



Digital pathology, deep learning, and cancer: a narrative review

Darnell K. Adrian Williams Jr¹, Gillian Graifman², Nowair Hussain³, Maytal Amiel², Priscilla Tran², Arjun Reddy⁴, Ali Haider⁵, Bali Kumar Kavitesh⁶, Austin Li², Leael Alishahian², Nichelle Perera², Corey Efros², Myoungmee Babu⁷, Mathew Tharakan⁸, Mill Etienne⁹, Benson A. Babu^{2,10^}

¹Medical Scientist Training Program, Albert Einstein College of Medicine, Bronx, NY, USA; ²New York Medical College, Valhalla, NY, USA; ³Department of Internal Medicine, Overlook Medical Center, Summit, NJ, USA; ⁴Applied Mathematics & Statistics Stony Brook University, Stony Brook, NY, USA; ⁵Department of Artificial Intelligence, Yeshiva University, New York, NY, USA; ⁶Centre for Frontier AI Research (CFAR), Agency for Science, Technology, and Research (A*STAR), Singapore, Singapore; ⁷Artificial Intelligence and Mathematics, New York City Department of Education, New York, NY, USA; ⁸Stonybrook Medical Center, Stony Brook, NY, USA; ⁹Department of Neurology, New York Medical College, Valhalla, NY, USA; ¹⁰Department of Hospital Medicine, Wyckoff, Medical Center, New York, NY, USA

Contributions: (I) Conception and design: BA Babu; (II) Administrative support: M Babu, M Etienne, M Tharakan; (III) Provision of study materials or patients: None; (IV) Collection and assembly of data: All authors; (V) Data analysis and interpretation: DKA Williams Jr, A Haider, A Reddy, BK Kavitesh, BA Babu; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Benson A. Babu, MD/MBA. New York Medical College, 40 Sunshine Cottage Rd, Valhalla, NY 10595, USA; Department of Hospital Medicine, Wyckoff Medical Center, 374 Stockholm St., Brooklyn, New York, NY 11237, USA. Email: zero-click@zero-click.io.

Background and Objective: Cancer is a leading cause of morbidity and mortality worldwide. The emergence of digital pathology and deep learning technologies signifies a transformative era in healthcare. These technologies can enhance cancer detection, streamline operations, and bolster patient care. A substantial gap exists between the development phase of deep learning models in controlled laboratory environments and their translations into clinical practice. This narrative review evaluates the current landscape of deep learning and digital pathology, analyzing the factors influencing model development and implementation into clinical practice.

Methods: We searched multiple databases, including Web of Science, Arxiv, MedRxiv, BioRxiv, Embase, PubMed, DBLP, Google Scholar, IEEE Xplore, Semantic Scholar, and Cochrane, targeting articles on whole slide imaging and deep learning published from 2014 and 2023. Out of 776 articles identified based on inclusion criteria, we selected 36 papers for the analysis.

Key Content and Findings: Most articles in this review focus on the in-laboratory phase of deep learning model development, a critical stage in the deep learning lifecycle. Challenges arise during model development and their integration into clinical practice. Notably, lab performance metrics may not always match real-world clinical outcomes. As technology advances and regulations evolve, we expect more clinical trials to bridge this performance gap and validate deep learning models' effectiveness in clinical care. High clinical accuracy is vital for informed decision-making throughout a patient's cancer care.

Conclusions: Deep learning technology can enhance cancer detection, clinical workflows, and patient care. Challenges may arise during model development. The deep learning lifecycle involves data preprocessing, model development, and clinical implementation. Achieving health equity requires including diverse patient groups and eliminating bias during implementation. While model development is integral, most articles focus on the pre-deployment phase. Future longitudinal studies are crucial for validating models in real-world settings post-deployment. A collaborative approach among computational pathologists, technologists, industry, and healthcare providers is essential for driving adoption in clinical settings.

Keywords: Artificial intelligence (AI); deep learning (DL); digital pathology (DP); computational pathology; cancer

[^] ORCID: 0000-0003-1015-2393.

Submitted Jun 05, 2023. Accepted for publication Mar 24, 2024. Published online May 22, 2024.

doi: 10.21037/tcr-23-964

View this article at: <https://dx.doi.org/10.21037/tcr-23-964>

Introduction

Cancer remains one of the leading causes of morbidity and mortality worldwide (1,2). In 2019, the World Health Organization estimated that cancer is the leading cause of death globally. According to the National Center for Health Statistics, there will be 1,958,310 newly diagnosed cancer cases in 2023 (3). The American Cancer Society projects 609,820 cancer deaths in 2023 (3).

Screening and treatment have improved but remain a considerable primary public health concern. Public health systems worldwide spend \$200 billion on cancer-related costs (4). By 2030, there will be 19.3 million new cancer cases and 10 million cancer deaths (5). As these numbers continue to rise, there is a need for enhanced efforts to diagnose and treat cancer for all patient populations. To ensure equitable access to care for all patient populations, addressing and overcoming disparities and biases in medical care is essential, thus providing patients with the care they deserve (6). Four principles underlie health equity: equal access, equal utilization of resources, health equality, and distribution according to need (7). Health disparities can be reduced by addressing multi-level structural determinants and ensuring access to disadvantaged populations (8). The Joint Commission and the Institute for Healthcare Improvement oversee initiatives and prioritize inclusion and fairness in clinical practice (9,10).

Cutting-edge healthcare innovations that transform glass slides into digitized formats, when coupled with artificial intelligence (AI) and telecommunication systems, have the potential to ensure equitable access to cancer care, particularly in regions facing shortages of specialists. Digital pathology (DP) originated in the late 1960s, with telepathology (TP), a branch of DP, as one of its first uses. TP involves sending digital images over a secure long-range network. In 1986, Ronald Weinstein, MD, coined the term “telepathology” (11). Since then, validation studies have shown DP and conventional light microscopy to be highly accurate (11-15). Diagnosis does not require an onsite pathologist. TP has also eradicated the need for long-distance travel for specialized care, effectively addressing pathologist shortages and resource limitations within these networks. This innovation promotes equity in healthcare by

making specialized care more accessible. TP also facilitates international expert consultations, enhances provider communication, and ensures seamless handoffs. Since the 1960s, the evolution of digital images has been remarkable, progressing from expensive devices to sophisticated robotic microscopes and autonomous robotic whole slide image (WSI) intelligent scanners with integrated storage and retrieval systems, significantly enhancing workflow efficiency (16). The appeal of digitizing slides with WSI technology lies in its convenience, portability, and the power to manipulate and analyze pixels. Furthermore, WSI systems are utilized for virtual education and novel research and serve as indispensable tools for primary cancer diagnosis (17).

In the era of precision medicine, integrating deep learning (DL) with these technologies is crucial for enhancing health equity and advancing the field. This can be achieved by developing DL models to identify algorithmic biases in cancer care (18). Furthermore, digital slides have been successfully utilized in pathology, leveraging computational pathology techniques to deliver faster and more precise outcomes. (19-21). DL algorithms in DP can help pathologists identify complex patterns, classify image features, and provide quantitative assessment and predictive analytics. By harnessing advanced digital image management systems using cloud-based technology, healthcare organizations can facilitate faster, quality workflows (16,17). As this technology progresses, the anticipated global revenue for DP is projected to soar to a remarkable \$2,045.9 million by 2029, boasting a robust Compound Annual Growth Rate of 12.6% (22). Amidst this rapid expansion, a significant opportunity emerges to enhance clinical outcomes in precision oncology. This article offers an overview of the literature surrounding DP and DL models in oncology. The DL development lifecycle encompasses critical stages, including data preprocessing, model development, deployment, and continuous management within clinical practice. Additionally, we delve into the factors that influence model development. We present this article in accordance with the Narrative Review reporting checklist (available at <https://tcr.amegroups.com/article/view/10.21037/tcr-23-964/rc>).

Table 1 The search strategy summary

Items	Specification
Date of search	The first search was conducted on 9/18/2022. The last search was conducted on 5/18/2023
Databases and other sources searched	The databases searched were Embase, PubMed, DBLP, Google Scholar, IEEE Xplore, Cochrane database, Web of Science, Arxiv, BioRxiv, MedRxiv, and Semantic Scholar
Search terms used	(I) Boolean logic: connecting words like “AND”, “OR”, and “NOT” in various combinations to expand or narrow down search results (II) Fuzzy logic: search terms like “Digital Pathology” NEAR “Deep Learning” or “Whole Slide Imaging” WITHIN 5 words of “Deep Learning” to search for particular articles (III) Truncation: we searched for terms that began with a specific string by placing an asterisk (*) at the end of a root word. The word string used in the search was digital pathology, deep learning pathology, cancer, digital AI pathology, artificial intelligence, and cancer
Timeframe	2014–2023
Inclusion and exclusion criteria	Inclusion criteria: age greater than 18 year old; DL novel models; peer & non-peer reviewed publications; cancer Exclusion criteria: pediatric population; cytopathology; stereology; non-cancer; review articles
Selection process	Using the search criteria, 776 articles were found. Two independent reviewers (B.A.B.) and (A.R.) selected articles that fit the inclusion criteria. Thirty-six papers met the study eligibility criteria out of 776. A third reviewer (B.K.K.) served as a tiebreaker to resolve disagreements

AI, artificial intelligence; DL, deep learning.

Methods

This study follows the PICO Framework. Problem: human samples for cancer diagnosis. Intervention: WSI and DL. Comparison (Evaluation): model evaluation. It is not possible to compare each model directly due to the differences in data sets, algorithms, and output metrics used in each study—outcomes: model performance metrics such as accuracy, F1 score, and others to measure model performance.

We searched Embase, PubMed, DBLP, Google Scholar, IEEE Xplore, Cochrane database, Web of Science, ArXiv, BioRxiv, MedRxiv, and Semantic Scholar.

Articles were published between 2014 and 2023, using terms (I) Boolean Logic: connecting words like “AND”, “OR”, and “NOT” in various combinations to expand or narrow down search results; (II) Fuzzy Logic: search terms like “Digital Pathology” NEAR “Deep Learning” or “Whole Slide Imaging” WITHIN five words of “Deep Learning” to search for particular articles; (III) truncation: we searched for terms that began with a specific string by placing an asterisk (*) at the end of a root word. The word string used in the search were digital pathology, deep learning pathology, cancer, digital AI pathology, artificial intelligence, and cancer.

Studies on cancer detection utilizing WSI and DL

were considered eligible (*Table 1*). Employing the specified search criteria, a total of 776 articles were identified. Two independent reviewers, B.A.B. and A.R., screened the articles to determine eligibility based on the inclusion criteria. Out of the initial pool, 36 papers were deemed eligible for inclusion in the study. In cases of disagreement, a third reviewer, B.K.K., served as a tiebreaker to reach a consensus. *Figure 1* shows the flow chart of studies using PRISMA guidelines.

Discussion

Model types

The upcoming section offers a categoric overview of the DL models used in the literature. This will include exploring the various types of DL methods and analyzing their model development processes. By examining these aspects, we aim to provide a thorough understanding of the capabilities of these DL models in cancer detection. The selected literature is denoted in *Table 2*.

High-resolution slide images are gigapixels in size and require more memory and computational resources, so they cannot be incorporated directly into deep-learning models. Splitting high-resolution slide images into patches and developing patch aggregation strategies can help deep-

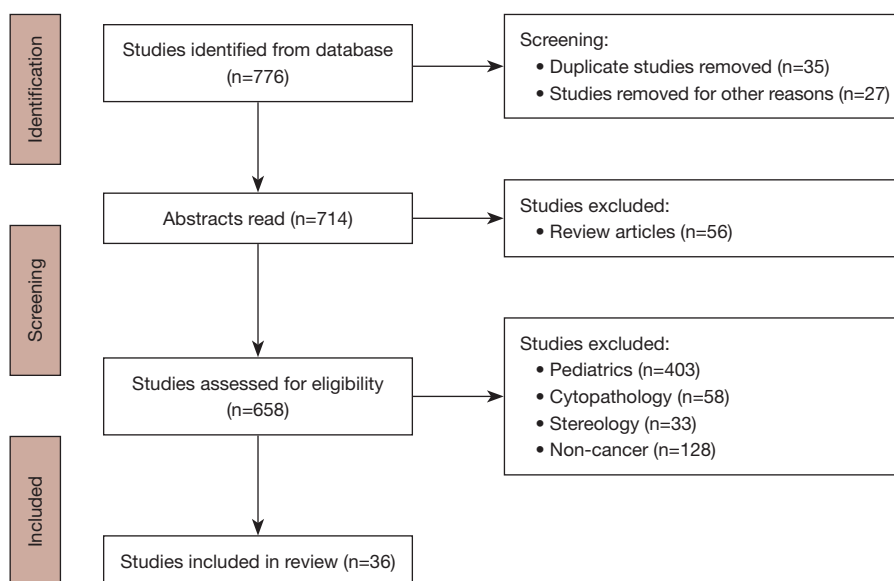


Figure 1 Flow chart of studies using PRISMA guidelines.

learning models analyze large image datasets efficiently.

Slides can be divided into patches to train DL models (30). Models can be taught to recognize patch features and patterns. After training, models can predict labels for new patches or entire images and reconstruct the source image using patch aggregation strategies. Patch-based aggregation optimizes memory usage and DL efficiency.

Patch-based aggregation methods

Patch-based aggregation in machine learning is often used in image processing and computer vision tasks. It involves dividing an image into smaller, fixed-size pieces or “patches”. Each of these patches is then analyzed and processed independently (50,58). This approach allows for a more detailed and localized understanding of the image data; learning about each patch may capture unique features, whereas the patches themselves are part of the bigger picture and may or may not contain the region of interest for tumor analysis. Patch-based aggregation is beneficial in tasks like image classification, object detection, and texture analysis. For instance, histopathology imaging can help identify specific features in WSI, such as tumors or other abnormalities (29). The algorithm can make more accurate predictions or identifications by focusing on small image areas at a time. Moreover, this method can be combined with various machine learning models, such as convolutional

neural networks (CNN). In such combinations, each patch is fed into the network, which learns to identify patterns or features within these smaller segments of the image. This can lead to more nuanced and detailed image analysis than processing the entire image (26,35).

Multiple instance learning (MIL)

During training, bags are labeled in MIL. Each bag has multiple instances, but only the bag label is known (25,26,30,35,36,57,59). A bag label is valid when the precise location of each instance is not essential, but the overall composition is. Bags are treated as single entities without explicitly modeling individual instances. For optimal performance, MIL requires large amounts of data; a minimum of 10,000 slides is recommended (30,57). MILs require many data points to capture patterns. Spatial relationships between instances are another MIL limitation.

Cancer classification using WSIs poses challenges because of their large pixel size and the annotation limitations when used on CNN (26). Several techniques have been developed to overcome this limitation. An end-to-end part learning-based approach can learn diverse and discriminative features to predict prostate and basal cell carcinomas (26). Furthermore, it defines multi-label lung cancer architectural subtypes for clinical decision support (26). Similarly, Cluster-to-Conquer overcomes the computational and algorithmic challenges posed by

Table 2 The information of the included 36 articles

Study	Model	Performance metrics	Training data	Problem type	Cancer type	Staining type
Wang <i>et al.</i> (23)	Patch-based with various DL models	Two evaluation metrics—lesion-based (based on correctly identifying cancer cells in WSI): 0.7051; slide-based: 0.925 ROC	CAMELYON-16 dataset: 400	Classification	Breast	Pan cytokeratin
Wang <i>et al.</i> (24)	Patch-based paired with FCN-RF (context-aware block selection)	M4-CNN weighted	SUCC dataset: training, 355; testing, 1,402	Classification	Lung	H&E
Chikontwe <i>et al.</i> (25)	Spatial contextual information MIL patch; aggregation; center embedding	Norm3-based RF accuracy: 0.973 F1: 92.36%; precision: 92.54%; recall: 92.31%; accuracy: 92.31%	173 WSI	Classification	Colon	H&E
Xie <i>et al.</i> (26)	MIL patch-based end-to-end (learning diverse discriminative)	Accuracy: 0.986	BCC dataset: training, 6,900; valid, 1,487; test, 1,575 Prostate needle biopsies dataset: training, 8,521; valid, 1,827; test, 1,811	Classification	Prostate	H&E
Li <i>et al.</i> (27)	2 stage patch based MIMS CNN scale invariant	Accuracy: 0.986	339 3D OCT: training, 67%; test, 33%	Classification	Colorectal	H&E
Tsuneki and Kanavati (28)	Patch aggregation weakly supervise EfficientNet B1	Best model TL-colon poorly ADC-2 (×20, 512) Accuracy: TURP (Hospital A-B: 0.916; hospital A: 0.969; hospital B: 0.874) TCGA: 0.821 (Hospital A-C: 0.844)	Training: hospital A, 281; hospital B, 739 Validation on: hospital A and B, 20 Test: hospitals A and B, 500	Classification	Stomach, colon, lung, and breast	H&E
Kanavati <i>et al.</i> (29)	Patch aggregation RNN/CNN EfficientNet B1	DCIS: test set1: WSI + RNN: ROC AUC: 0.937 IDC: test set1: WSI + RNN: ROC AUC: 0.977 DCIS: test set2: WSI + RNN: ROC AUC: 0.960 IDC: test set2: WSI + RNN: ROC AUC: 0.959	Total 3,672 WSI, total biopsy 2,101, total surgery 1,571 Test: 1,930, 1,065, 865; train: 1,652, 978, 674; valid: 90, 58, 32	Classification	Breast	H&E
Campanella <i>et al.</i> (30)	MIL-RNN	AUC 0.98	Skin, prostate, and breast slides: 44,732	Classification	Prostate, basal cell carcinoma, and breast	H&E
Shao <i>et al.</i> (31)	TransMIL	TCGA NSCLC: AUC 96.03%; TCGA RCC: 98.82%	993 NSCLC; 884 RCC	Classification	Kidney, breast, and lung	H&E
Zhang <i>et al.</i> (32)	DTFD-MIL	CAMELYON-16 accuracy 90.8%; F1 88.2%; AUC 94.6% TCGA accuracy 89.4%; F1 89.1%; AUC 96.1%	CAMELYON-16 patches: 3.7 million TCGA patches: 8.3 million	Classification	Lung	H&E
Wang <i>et al.</i> (33)	Second order MIL	AUC 96%	130 WSI	Classification	Breast	H&E
Hashimoto <i>et al.</i> (34)	Domain adversarial MIL	ACC 87%; precision 97%; recall 81%	Malignant lymphoma slides: 196	Classification	Lymphoma	H&E
Lu <i>et al.</i> (35)	Cluster constraint attention multiple instance learning	AUC: >0.95	Variable sizes used: 884, 1,967, 899	Classification	Lung	H&E
Sharman <i>et al.</i> (36)	Cluster-2-conquer	C2C (WSI + Patch + KLD loss) Accuracy: 86.2%	Gastrointestinal dataset 413 high-resolution images Train: 65%; valid: 15%; test: 20%	Classification	Gastrointestinal	H&E
Guan <i>et al.</i> (37)	Node aligned graph-based MIL	Accuracy: 0.896; AUC ROC: 0.946	Patches for NSCLC: 13,904; patches for RCC: 14,116	Classification	Lung and renal	H&E
Xu <i>et al.</i> (38)	Graph-based	Accuracy: 27.9%; AUC ROC: 0.78±0.03 (mean ± std)	SIFT flow segmentation dataset with 2,688 images and 33 classes TCGA BRCA 20% test and 80% train and validation	Segmentation	Lung and breast	H&E
Lu <i>et al.</i> (39)	Graph-based	TCGA AUC: 75% Two independent data set AUC: 80%	TCGA two independent data sets Training: 80%; test: 20%	Classification	Breast	DAB

Table 2 (continued)

Table 2 (continued)

Study	Model	Performance metrics	Training data	Problem type	Cancer type	Staining type
Zheng <i>et al.</i> (40)	Graph-based	Best models with accuracy CPTAC: 2-label: Resnet + GT: 0.935±0.010 3-label: CL + GraphAtt: 0.835±0.022 TCGA 2-label CL + GraphAtt: 0.911±0.011 3-label CL + GraphAtt: 0.797±0.026	CPTAC: 2,071 WSI's TCGA: 288 WSI's NLST: 665	Classification	Lung	H&E
Lu <i>et al.</i> (41)	Attention based MIL with contrastive learning self-supervised model	AUC ROC 0.968 ± 0.022	H&E strained breast cancer with 400 images of size 2,048 × 1,536 Train: 300 images; valid: 100 images	Classification	Breast	H&E
Ilse <i>et al.</i> (42)	Attention based MIL	Breast: AUC ROC 0.799; colon: AUC ROC 0.968	58 breast; 100 colon	Classification	Breast and colon	H&E
Li <i>et al.</i> (43)	Dual-stream MIL self-supervised contrastive learning	Cameylon-16 accuracy 86.82%, AUC 89.4% TCGA accuracy 91%; AUC 96%	TCGA CAMELYON-16, 271 training, 129 test	Classification	Breast and lung	H&E
Khened <i>et al.</i> (44)	Ensemble	Dice score: 0.782	Datasets with train CAMELYON-16: 270, 129; CAMELYON-17: 500, 500; DigestPath: 660, 212; PAIP: 50, 40	Segmentation	Breast, colon, and liver	H&E
Kalra <i>et al.</i> (45)	MEM invariant representation	TCGA accuracy 84.84%	TCGA 2,580 WSI	Classification	Lung adenocarcinoma and lung squamous cell carcinoma	H&E
Chen <i>et al.</i> (46)	SISH self-supervised MIL	SISH Macro average: 45.51%	WSI: 22,385 Anatomic sites: 13 Subtypes: 56	Classification	37 cancers including lung carcinoma	Variability According to cancer
Niehues <i>et al.</i> (47)	Self-supervised attention based MIL	attMIL AUROC: 0.94±0.02	The dataset consists of 30 patients' data The performance is then compared with previous studies	Classification	Rectal and colon	H&E
Yao <i>et al.</i> (48)	Unsupervised siamese network		1,500 Australian colorectal cancer patients Training: 70%; valid: 10%; test: 10%	Classification	Lung	H&E
Muhammad <i>et al.</i> (49)	Unsupervised deep convolutional autoencoder-based clustering model	Autoencoder with combined MSE loss and reconstruction clustering error Best cluster: 13	246 ICC slides	Clustering	Breast and prostate	H&E
Zhu <i>et al.</i> (50)	Unsupervised K means clustering	WSISA score: 0.703 (NLST), 0.638 (TCGA LUSC), 0.60 (TCGA GBM) WSISA score: 0.703 (NLST), 0.638 (TCGA LUSC), 0.603 (TCGA GBM) WSISA score: 0.440 (NLST), 0.397 (TCGA LUSC), 0.645 (TCGA GBM)	NLST dataset: 404 patients, 1,104 WSI's TCGA dataset is divided into two parts TCGA LUSC: 121 patients, 485 WSI's TCGA GBM: 126 patients WSI's training: 65%; validation: 25%; test: 20%	Clustering	Lung	H&E
Olsson <i>et al.</i> (51)	Conformal prediction	With conformal: 0.1%; error detects: 85% Unreliable predictions without conformal: 2% Error: 25% (error in atypical prostate tissue)	Prostate slides: 7,788; training: 3,059	Classification	Entropy, pseudo hyperplastic, small-cell cancer	H&E
Tsuneki <i>et al.</i> (52)	Weakly Supervised Multi-organ classification	Best Model x10 EfficientNet B1: average accuracy: 0.903	Training 1k, 2k, and 4k WSI's validation on set is randomly chosen Test dataset: WSI's: 4,896	Classification	Prostate	H&E

Table 2 (continued)

Table 2 (continued)

Study	Model	Performance metrics	Training data	Problem type	Cancer type	Staining type
Mohammadi <i>et al.</i> (53)	Weakly supervised	CLAM Accuracy: 87.04%; AUC: 95.06%	iCAIRD endometrial Train: 1,497; valid: 499; test: 911	Classification	Endometrial	H&E
Aswolinskiy <i>et al.</i> (54)	Neural image compression	TCGA AUC: 94%, TCIA AUC: 94% Independent datasets AUC: 84–98%	WSI: 2,000	Segmentation	Lung	H&E
Tellez <i>et al.</i> (55)	Neural compression	Accuracy: 0.725	Rectum: 74 WSI CAMELYON-16: 60 WSI Tupac-16: 40 WSI training: 50,000 (extracted)	Compression and analysis	Breast	H&E
Sornapudi <i>et al.</i> (56)	Image segmentation	CIN model test results IoU5/IoU6 F1: 93.5% Accuracy: 96.5%; AUC: 95.5%	3 datasets with 50 WSI's each IoU: 2,998 True ROIs: 20,841 False ROIs IoU5: 4,915 True ROIs: 12,595 False ROIs IoU6: 4,106 True ROIs: 8,601	Segmentation	Cervical	H&E
Schmitt <i>et al.</i> (57)	CNN	ResNet 50 + CNN UDA: 50+ accuracy 56.1% Slide preparation date 100% slide origin	Total: 42; images: 7 Training: 80%; test: 20%	Classification	Breast, lung, skin, and gastrointestinal	H&E

DL, deep learning; WSI, whole slide image; ROC, receiver operator characteristic; FCN-RF, fully convolutional neural network; RF, receptive field; SUCC, squamous cell carcinoma; CNN, convolutional neural network; MIL, multiple instance learning; BCC, basal cell carcinoma; MIMS, Multi-Instance Multi-Scale; OCT, optical coherence tomography; TURP, transurethral resection of the prostate; TCGA, The Cancer Genome Atlas; RNN, recurrent neural network; DCIS, ductal carcinoma in situ; AUC, area under the curve; IDC, intraductal carcinoma; NSCLC, non-small cell carcinoma; RCC, renal cell carcinoma; DTFD, double-tier feature distillation; SIFT, Scale-Invariant Feature Transform; DAB, 3,3'-diaminobenzidine; CPTAC, clinical proteomic tumor analysis consortium; GT, Graph-Transformer; NLST, National Lung Screening Trial; MEM, Memory-based Exchangeable Model; SISH, self-supervised image search for histology; attMIL, attention-based MIL; MSE, mean square error; ICC, intrahepatic cholangiocarcinoma; WSISA, Whole Slide Histopathology Images Survival Analysis; LUSC, lung squamous cell carcinoma; GBM, glioblastoma multiforme; CLAM, Cluster Attention Multiple Instance Learning; iCAIRD, Industrial Centre for Artificial Intelligence Research in Digital Diagnostics; TCIA, The Cancer Imaging Archive; CIN, Cervical Intraepithelial Neoplasia; IoU, Intersection over Union; ROI, region of interest; UDA, unsupervised data augmentation.

gigapixel-sized WSIs and the lack of MIL annotation (36). CNN encoders and aggregation improve classification accuracy by learning slide-level label representations (36).

In small sample sizes and the large size of WSIs, the double-tier feature distillation MIL (DTFD-MIL) uses pseudo bags to virtually enlarge the number of bags and utilize a double-tier MIL model to improve feature representation (32). This method outperforms other existing methods on CAMELYON-16 and The Cancer Genome Atlas (TCGA) lung cancer datasets (32).

Whole Slide Histopathology Images Survival Analysis (WSISA) is an aggregation method for predicting cancer survival in cases that are too computationally complex for traditional survival models (50). Through adaptive sampling, WSISA extracts hundreds of patches from each WSI. An aggregation method makes patient-level predictions based on cluster-level deep convolutional survival (50). The process is evaluated through experiments on different datasets related to Non-small-cell lung cancer and Glioma, demonstrating its ability to significantly improve prediction performance compared to existing state-of-the-art survival methods (50).

Using MIL approaches, an additional center-embedded aggregation technique accurately classifies colon cancer dataset images. This method learns both instance and bag-level embeddings by hierarchical pooling of features (25).

Spatially aware MIL (saMIL)

A need for more spatial information limits MIL's applications. SaMIL has been proposed to address this problem (59). CNN extracts features from instances within each bag. Combining these features creates a more accurate and interpretable model.

The dual-stream MIL with contrastive learning learns meaningful features from local and global pathology images (43). Two streams of the model are trained, one focusing on local patches and the other on the whole image, and a contrastive loss function is used to encourage the model to learn similar representations (43).

The multi-task MIL (MT-MIL) algorithm handles multiple related tasks simultaneously to capture complex relationships. MT-MIL assigns each task a different bag-level label, and the goal is to learn a model to predict these labels accurately, as used in the study looking at both the diagnosis and prognosis of early-stage invasive breast carcinoma (58).

Global and local features are extracted using a Multiscale

Domain-adversarial Multiple instance CNN, automatically detecting tumor-specific features in a WSI (34). It addresses the difficulties associated with annotating tumor regions in WSIs, extracting global and local image features, and detecting image features against differences in staining conditions among hospitals/specimens for malignant lymphoma use cases (34).

Similarly, a framework for correlating MIL is (TransMIL), incorporating morphological and spatial information (31). The proposed method achieved faster convergence than state-of-the-art methods in various experiments (31). This framework improves weakly supervised tumor classification and cancer subtype identification (31).

A second-order learning model (SoMIL) with an attention mechanism and recurrent neural network (RNN) learns the bag and extracts instance-level feature information trained on the breast cancer lymph node metastasis dataset (33).

Graph-based

Using graph theory, graph-based methods model spatial relationships at the local and global levels to better define cells within the tissue structure and its functional relationships (38,39). Graph-based methods such as the SlideGraph model predict HER2 status in breast cancer and accurately identify and predict HER2-positive regions (39).

Node-aligned graph convolutional network (NAGCN) representation and classification addresses the large gigapixel size of WSI (37). Prior approaches MIL combined with graph convolutional network (GCN), but non-ordered pooling may lose valuable information. Using a global-to-local clustering strategy, NAGCN builds correspondence across different WSIs, representing them with rich local structural information and global distribution (37). It performs better in cancer subtype classification datasets and can be applied to improve WSI representation (37).

The Graph-Transformer (GT) framework called GTP interprets morphological and spatial information for disease grade prediction (40). Contextual information is crucial in disease grading, overcoming the patch-based method's limitations (40). GTP distinguishes adenocarcinoma and squamous cell carcinoma from normal histology (40). A graph-based saliency mapping technique called GraphCAM was also introduced, highlighting WSI regions associated with a class label (40).

Attention-based MIL

An attention-based deep MIL for learning Bernoulli distributions of bag labels, where neural networks fully parameterize the bag label probability (42). Compared to other methods, the proposed method outperforms two real-life histopathology datasets without sacrificing interpretability (42). Pathologists highlight each instance's contribution to each bag label to identify disease markers in large histopathological images (42).

Only sets are labeled on individual data instances in a permutation-invariant neural network called the Memory-based Exchangeable Model (MEM). Based on input sequences embedded in high-level features, the model learns interdependencies among instances using a self-attention mechanism (45). For the classification of two subtypes of lung cancer, the model achieved an accuracy of 84.84% by relying on toy datasets, point cloud classification, and lung WSIs (45). The MEM model can classify histopathology images into different cancer subtypes (45).

The attention-based deep MIL also accurately classifies breast and colon cancer, and it also assists with cancer survival prediction, including high accuracy in predicting the survival of breast cancer patients (42,48).

Weakly supervised learning

Annotation by hand takes time, is tedious, and needs to be more scalable. MIL methods handle weakly labeled or unlabeled data by training on bags of instance (25,26,30,32,35,36,57). Classifying tumors with weakly supervised DL using patch-based models with regions of interest does not require manual annotation (24,35). Weakly labeled semantic segmentation is where only image tags are available as class annotations (19). Latent structured prediction encodes the presence and absence of classes and assigns semantic labels to superpixels (19). It shows improved accuracy in per-class classification compared to state-of-the-art methods (19).

Multi-class learning improves the model's capability to classify complex structures. For weakly supervised image segmentation, representative structure cues in WSI identify glandular regions in endometrial cancer images (53).

Self-supervised learning

Studies using self-supervised learning overcome the labeling bottleneck. One study combines self-supervised feature

learning using contrastive predictive coding (CPC) with regularized attention-based MIL (41). It achieves state-of-the-art performance for binary classification of breast cancer histology images with high accuracy and an area under the receiver operator characteristic (ROC) curve reporting (41). Similarly, in dual-stream MIL networks, MIL-based methods effectively address WSI classification without localized annotations (41).

Its model accuracy can be improved with a novel MIL aggregator that models the relations of the instances in a dual-stream architecture with trainable distance measurement. Self-supervised contrastive learning was used to extract good representations for MIL, and a pyramidal fusion mechanism for multiscale WSI features was used to improve the accuracy of classification and localization (43).

Self-supervised contrastive learning can extract good representations for MIL and reduce cost bags' prohibitive memory costs (43). The classification accuracy and localization are further improved through pyramidal fusion (43).

The study also uses SISH (self-supervised image search for histology) self-supervised DL to search WSI (46). The self-supervised model achieves constant search speed after being trained with only slide-level labels (46). The model encodes WSIs into discrete latent representations. Furthermore, it leverages a tree data structure for fast searching followed by an uncertainty-based ranking algorithm for image retrieval. It identifies similar regions across multiple large diverse WSI datasets with strong performance accuracy (46).

Lastly, the study uses a self-supervised approach that highly accurately predicts biomarkers from WSI across various institutions and scanner types (47). It utilizes self-supervised, attention-based MIL (attMIL) to train models capable of identifying vital morphological characteristics within the WSIs. This methodology allows the model to concentrate on specific regions of the WSI and obtain insights from the data (47).

Unsupervised learning

In unsupervised learning, models discover patterns from unlabeled data on their own. A study identified intrahepatic cholangiocarcinoma subtypes using unsupervised clustering (49). Deep convolutional autoencoders group tumor morphologies based on visual similarity (49). It identifies patterns in cancer tissue images without knowing the type of tumor (49). Tumor stroma is crucial in tumor

growth, angiogenesis, and metastasis; it was used to stratify ductal carcinoma *in situ* (49). These models predict overall survival (49).

Transfer learning

Transfer learning in DP involves using pre-trained CNNs trained on large-scale image data sets and then fine-tuning them on pathology image datasets. Transfer learning aims to overcome the limitations of pathology slide annotations and, by doing so, may improve performance and save time and computational costs (60-64). At times, performance declines when using specific pre-trained models, for example, ones trained on ImageNet, because it may incorporate the significant differences in the natural and the pathology image datasets, resulting in performance inefficiencies (65).

Prostate carcinoma is detected from transurethral resection of the prostate (TURP) images using DL trained on large WSI datasets, then fine-tuned on smaller datasets. Using transfer and weakly supervised learning, they trained DL models to classify TURP WSIs into prostate adenocarcinoma and benign lesions. With these results, these DL models are suitable for diagnostic workflows (52).

The DL breast ductal carcinoma *in situ* classification model trains a CNN on a large, annotated image dataset and uses transfer learning to improve performance (66). It reports improved accuracy compared to traditional machine learning (66).

Ensemble learning

Ensemble learning combines DL models to improve classification performance (44). Ensemble methods reduce the risk of over-fitting and improve overall model accuracy by training multiple models with different architectures or input data (44). One study's system's predictions may assist clinicians in making medical decisions and managing treatment (44).

Other algorithms and techniques

NN192 machine learning algorithm

Acs *et al.* (67) is a significant study in DP, focusing on developing an automated NN192 algorithm for assessing tumor-infiltrating lymphocytes (TILs) in melanoma. The study addresses the challenges of standardization and subjectivity in TIL assessment by utilizing hematoxylin-

eosin-stained sections. The automated scoring system, validated through a retrospective analysis of 641 melanoma patients, demonstrates that higher TIL scores correlate with better disease-specific overall survival. This highlights the potential of the automated TIL scoring system as an independent prognostic marker in melanoma. Furthermore, the study showcases the transformative power of digital technologies in pathology, improving the accuracy and efficiency of pathological assessments. This research contributes to melanoma prognosis and sets a precedent for the broader application of DP in cancer care and research (67).

Aung *et al.* (68) examine how DP, specifically in melanoma, can improve cancer diagnosis and prognosis. Using the NN192 machine learning algorithm, the study achieves an objective and precise assessment of TILs compared to traditional visually based evaluations. The analysis of TILs in melanoma samples highlights the potential of DP to provide consistent and reproducible results. This advancement is crucial in personalized medicine, where accurate biomarker assessment is essential for tailoring cancer treatment. The study emphasizes integrating digital technologies into pathology for more precise diagnoses and better patient outcomes in oncology.

Neural image compression (NIC)

NIC reduces gigapixel WSI images to extremely compact representations for training CNNs to predict image labels (54,55,69,70). Three encoding tools transform low-level vector embeddings of high-resolution WSIs: contrastive learning, reconstructional error minimization, and bidirectional generative adversarial network (BiGAN). BiGANs outperformed the other two unsupervised encoding mechanisms, with a Spearman correlation of 0.521 (55). NIC builds CNN for gigapixel image analysis based on weak image-level labels (55). The first step is compressing gigapixel images using an unsupervised neural network, retaining high-level information, and suppressing pixel noise (55). CNN is trained on these compressed image representations to predict image-level labels, avoiding manual annotations (55). Two public histopathology datasets were evaluated with NIC and found to integrate global and local visual information while attending to areas of the input gigapixel images that overlap with human annotations (55). NIC classifies non-small cell lung cancer subtypes with high accuracy (54). In addition, compression algorithms can preserve important image features and accuracy (70).

Conformal prediction

Conformal prediction is a mathematical framework to assess the reliability of prediction systems for diagnosing and grading prostate biopsies. A model is trained to estimate its accuracy, and conformal prediction intervals are generated to quantify its uncertainty (68). Conformal prediction may improve the reliability and interpretability of AI-assisted pathology diagnosis (68). Conformal prediction led to a lower rate of incorrect cancer diagnoses and flagged a higher percentage of unreliable predictions than AI systems without conformal prediction (68). Conformal prediction can help augment medical providers' decision-making in the clinical setting (51).

Elements impacting model development

Factors that should be considered arise when developing a DL model (71-75). Awareness of these factors is crucial to ensure a successful development process.

Reducing GPU use and memory consumption with sustainable AI technique

A DL model for WSI often needs a lot of computing power, especially when working with large-pixel data. GPU usage and memory consumption can be high, leading to longer training times and higher costs. DL techniques have been developed to address this issue, limiting GPU use and memory consumption. Compressed models, transfer learning, data augmentation, and batch normalization have emerged as effective techniques that reduce GPU usage and memory consumption while improving model accuracy, training times, and energy costs, reducing the carbon footprint (76,77). Preparation variability slide prep, staining, tissue preparation protocols, and scanners can affect DL performance. This involves processing enormous amounts of data with these variations, labeling ground truth, comprehensive image pre-processing, denoising, and WSI normalization. Normalization and data augmentation also adjust for non-cancer cells, necrosis, and inflammation. MIL, attention, and graph-based methods capture spatial relationships between these cells. Furthermore, malignant cells can be challenging to identify and classify due to non-cancerous cells and tissue artifacts; this is addressed by handcrafted feature extraction.

Data augmentation

Data augmentation involves rotating, flipping, and scaling existing data to generate new training data. As a result, DL models perform well and require less training data, and it helps prevent overfitting and reduce memory and GPU utilization.

Defining accurate ground truth

We need ground truth and a labeled set of data points to train, test, and validate DL models to compare model outputs. The accuracy of ground truth impacts model accuracy. Experts traditionally label ground truth. Pathologists manually highlight regions of interest in images and train DL models to detect similar areas in other images. Expert annotation verification for large datasets is time-consuming, expensive, and not scalable. Experts may also need help labeling consistently.

Labeling annotation

Labeling slide annotation is time-consuming and requires much effort from experts. Moreover, the amount of labeled data available may be limited, impacting the performance of DL algorithms. Recent progress in developing weakly, semi, self-supervised, and unsupervised machine-learning clustering techniques can be used to analyze WSI data without labeled data (24-26,30,32,35,36,41,43,46,47,49,53,57). In semi-supervised learning, a limited amount of labeled data is combined with a more significant amount of unlabeled data to train models; it allows DL models to perform better with less labeled data required.

Domain adaptation and diversity

Domain adaptation refers to the ability of a model to perform well on data from a target domain that may differ from the domain in which the model was initially trained. Poor domain adaptation is a significant challenge in WSI DL implementation, attributed to the high-resolution large gigapixel size of the images and annotation challenges. DL models may lose performance accuracy (35,78-80).

Data diversity, sensors, and patients must be considered when developing DL models for WSI. Managing data

diversity is challenging since slide images can come from different hospitals, laboratories, and clinics. These differences in quality, resolution, and staining can affect DL models.

It is crucial to consider sensor type when training DL models. Sensors can also affect image quality. Images acquired by fluorescence microscopy may differ from those acquired by brightfield microscopy.

Finally, consider the patients; different patient populations must be represented in a diverse dataset.

Structured prediction global vs. local features

Structured predictions capture spatial and heterogeneous cell relationships in various spatial scales. Global cell features describe the overall distribution of cells in a tissue, while local cell features describe their immediate neighbors. This model captures complex spatial relationships between cells by combining these features. Interactions between cells are complex and dynamic (81). The properties of cells vary depending on where they are in tissues. By combining global and local features and examining their cellular relationships, structured prediction makes accurate predictions about individual cells' malignant behavior.

High spatial resolution

A substantial amount of spatial data generated by WSI can make the development of accurate DL algorithms challenging. Studies reviewed noted that image compression and patch-based techniques can overcome this limitation.

Compressed models

Model compression minimizes the number of parameters (54,55,69,70). A pruning technique removes less meaningful connections between nodes, and a quantization technique reduces the precision of the weights and activations. Compressed models use less memory, can be trained faster, reduce GPU usage, and improve digital storage (54,55,69,70).

Batch normalization

During training, batch normalization and activations at each neural network layer are normalized to improve stability and convergence. Using batch normalization, DL models can be trained faster and more accurately and use less memory.

Transfer learning

A pre-trained model is fine-tuned in transfer learning for a specific task (60-64,82). This approach can significantly reduce training time, GPU, and memory use. Transfer learning is especially effective for tasks with similar features and datasets. Recent developments in resource-efficient neuroevolutionary multi-tasking have shown promising results within machine learning pipelines, including AutoML (76). Gradient-free evolutionary optimizers have emerged as a powerful alternative to traditional DL, as demonstrated by OpenAI (83).

Combining ResNet50 (residual network 50), a CNN that is 50 layers deep, with weakly supervised or unsupervised techniques in DL WSI improves image classification performance and addresses gradient explosion. Training with a pre-trained ResNet50's backbone is faster and more efficient. The model has already learned general features from a large dataset, requiring fewer data and iterations to reach good accuracy. This is useful for small datasets or scenarios requiring real-time inference. The quality and diversity of the training data determine the model's performance. Data preparation and selection are critical to achieving optimal results. Not all pre-trained models are suitable for pathology. For example, DL models pre-trained on ImageNet at certain times may not be ideal for the computational pathology task (65). The ImageNet models are trained on natural images, whereas pathological images have unique features. To achieve optimal performance, pre-trained models may need to be fine-tuned or trained from scratch. Furthermore, this model may not capture all the variations in pathology images. Artifacts, stains, and other factors may affect the accuracy of DL models in pathology images. Careful evaluation of pre-trained models on pathology images and considering alternatives is essential.

Rare conditions

Rare conditions are underrepresented in training data, affecting DL models' ability to diagnose them. Techniques in transfer learning, data augmentation, synthetic data, and federated learning address these limitations.

Healthcare bias, inclusion, fairness, and equity

DL digital services should incorporate inclusion, fairness, bias detection algorithms, metrics, and strict governing protocols during the DL model development lifecycle

(6,18,84,85). DL model development workflows should strive to obtain model parity. Fletcher describes three aspects of DL model development: fairness, appropriateness, and bias (85). To limit disparities, these elements must be addressed when developing DL models: religion, economic status, ethnicity, race, and gender should also be incorporated into these algorithms (84). If left unaddressed, AI models may demonstrate bias, as noted by (84). Feedback algorithms can rectify skewed patient datasets (84). Local governance committees and agencies, such as the Joint Commission and Institute for Healthcare Improvement, provide ongoing oversight to ensure equitable conditions. Continuous monitoring of quality data underscores the importance of equity in clinical practice.

False positives management

DL models for WSI improve sensitivity to detect cancer. However, with this improvement, false positive detection increases. While the models are better at detecting cancer, they may also flag areas as cancerous when they are not. Training the models on more extensive and diverse datasets may mitigate the risk of false positives and improve their accuracy. A study revealed that techniques can reduce and improve false positives (86). When interpreting the results of these models, it is vital to consider false positive detections.

Data privacy

Privacy and security concerns arise with WSI and DL technologies. As WSI deals with large amounts of sensitive patient information, strong data privacy policies, encryption, and access control are indispensable. Differential privacy methods such as encryption and federated learning protect individual privacy while allowing accurate data analysis (87-89).

Conclusions

Data pre-processing and model development represent only a small fraction of the expansive DL lifecycle. A truly seamless journey from model development to clinical production environment is ideal. To achieve this seamlessly on a healthcare enterprise scale within DP, a robust AI technical infrastructure, expert personnel, and a governance system for continuous oversight are needed.

The digitalization of pathology has trailed behind its

radiology counterpart, which transitioned from analog films to digital systems decades ago. This lag is evident in the field of DL within DP compared to radiology, as demonstrated by the numerous algorithm approvals by the FDA. While the technical intricacies of model development are vital, as discussed in this review, they are but one piece of the puzzle.

Future longitudinal clinical trials during the post-deployment production phase of the DL lifecycle are essential. These studies will offer invaluable insights into the real-world clinical impact of DL on cancer care, guiding us toward a future where innovative technology and healthcare intersect seamlessly to benefit patients.

Being a narrative review, this study design may exhibit more bias and is not intended to be as comprehensive as other methods.

Acknowledgments

Funding: None.

Footnote

Reporting Checklist: The authors have completed the Narrative Review reporting checklist. Available at <https://tcr.amegroups.com/article/view/10.21037/tcr-23-964/rc>

Peer Review File: Available at <https://tcr.amegroups.com/article/view/10.21037/tcr-23-964/prf>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://tcr.amegroups.com/article/view/10.21037/tcr-23-964/coif>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. WHO International Agency for Research on Cancer. Estimated number of deaths in 2020, all cancers, sexes, and ages. Cancer today. 2020. Available online: <https://gco.iarc.fr/today/online-analysis-pie>
2. American Cancer Society. The global cancer burden why global cancer rates are rising. Available online: <https://rb.gy/8ztpz6>
3. Siegel RL, Miller KD, Wagle NS, et al. Cancer statistics, 2023. *CA Cancer J Clin* 2023;73:17-48.
4. Chow RD, Bradley EH, Gross CP. Comparison of Cancer-Related Spending and Mortality Rates in the US vs 21 High-Income Countries. *JAMA Health Forum* 2022;3:e221229.
5. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* 2021;71:209-49.
6. Hall WJ, Chapman MV, Lee KM, et al. Implicit Racial/Ethnic Bias Among Health Care Professionals and Its Influence on Health Care Outcomes: A Systematic Review. *Am J Public Health* 2015;105:e60-76.
7. Culyer AJ, Wagstaff A. Equity and equality in health and health care. *J Health Econ* 1993;12:431-57.
8. Brown AF, Ma GX, Miranda J, et al. Structural Interventions to Reduce and Eliminate Health Disparities. *Am J Public Health* 2019;109:S72-8.
9. The Joint Commission. Take 5: The joint commission's new diversity and inclusion activities. Available online: <https://shorturl.at/emL89>
10. Strategies for Quality Care. A framework for achieving value-based care. Available online: <https://www.strategiesforqualitycare.com/quadruple-aim>
11. Nishat R, Ramachandra S, Behura SS, et al. Digital cytopathology. *J Oral Maxillofac Pathol* 2017;21:99-106.
12. Snead DR, Tsang YW, Meskiri A, et al. Validation of digital pathology imaging for primary histopathological diagnosis. *Histopathology* 2016;68:1063-72.
13. Goacher E, Randell R, Williams B, et al. The Diagnostic Concordance of Whole Slide Imaging and Light Microscopy: A Systematic Review. *Arch Pathol Lab Med* 2017;141:151-61.
14. Babawale M, Gunavardhan A, Walker J, et al. Verification and Validation of Digital Pathology (Whole Slide Imaging) for Primary Histopathological Diagnosis: All Wales Experience. *J Pathol Inform* 2021;12:4.
15. Mukhopadhyay S, Feldman MD, Abels E, et al. Whole Slide Imaging Versus Microscopy for Primary Diagnosis in Surgical Pathology: A Multicenter Blinded Randomized Noninferiority Study of 1992 Cases (Pivotal Study). *Am J Surg Pathol* 2018;42:39-52.
16. Pannu V, Sprinkle S, Stemm K, et al. Large scaled digital pathology clinical operations for precision medicine -reflection and aspiration. 2023. Available online: <https://dpa.planion.com/Web/User/PDFViewer?ACCOUNT=DPA&conf=Pv2023&ssoOverride=OFF&ckey=&PDFID=6f364eb5-4892-4fbd-90f8-aa62b3f1d525&AUDIOID=&VIDEOID=>
17. Patel A, Balis UGJ, Cheng J, et al. Contemporary Whole Slide Imaging Devices and Their Applications within the Modern Pathology Department: A Selected Hardware Review. *J Pathol Inform* 2021;12:50.
18. Thomasian NM, Eickhoff C, Adashi EY. Advancing health equity with artificial intelligence. *J Public Health Policy* 2021;42:602-11.
19. Bera K, Schalper KA, Rimm DL, et al. Artificial intelligence in digital pathology - new tools for diagnosis and precision oncology. *Nat Rev Clin Oncol* 2019;16:703-15.
20. Pallua JD, Brunner A, Zelger B, et al. The future of pathology is digital. *Pathol Res Pract* 2020;216:153040.
21. Azam AS, Miligy IM, Kimani PK, et al. Diagnostic concordance and discordance in digital pathology: a systematic review and meta-analysis. *J Clin Pathol* 2021;74:448-55.
22. GlobeNewswire. With 12.6% CAGR, digital pathology market size to surpass USD 2045.9 million by 2029. Available online: <https://shorturl.at/vBIYZ>
23. Wang D, Khosla A, Gargeya R, et al. Deep learning for identifying metastatic breast cancer. *Arxiv Preprint Arxiv:160605718* 2016.
24. Wang X, Chen H, Gan C, et al. Weakly Supervised Deep Learning for Whole Slide Lung Cancer Image Analysis. *IEEE Trans Cybern* 2020;50:3950-62.
25. Chikontwe P, Kim M, Nam SJ, et al. Multiple instance learning with center embeddings for histopathology classification. In: Martel AL, Abolmaesumi P, Stoyanov D, et al. editors. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, Cham: Springer International Publishing;2020:519-28.
26. Xie C, Muhammad H, Vanderbilt CM, et al. Beyond Classification: Whole Slide Tissue Histopathology Analysis By End-To-End Part Learning. In: Arbel T, Ben AI, de Bruijne M, et al. editors. *Proceedings of the Third Conference on Medical Imaging with Deep Learning*,

- Breckenridge, Colorado, USA, PMLR, 2020:843-56.
27. Li S, Liu Y, Sui X, et al. Multi-instance multi-scale CNN for medical image classification. In: Shen D, Liu T, Peters TM, et al., editors. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, Cham: Springer International Publishing; 2019:531-9.
 28. Tsuneki M, Kanavati F. Weakly supervised learning for multi-organ adenocarcinoma classification in whole slide images. *PLoS One* 2022;17:e0275378.
 29. Kanavati F, Ichihara S, Tsuneki M. A deep learning model for breast ductal carcinoma in situ classification in whole slide images. *Virchows Arch* 2022;480:1009-22.
 30. Campanella G, Hanna MG, Geneslaw L, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat Med* 2019;25:1301-9.
 31. Shao Z, Bian H, Chen Y, et al. Transmil: Transformer based correlated multiple instance learning for whole slide image classification. *Adv Neural Inf Process Syst* 2021;34:2136-47.
 32. Zhang H, Meng Y, Zhao Y, et al. DTFD-MIL: Double-tier feature distillation multiple instance learning for histopathology whole slide image classification. In: *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Piscataway, New Jersey, IEEE, 2022:18780-90.
 33. Wang Q, Zou Y, Zhang J, et al. Second-order multi-instance learning model for whole slide image classification. *Phys Med Biol* 2021.
 34. Hashimoto N, Fukushima D, Koga R, et al. Multi-scale domain-adversarial multiple-instance CNN for cancer subtype classification with unannotated histopathological images. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, Seattle, WA, USA, IEEE, 2020:3852-61.
 35. Lu MY, Williamson DFK, Chen TY, et al. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nat Biomed Eng* 2021;5:555-70.
 36. Sharma Y, Shrivastava A, Ehsan L, et al. Cluster-to-conquer: A framework for end-to-end multi-instance learning for whole slide image classification. In: Heinrich M, Dou Q, de Bruijne M, et al. editors. *Proceedings of the Fourth Conference on Medical Imaging with Deep Learning*, Breckenridge, Colorado, USA, PMLR, 2021:682-98.
 37. Guan Y, Zhang J, Tian K, et al. Node-aligned graph convolutional network for whole-slide image representation and classification. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, LA, USA, IEEE, 2022:18813-23.
 38. Xu J, Schwing AG, Urtasun R. Tell me what you see, and I will show you where it is. In: *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Piscataway, NJ, IEEE, 2014:3190-7.
 39. Lu W, Toss M, Dawood M, et al. SlideGraph(+): Whole slide image level graphs to predict HER2 status in breast cancer. *Med Image Anal* 2022;80:102486.
 40. Zheng Y, Gindra RH, Green EJ, et al. A Graph-Transformer for Whole Slide Image Classification. *IEEE Trans Med Imaging* 2022;41:3003-15.
 41. Lu MY, Chen RJ, Mahmood F. Semi-supervised breast cancer histology classification using deep multiple instance learning and contrast predictive coding (conference presentation). In: *Medical imaging 2020: digital pathology*, Bellingham, Washington, SPIE, 2020:113200J.
 42. Ilse M, Tomczak J, Welling M. Attention-based deep multiple instance learning. In: Dy J, Krause A, editors. *Proceedings of the 35th International Conference on Machine Learning*, Breckenridge, Colorado, USA, PMLR, 2018:2127-36.
 43. Li B, Li Y, Eliceiri KW. Dual-stream Multiple Instance Learning Network for Whole Slide Image Classification with Self-supervised Contrastive Learning. *Conf Comput Vis Pattern Recognit Workshops* 2021;2021:14318-28.
 44. Khened M, Kori A, Rajkumar H, et al. A generalized deep learning framework for whole-slide image segmentation and analysis. *Sci Rep* 2021;11:11579.
 45. Kalra S, Adnan M, Taylor G, et al. Learning permutation invariant representations using memory networks. In: Vedaldi A, Bischof H, Brox T, et al., editors. *Computer Vision – ECCV 2020*, Cham: Springer International Publishing; 2020:677-93.
 46. Chen C, Lu MY, Williamson DFK, et al. Fast and scalable search of whole-slide images via self-supervised deep learning. *Nat Biomed Eng* 2022;6:1420-34.
 47. Niehues JM, Quirke P, West NP, et al. Generalizable biomarker prediction from cancer pathology slides with self-supervised deep learning: A retrospective multi-centric study. *Cell Rep Med* 2023;4:100980.
 48. Yao J, Zhu X, Jonnagaddala J, et al. Whole slide images based cancer survival prediction using attention guided deep multiple instance learning networks. *Med Image Anal* 2020;65:101789.
 49. Muhammad H, Sigel CS, Campanella G, et al. Towards unsupervised cancer subtyping: Predicting prognosis using

- a histologic visual dictionary. Arxiv Prepr Arxiv:1903.05257 2019.
50. Zhu X, Yao J, Zhu F, et al. WSISA: Making survival prediction from whole slide histopathological images. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Piscataway, New Jersey, IEEE, 2017:6855-63.
 51. Olsson H, Kartasalo K, Mulliqi N, et al. Estimating diagnostic uncertainty in artificial intelligence assisted pathology using conformal prediction. *Nat Commun* 2022;13:7761.
 52. Tsuneki M, Abe M, Kanavati F. Transfer Learning for Adenocarcinoma Classifications in the Transurethral Resection of Prostate Whole-Slide Images. *Cancers (Basel)* 2022;14:4744.
 53. Mohammadi M, Cooper J, Arandelović O, et al. Weakly supervised learning and interpretability for endometrial whole slide image diagnosis. *Exp Biol Med (Maywood)* 2022;247:2025-37.
 54. Aswolinskiy W, Tellez D, Raya G, et al. Neural image compression for non-small cell lung cancer subtype classification in H&E stained whole-slide images. *Proc. SPIE 11603, Medical Imaging 2021: Digital Pathology* doi: 10.1117/12.2581943.
 55. Tellez D, Litjens G, van der Laak J, et al. Neural Image Compression for Gigapixel Histopathology Image Analysis. *IEEE Trans Pattern Anal Mach Intell* 2021;43:567-78.
 56. Sornapudi S, Hagerty J, Stanley RJ, et al. EpithNet: Deep Regression for Epithelium Segmentation in Cervical Histology Images. *J Pathol Inform* 2020;11:10.
 57. Schmidt A, Silva-Rodríguez J, Molina R, et al. Efficient cancer classification by coupling semi supervised and multiple instance learning. *IEEE Access* 2022;10:9763-73.
 58. Liu J, Ge R, Wan P, et al. Multi-task multi-instance learning for jointly diagnosis and prognosis of early-stage breast invasive carcinoma from whole-slide pathological images. In: Frangi A, de Bruijne M, Wassermann D, et al. editors. *Information Processing in Medical Imaging*, Cham: Springer Nature Switzerland; 2023:145-57.
 59. Wölflein G, Magister LC, Liò P, et al. Deep Multiple Instance Learning with Distance-Aware Self-Attention. *Arxiv Preprint Arxiv:2305.10552* 2023.
 60. Wahab N, Khan A, Lee YS. Transfer learning based deep CNN for segmentation and detection of mitoses in breast cancer histopathological images. *Microscopy (Oxf)* 2019;68:216-33.
 61. Bayramoglu N, Heikkilä J. Transfer learning for cell nuclei classification in histopathology images. In: Hua G, Jégou H, editors. *Computer Vision – ECCV 2016 Workshops*, Cham: Springer International Publishing; 2016:532-9.
 62. Ahmed S, Shaikh A, Alshahrani H, et al. Transfer Learning Approach for Classification of Histopathology Whole Slide Images. *Sensors (Basel)* 2021;21:5361.
 63. Alinsaf S, Lang J. Histological image classification using deep features and transfer learning. In: editors. 2020 17th Conference on Computer and Robot Vision (CRV), Piscataway, NJ, IEEE, 2020:101-8.
 64. Xu Y, Jia Z, Wang LB, et al. Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features. *BMC Bioinformatics* 2017;18:281.
 65. Rai T, Morisi A, Bacci B, et al. An investigation of aggregated transfer learning for classification in digital pathology. In: *Medical Imaging 2019: Digital Pathology*, Bellingham, Washington, SPIE, 2019:109560U.
 66. Kanavati F, Tsuneki M. Breast Invasive Ductal Carcinoma Classification on Whole Slide Images with Weakly-Supervised and Transfer Learning. *Cancers (Basel)* 2021;13:5368.
 67. Acs B, Ahmed FS, Gupta S, et al. An open source automated tumor infiltrating lymphocyte algorithm for prognosis in melanoma. *Nat Commun* 2019;10:5440.
 68. Aung TN, Shafi S, Wilmott JS, et al. Objective assessment of tumor infiltrating lymphocytes as a prognostic marker in melanoma using machine learning algorithms. *EBioMedicine* 2022;82:104143.
 69. Keighley J, de Kamps M, Wright A, et al. Digital pathology whole slide image compression with vector quantized variational autoencoders. In: *Medical Imaging 2023: Digital and Computational Pathology*, Bellingham, Washington, SPIE, 2023:344-53.
 70. Chen Y, Janowczyk A, Madabhushi A. Quantitative Assessment of the Effects of Compression on Deep Learning in Digital Pathology Image Analysis. *JCO Clin Cancer Inform* 2020;4:221-33.
 71. Rizzo PC, Girolami I, Marletta S, et al. Technical and Diagnostic Issues in Whole Slide Imaging Published Validation Studies. *Front Oncol* 2022;12:918580.
 72. Patel AU, Shaker N, Erck S, et al. Types and frequency of whole slide imaging scan failures in a clinical high throughput digital pathology scanning laboratory. *J Pathol Inform* 2022;13:100112.
 73. Abdelsamea MM, Zidan U, Senousy Z, et al. A survey on artificial intelligence in histopathology image analysis. *WIREs Data Min Knowl Discov* 2022;12:e1474.

74. Zarella MD, Rivera Alvarez K. High-throughput whole-slide scanning to enable large-scale data repository building. *J Pathol* 2022;257:383-90.
75. Nakagawa K, Moukheiber L, Celi LA, et al. AI in Pathology: What could possibly go wrong? *Semin Diagn Pathol* 2023;40:100-8.
76. Bali KK, Ong YS, Gupta A, et al. Multifactorial evolutionary algorithm with online transfer parameter estimation: MFEA-II. *IEEE Trans Evol Comput* 2020;24:69-83.
77. van Wynsberghe A. Sustainable AI: AI for sustainability and the sustainability of AI. *AI and Ethics* 2021;1:213-8.
78. Stacke K, Eilertsen G, Unger J, et al. Measuring Domain Shift for Deep Learning in Histopathology. *IEEE J Biomed Health Inform* 2021;25:325-36.
79. Walker AJ. Adaptive Domain Generalization for Digital Pathology Images. Minnesota, Doctoral Dissertation, University of Minnesota, 2022.
80. Falahkheirkhah K, Lu A, Alvarez-Melis D, et al. Domain adaptation using optimal transport for invariant learning using histopathology datasets. *Arxiv Prepr Arxiv:2303.02241* 2023.
81. Deng Y, Feng M, Jiang Y, et al. Development of pathological reconstructed high-resolution images using artificial intelligence based on whole slide image. *MedComm (2020)* 2020;1:410-7.
82. Weiss K, Khoshgoftaar TM, Wang D. A survey of transfer learning. *J Big Data* 2016;3:1-40.
83. Salimans T, Ho J, Chen X, et al. Evolution strategies as a scalable alternative to reinforcement learning. *Arxiv Prepr Arxiv:1703.03864* 2017.
84. Pagano TP, Loureiro RB, Lisboa FVN, et al. Bias and unfairness in machine learning models: A systematic review on datasets, tools, fairness metrics, and identification and mitigation methods. *Big Data and Cognitive Computing* 2023;7:15.
85. Fletcher RR, Nakeshimana A, Olubeko O. Addressing Fairness, Bias, and Appropriate Use of Artificial Intelligence and Machine Learning in Global Health. *Front Artif Intell* 2021;3:561802.
86. Janowczyk A, Madabhushi A. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *J Pathol Inform* 2016;7:29.
87. Holub P, Müller H, Bil T, et al. Privacy risks of whole-slide image sharing in digital pathology. *Nat Commun* 2023;14:2577.
88. Lu MY, Chen RJ, Kong D, et al. Federated learning for computational pathology on gigapixel whole slide images. *Med Image Anal* 2022;76:102298.
89. Truhn D, Tayebi Arasteh S, Saldanha OL, et al. Encrypted federated learning for secure decentralized collaboration in cancer image analysis. *Med Image Anal* 2024;92:103059.

Cite this article as: Williams DKA Jr, Graifman G, Hussain N, Amiel M, Tran P, Reddy A, Haider A, Kavitesh BK, Li A, Alishahian L, Perera N, Efros C, Babu M, Tharakan M, Etienne M, Babu BA. Digital pathology, deep learning, and cancer: a narrative review. *Transl Cancer Res* 2024;13(5):2544-2560. doi: 10.21037/tcr-23-964