there is potential for the development of AI algorithms that not only support pathologists but also aim to alleviate the burden of mundane, repetitive tasks to enhance diagnostic accuracy and grading. With the advent of the new era of deep learning-assisted pathology, data banking, integration, and cloud laboratories are emerging as essential elements of daily pathology practice. Furthermore, pathologists, data scientists, and enterprises are increasingly collaborating to amalgamate genomics, proteomics, bioinformatics, and computational algorithms into vast and complex clinical datasets.

This convergence of disciplines is paving the way for computational pathology to provide critical insights into disease diagnosis, prognosis, and treatment. Through this approach, computational pathology has the potential to revolutionize the field by offering a comprehensive understanding of diseases at a molecular level, thereby enabling personalized and precision medicine.

However, it is important to acknowledge that numerous technological and ethical challenges must be addressed to fully realize the potential of computational pathology. These challenges include ensuring data privacy and security, standardizing data formats, and developing robust and interpretable AI models. As a synergistic system, computational pathology promises to enhance

workflow efficiency, enabling clinical teams to share and analyze imaging data across a broader platform, thus fostering collaboration and innovation.

In this context, we critically assess the current state of research in computational pathology and outline the direction for future developments. It is imperative to address the obstacles to the widespread integration of AI in cancer care to unlock the full potential of these technologies in improving patient outcomes. In summary, this research highlights the importance of embracing the opportunities and challenges presented by AI and computational pathology to advance the field and ultimately enhance patient care in oncology.

Author Contributions: Conceptualization, Literature review, Z.R., A.A. and A.U.; formal analysis and investigation, Z.R., Y.B., A.M., A.A., B.M., writing—review and editing Z.R., A.A., Y.B., A.M. and A.U. All authors have read and agreed to the published version of the manuscript.

Funding: The study was funded by the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan grant AP19679717 for Z.R., Y.B., A.A., A.M. and A.U. The funders had no role in study design, analysis, decision to publish, or preparation of the manuscript. All authors had full access to all the data in the study, and the lead authors (Z.R., Y.B., A.M., A.U.) had final responsibility for the decision to submit manuscript for publication.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We thank all the authors and Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan for providing the grant that made this research possible.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin.* 2021;71(3):209–49.
2. Mattiuzzi C, Lippi G. Current Cancer Epidemiology. *J Epidemiol Glob Health.* 2019 Dec;9(4):217–22.
3. Lopes-Júnior LC, Veronez LC. Personalized Care for Patients with Cancer in the Precision-Medicine Era. *Int J Environ Res Public Health.* 2023 Feb 9;20(4):3023.
4. Prewitt JM, Mendelsohn ML. The analysis of cell images. *Ann N Y Acad Sci.* 1966 Jan 31;128(3):1035–53.
5. McCarthy J, Minsky ML, Rochester N, Shannon CE. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. *AI Mag.* 2006 Dec 15;27(4):12–12.
6. Yao X. Evolving artificial neural networks. *Proc IEEE.* 1999 Sep;87(9):1423–47.
7. Arbib MA. *The Handbook of Brain Theory and Neural Networks.* MIT Press; 2003. 1328 p.
8. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. In: *Advances in Neural Information Processing Systems [Internet].* Curran Associates, Inc.; 2012 [cited 2023 Oct 23]. Available from: https://proceedings.neurips.cc/paper_files/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html
9. Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. *Science.* 2006 Jul 28;313(5786):504–7.
10. Taylor CR, Levenson RM. Quantification of immunohistochemistry--issues concerning methods, utility and semiquantitative assessment II. *Histopathology.* 2006 Oct;49(4):411–24.
11. Wolff AC, Hammond MEH, Hicks DG, Dowsett M, McShane LM, Allison KH, et al. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. *J Clin Oncol Off J Am Soc Clin Oncol.* 2013 Nov 1;31(31):3997–4013.
12. Herbst RS, Soria JC, Kowanetz M, Fine GD, Hamid O, Gordon MS, et al. Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients. *Nature.* 2014 Nov 27;515(7528):563–7.
13. Tumei PC, Harview CL, Yearley JH, Shintaku IP, Taylor EJM, Robert L, et al. PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature.* 2014 Nov 27;515(7528):568–71.
14. Van Cutsem E, Köhne CH, Láng I, Folprecht G, Nowacki MP, Cascinu S, et al. Cetuximab plus irinotecan, fluorouracil, and leucovorin as first-line treatment for metastatic colorectal cancer: updated analysis of

- overall survival according to tumor KRAS and BRAF mutation status. *J Clin Oncol Off J Am Soc Clin Oncol*. 2011 May 20;29(15):2011–9.
15. Münger K, Baldwin A, Edwards KM, Hayakawa H, Nguyen CL, Owens M, et al. Mechanisms of human papillomavirus-induced oncogenesis. *J Virol*. 2004 Nov;78(21):11451–60.
 16. Kastrinos F, Syngal S. Inherited Colorectal Cancer Syndromes. *Cancer J Sudbury Mass*. 2011 Nov;17(6):405–15.
 17. Al-Tashi Q, Saad MB, Muneer A, Qureshi R, Mirjalili S, Sheshadri A, et al. Machine Learning Models for the Identification of Prognostic and Predictive Cancer Biomarkers: A Systematic Review. *Int J Mol Sci*. 2023 Apr 24;24(9):7781.
 18. Ahmad A, Imran M, Ahsan H. Biomarkers as Biomedical Bioindicators: Approaches and Techniques for the Detection, Analysis, and Validation of Novel Biomarkers of Diseases. *Pharmaceutics*. 2023 May 31;15(6):1630.
 19. Califf RM. Biomarker definitions and their applications. *Exp Biol Med*. 2018 Feb;243(3):213–21.
 20. Das S, Dey MK, Devireddy R, Gartia MR. Biomarkers in Cancer Detection, Diagnosis, and Prognosis. *Sensors*. 2023 Dec 20;24(1):37.
 21. Meehan J, Gray M, Martínez-Pérez C, Kay C, Pang LY, Fraser JA, et al. Precision Medicine and the Role of Biomarkers of Radiotherapy Response in Breast Cancer. *Front Oncol*. 2020 Apr 24;10:628.
 22. Li MX, Sun XM, Cheng WG, Ruan HJ, Liu K, Chen P, et al. Using a machine learning approach to identify key prognostic molecules for esophageal squamous cell carcinoma. *BMC Cancer*. 2021 Aug 9;21(1):906.
 23. Awad M, Fraihat S. Recursive Feature Elimination with Cross-Validation with Decision Tree: Feature Selection Method for Machine Learning-Based Intrusion Detection Systems. *J Sens Actuator Netw*. 2023 Oct;12(5):67.
 24. Li F, Yang Y, Wei Y, He P, Chen J, Zheng Z, et al. Deep learning-based predictive biomarker of pathological complete response to neoadjuvant chemotherapy from histological images in breast cancer. *J Transl Med*. 2021 Aug 16;19:348.
 25. Lara H, Li Z, Abels E, Aeffner F, Bui MM, ElGabry EA, et al. Quantitative Image Analysis for Tissue Biomarker Use: A White Paper From the Digital Pathology Association. *Appl Immunohistochem Mol Morphol*. 2021 Aug;29(7):479–93.
 26. Xu H, Cong F, Hwang TH. Machine Learning and Artificial Intelligence-driven Spatial Analysis of the Tumor Immune Microenvironment in Pathology Slides. *Eur Urol Focus*. 2021 Jul 1;7(4):706–9.
 27. Agarwal N, Nupur, Paul PK, Mishra SK. Artificial Intelligence and Machine Learning for Analysis of Multi-omics. In: Mani I, Singh V, editors. *Multi-Omics Analysis of the Human Microbiome: From Technology to Clinical Applications* [Internet]. Singapore: Springer Nature; 2024 [cited 2024 May 30]. p. 339–54. Available from: https://doi.org/10.1007/978-981-97-1844-3_16
 28. Kourou K, Exarchos KP, Papaloukas C, Sakaloglou P, Exarchos T, Fotiadis DI. Applied machine learning in cancer research: A systematic review for patient diagnosis, classification and prognosis. *Comput Struct Biotechnol J*. 2021 Oct 6;19:5546–55.
 29. Pan F, Feng L, Liu B, Hu Y, Wang Q. Application of radiomics in diagnosis and treatment of lung cancer. *Front Pharmacol*. 2023 Nov 1;14:1295511.
 30. Hoyt CC. Multiplex Immunofluorescence and Multispectral Imaging: Forming the Basis of a Clinical Test Platform for Immuno-Oncology. *Front Mol Biosci*. 2021 Jun 2;8:674747.
 31. Rakha EA, Vougas K, Tan PH. Digital Technology in Diagnostic Breast Pathology and Immunohistochemistry. *Pathobiology*. 2021 Dec 30;89(5):334–42.
 32. Korbar B, Olofson AM, Miraflor AP, Nicka CM, Suriawinata MA, Torresani L, et al. Deep Learning for Classification of Colorectal Polyps on Whole-slide Images. *J Pathol Inform*. 2017 Jan 1;8(1):30.
 33. Bychkov D, Linder N, Turkki R, Nordling S, Kovanen PE, Verrill C, et al. Deep learning based tissue analysis predicts outcome in colorectal cancer. *Sci Rep*. 2018 Feb 21;8(1):3395.
 34. Litjens G, Bandi P, Ehteshami Bejnordi B, Geessink O, Balkenhol M, Bult P, et al. 1399 H&E-stained sentinel lymph node sections of breast cancer patients: the CAMELYON dataset. *GigaScience*. 2018 Jun 1;7(6):giy065.
 35. Tolkach Y, Dohmgorgen T, Toma M, Kristiansen G. High-accuracy prostate cancer pathology using deep learning. *Nat Mach Intell*. 2020 Jul;2(7):411–8.
 36. Wildeboer RR, van Sloun RJG, Wijkstra H, Mischi M. Artificial intelligence in multiparametric prostate cancer imaging with focus on deep-learning methods. *Comput Methods Programs Biomed*. 2020 Jun 1;189:105316.
 37. Jaroensri R, Wulczyn E, Hegde N, Brown T, Flament-Auvigne I, Tan F, et al. Deep learning models for histologic grading of breast cancer and association with disease prognosis. *Npj Breast Cancer*. 2022 Oct 4;8(1):1–12.
 38. Quantitative nuclear histomorphometry predicts oncotype DX risk categories for early stage ER+ breast cancer | BMC Cancer | Full Text [Internet]. [cited 2023 Oct 24]. Available from: <https://bmccancer.biomedcentral.com/articles/10.1186/s12885-018-4448-9>

39. Lee K, Lockhart JH, Xie M, Chaudhary R, Slebos RJC, Flores ER, et al. Deep Learning of Histopathology Images at the Single Cell Level. *Front Artif Intell*. 2021 Sep 10;4:754641.
40. Kalra S, Tizhoosh HR, Choi C, Shah S, Diamandis P, Campbell CJV, et al. Yottixel – An Image Search Engine for Large Archives of Histopathology Whole Slide Images. *Med Image Anal*. 2020 Oct 1;65:101757.
41. Sukmandhani AA, Lukas L, Heryadi Y, Suparta W, Wibowo A. Classification Algorithm Analysis for Breast Cancer. *E3S Web Conf*. 2023;388:02012.
42. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017 Feb;542(7639):115–8.
43. Sundahl N, Lievens Y. Radiotherapy for oligometastatic non-small cell lung cancer: a narrative review. *Transl Lung Cancer Res*. 2021 Jul;10(7):3420–31.
44. Zhang B, Shi H, Wang H. Machine Learning and AI in Cancer Prognosis, Prediction, and Treatment Selection: A Critical Approach. *J Multidiscip Healthc*. 2023 Jun 26;16:1779–91.
45. Ertosun MG, Rubin DL. Automated Grading of Gliomas using Deep Learning in Digital Pathology Images: A modular approach with ensemble of convolutional neural networks. *AMIA Annu Symp Proc AMIA Symp*. 2015;2015:1899–908.
46. Liao J, Li X, Gan Y, Han S, Rong P, Wang W, et al. Artificial intelligence assists precision medicine in cancer treatment. *Front Oncol*. 2023 Jan 4;12:998222.
47. Sparano JA, Gray RJ, Makower DF, Pritchard KI, Albain KS, Hayes DF, et al. Adjuvant Chemotherapy Guided by a 21-Gene Expression Assay in Breast Cancer. *N Engl J Med*. 2018 Jul 12;379(2):111–21.
48. Pauls M, Chia S. Clinical Utility of Genomic Assay in Node-Positive Early-Stage Breast Cancer. *Curr Oncol*. 2022 Jul 20;29(7):5139–49.
49. Mook S, Schmidt MK, Weigelt B, Kreike B, Eekhout I, van de Vijver MJ, et al. The 70-gene prognosis signature predicts early metastasis in breast cancer patients between 55 and 70 years of age. *Ann Oncol*. 2010 Apr 1;21(4):717–22.
50. Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol*. 2015;19(1A):A68.
51. Doyle S, Feldman M, Tomaszewski J, Madabhushi A. A boosted Bayesian multiresolution classifier for prostate cancer detection from digitized needle biopsies. *IEEE Trans Biomed Eng*. 2012 May;59(5):1205–18.
52. Chromatin organisation and cancer prognosis: a pan-cancer study - The Lancet Oncology [Internet]. [cited 2023 Oct 24]. Available from: [https://www.thelancet.com/journals/lanonc/article/PIIS1470-2045\(17\)30899-9/fulltext](https://www.thelancet.com/journals/lanonc/article/PIIS1470-2045(17)30899-9/fulltext)
53. Liu Y, Kohlberger T, Norouzi M, Dahl GE, Smith JL, Mohtashamian A, et al. Artificial Intelligence-Based Breast Cancer Nodal Metastasis Detection: Insights Into the Black Box for Pathologists. *Arch Pathol Lab Med*. 2019 Jul;143(7):859–68.
54. Impact of Deep Learning Assistance on the Histopathologic Review of Lymph Nodes for Metastatic Breast Cancer - PubMed [Internet]. [cited 2023 Oct 24]. Available from: <https://pubmed.ncbi.nlm.nih.gov/30312179/>

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.