Lab Invest 103 (2023) 100255

LABORATORY INVESTIGATION

journal homepage: https://laboratoryinvestigation.org/

USCAP UNITED STATES AND CANADIAN ACADEMY OF PATHOLOGY
Creating a Better Pathologist

Review Article

Revolutionizing Digital Pathology With the Power of Generative Artificial Intelligence and Foundation Models

Asim Waqas[a,b,*], Marilyn M. Bui[a,c,d], Eric F. Glassy[e], Issam El Naqa[a], Piotr Borkowski[f,g], Andrew A. Borkowski[d,h,i], Ghulam Rasool[a,b,d,j]

[a] Department of Machine Learning, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida; [b] Department of Electrical Engineering, University of South Florida, Tampa, Florida; [c] Department of Pathology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida; [d] University of South Florida, Morsani College of Medicine, Tampa, Florida; [e] Affiliated Pathologists Medical Group, Inc., Rancho Dominguez, California; [f] Quest Diagnostics/Ameripath, Tampa, Florida; [g] Center of Excellence for Digital and AI-Empowered Pathology, Quest Diagnostics, Tampa, Florida; [h] James A. Haley Veterans' Hospital, Tampa, Florida; [i] National Artificial Intelligence Institute, Washington, District of Columbia; [j] Department of Neuro-Oncology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida

A R T I C L E I N F O

Article history:
Received 14 April 2023
Revised 6 September 2023
Accepted 19 September 2023
Available online 26 September 2023

Keywords:
artificial intelligence
computational and digital pathology
foundation models
large language models
multimodal data
vision-language models

A B S T R A C T

Digital pathology has transformed the traditional pathology practice of analyzing tissue under a microscope into a computer vision workflow. Whole-slide imaging allows pathologists to view and analyze microscopic images on a computer monitor, enabling computational pathology. By leveraging artificial intelligence (AI) and machine learning (ML), computational pathology has emerged as a promising field in recent years. Recently, task-specific AI/ML (eg, convolutional neural networks) has risen to the forefront, achieving above-human performance in many image-processing and computer vision tasks. The performance of task-specific AI/ML models depends on the availability of many annotated training datasets, which presents a rate-limiting factor for AI/ML development in pathology. Task-specific AI/ML models cannot benefit from multimodal data and lack generalization, eg, the AI models often struggle to generalize to new datasets or unseen variations in image acquisition, staining techniques, or tissue types. The 2020s are witnessing the rise of foundation models and generative AI. A foundation model is a large AI model trained using sizable data, which is later adapted (or fine-tuned) to perform different tasks using a modest amount of task-specific annotated data. These AI models provide in-context learning, can self-correct mistakes, and promptly adjust to user feedback. In this review, we provide a brief overview of recent advances in computational pathology enabled by task-specific AI, their challenges and limitations, and then introduce various foundation models. We propose to create a pathology-specific generative AI based on multimodal foundation models and present its potentially transformative role in digital pathology. We describe different use cases, delineating how it could serve as an expert companion of pathologists and help them efficiently and objectively perform routine laboratory tasks, including quantifying image analysis, generating pathology reports, diagnosis, and prognosis. We also outline the potential role that foundation models and generative AI can play in standardizing the pathology laboratory workflow, education, and training.

* Corresponding author.
E-mail address: Asim.Waqas@moffitt.org (A. Waqas).

ELSEVIER

0023-6837/$ - see front matter
https://doi.org/10.1016/j.labinv.2023.100255

Introduction

Conventional pathology methods have been crucial in diagnosing disease, heavily relying on examining tissue samples under a microscope. With technological advancements and a growing emphasis on precision medicine, digital pathology (DP) has emerged as a new approach for conducting precise quantitative assessments. DP involves utilizing whole-slide imaging (WSI) to digitize and analyze tissue samples using a computer. Computational pathology further builds on it and incorporates artificial intelligence (AI) and machine learning (ML) to enable the extraction of information that goes beyond what the human eye can perceive. The clinical responsibilities of pathologists, such as providing precise diagnoses and quantifying biomarkers for diagnosis, prognosis, and predictions, may be strengthened in terms of precision, reproducibility, and scalability by using AI-driven analysis tools. AI can address the challenging problems in pathology workflow, including the following: (1) increasing workload and staff shortages leading to physician burnout, (2) growing diagnostic complexity, including ever-expanding cancer protocols and biomarkers, (3) case variability, often involving rare diseases or overlapping morphologic changes, (4) issues with the quality of slides due to artifacts introduced by tissue folding, staining inconsistencies, and compression artifacts, and (5) lack of standardization, which hinders interoperability between different laboratories, platforms, image formats, and analysis tools.

AI is a broad field focused on simulating human intelligence by creating models and algorithms to automate various tasks, such as recognizing objects in images, understanding and generating natural language text, or making predictions based on historical data.[1] ML is a subset of AI that involves creating statistical and mathematical models and learning algorithms for recognizing patterns in the data.[2] Artificial neural networks that attempt to mimic the human brain's way of analyzing data have recently made significant progress.[3] The advancements made possible by artificial neural networks have revolutionized computer vision (a subfield of AI that deals with image processing) and natural language processing (a subfield of AI that deals with text and speech).[3] Although the initial adoption of these technologies in medicine and health care was slow, recently, medical imaging has been transforming at an unprecedented rate. Digital and computational pathology are also rapidly evolving on the research front, with the industry offering new AI-enabled technologies.[4-8]

Although task-specific traditional AI tools date back to the 1970s, the decade of 2010 saw a sharp rise in the research and development of narrow AI methods enabled by deep learning models, eg, convolutional neural networks (CNNs), recurrent neural networks (RNNs), and Transformers. These AI models eliminated the need for feature engineering using domain expertise, a defining characteristic of classical ML techniques widely known as pathomics in the pathology domain.[9] For a given task, the performance of these artificial neural network-based AI models surpassed previous AI techniques. Developing a task-specific AI starts with selecting a particular problem, eg, counting mitosis in a histopathology image, then curating and annotating relevant historical data, and finally, training the model by learning optimal parameters (or weights). Annotating (or equivalently labeling) data requires experts (pathologists) to carefully review each data sample and identify/define objects/patterns that help AI learn about the task during training. The performance of task-specific AI with supervised learning techniques strongly depends on the availability of large, high-quality, annotated training datasets. Despite boasting above-human performance, task-specific AI suffers from significant limitations, including the requirement for a large amount of expert-annotated datasets, the lack of performance generalization (eg, the AI may fail if used on images generated using a staining protocol different than the one used for generating training images), and the inability to use relevant data from other modalities, eg, patient demographics, laboratory data, or their prior disease history cannot help the model improve its prediction accuracy.[4,5] Analysis of task-specific AI in pathology through qualitative interviews of 24 professionals revealed such shortcomings in the existing tools, which hinder their broad integration in the decision-making processes of pathologists.[10] For further details, the reader is referred to the surveys reviewing the use of AI in pathology.[11,12]

The 2020s are witnessing the rise of foundation models and generative AI. Foundation models are very large task-agnostic AI models trained using unannotated (possibly multimodal) datasets and form the brain of generative AI.[13,14] A trained foundation model can be adapted to perform many different tasks using a modest amount of task-specific annotated data.[14] Training a foundation model may not require manually annotating large amounts of data because these models use self-, semi-, or unsupervised learning techniques. Foundation models can consume data from various modalities, including images (eg, WSIs), text (eg, pathology reports), and tabular data (eg, medical records). The well-known generative AI model, ChatGPT, is based on a foundation model called Generative Pretraining Transformer (GPT).[15-20] Foundation models hold much promise for quantitative image analysis, diagnosis and prognosis, pathology report generation, and questioning/answering with conversational use in pathology laboratory workflow.[13,14]

In this article, "AI in Digital Pathology" provides a brief overview of AI and ML models and advancements enabled by these task-specific AI models in computational pathology. We introduce various foundation models, their structure, characteristics, and limitations in "Foundation Models and Generative AI." "Transformers, Foundation Models, and Digital Pathology" outlines the transformative role that foundation models may play in the pathology laboratory workflow in the near future. We provide use cases delineating how a pathology generative AI based on a foundation model could serve as an expert companion pathologist that assists in efficiently and objectively performing routine laboratory tasks, including image analysis, presenting and justifying findings, quantifying the analysis, generating reports, performing prognostics, and making predictions. In the last section, we have outlined the potential role that foundation models and generative AI can play in pathology education and training.

Artificial Intelligence in Digital Pathology

AI comprises computational methods, statistical and mathematical models, and the implementation of various algorithms to mimic human-style intelligence. AI-based technologies have enabled pathologists and researchers to analyze large amounts of data with greater accuracy and speed, making the process of disease diagnosis faster and more precise.[4,21-23] AI has made it easier to identify patterns and biomarkers that were previously challenging to detect, leading to more personalized and targeted treatments.[24] We have provided definitions of the key terminologies used in this article in Table 1.

Digital Pathology

DP involves digitizing tissue specimens, allowing them to be analyzed and shared electronically. DP uses complex imaging systems to capture high-resolution images of tissue specimens,

Table 1
Definitions of key terminologies

Digital pathology (DP)	A comprehensive term that includes various tools and systems to digitize pathology slides and associated metadata, as well as their storage, review, analysis, and supporting infrastructure.
Computational pathology (CP)	A branch of pathology that utilizes computational techniques to analyze methods of studying disease through patient specimens. It may involve using AI methods to analyze data and extract meaningful information from digitized pathology images.
Artificial intelligence (AI)	The field of AI aims to simulate human intelligence in machines, allowing them to perform tasks such as learning, problem solving, and decision making.
Machine learning (ML)	It is a branch of AI that programs computers to optimize a performance criterion using sample data or past experience. It uses the theory of statistics to build learning models.
Artificial neurons	These are the fundamental building blocks of artificial neural networks. It is a mathematical function that receives one or more inputs, applies a weighted sum, adds a bias term, and applies a nonlinear activation function to the result. The output of the activation function is then passed on to the next layer of neurons.
Artificial neural network (ANN)	A computational model inspired by the structure and function of biological neural networks in the brain. It is a network of interconnected artificial neurons that work together to process information and make predictions or decisions.
Neural network architecture	The architecture of a neural network refers to its structure, which is determined by the number and arrangement of its layers, the number of neurons in each layer, and the connections between the neurons.
AI training	The process of teaching an AI system to learn patterns from data and make accurate predictions or decisions. The training process involves feeding large amounts of data into the AI system and adjusting its internal parameters to optimize performance.
Supervised learning	AI training technique that uses annotated data, ie, each data point is associated with a known target value. Goal is to learn a mapping between inputs and outputs such that trained AI can make accurate predictions on new, unlabeled data. If learning involves lesser labeled data compared with unlabeled samples, it is weakly supervised learning.
Self-supervised learning	This technique of training AI does not require explicit data annotations. The AI learns to solve the given task using the inherent structure in the data as the supervisory signal.
Unsupervised learning	A learning technique for finding patterns, relationships, or structure in the data, such as clusters or groups of similar data points, without any knowledge of the ground truth. Unlike self-supervised learning, which uses a supervisory signal implicit in the data, unsupervised learning does not use any supervisory signal.
Computer vision (CV)	A field of AI that enables computers and systems to derive meaningful information from digital images, videos, and other visual data.
Natural language processing (NLP)	An area of AI that deals with a wide range of computational methods and techniques for analyzing, understanding, and generating natural language text.
Multimodal AI	Multimodal AI refers to AI models that involve multiple data modalities, such as vision (images) and language (text) and require AI to integrate information across data modalities.
Convolutional neural networks (CNNs)	CNNs are types of artificial neural networks commonly used for image and video analysis. CNNs are designed to automatically and adaptively learn spatial hierarchies of features from input images by using multiple convolutional layers, followed by pooling layers and fully connected layers.
Recurrent neural networks (RNNs)	RNNs are specialized for processing sequential data, such as text, speech, or time series. RNNs are designed to capture context and dependencies between the elements of a sequence.
Graph neural networks (GNNs)	GNNs are neural networks that process data with a graph structure. GNNs analyze relationships between objects (nodes) and their mutual relationships (edges) by iteratively using message-passing algorithms to update the features, allowing the network to capture the relationships between nodes in the graph.
Transformers	Transformers are neural networks that use a self-attention mechanism (or equivalently scaled dot-product) to capture relationships between input elements, especially in long sequences. They can process and learn from all data types, including images, text, and speech.
Foundation models	Foundation models are an emerging class of AI trained on a vast quantity of unannotated data at scale resulting in a model that can be adapted to a wide range of downstream tasks with only a handful of annotated examples. They use transformer architecture and are the workhorse of generative AI models.
Generative AI models	These are models specialized in generating new data similar to the training data, such as images or text. Examples include Bayesian networks, Generative Adversarial Networks (GANs), and foundation models such as ChatGPT, GPT-4, Stable Diffusion, and Dall-E 2.

which can then be viewed and analyzed on a computer screen.[25] DP improves the accuracy and efficiency of pathology diagnoses by allowing pathologists to access and share images remotely, collaborate with other experts, and integrate computer-aided analysis tools. With the advent of DP, the amount of data generated has increased exponentially, enabling the automation of time-consuming processes, such as segmentation and mitotic counting.[26] Public data archives, such as The Cancer Genome Atlas,[27] Clinical Proteomic Tumor Analysis Consortium,[28] and The Cancer Imaging Archive,[29] host pathology image data for multiple cancer sites. This is possible only because of DP and other advancements.

WSI is the technology that allows high-resolution digital images of entire microscope slides to be created and viewed on a computer screen. This process involves scanning glass slides containing tissue samples or other specimens using specialized digital scanners. WSIs can capture the entire slide at very high magnification, allowing users to zoom in and examine specific regions of interest in great detail.[30] WSIs are usually too large for contemporary computers to analyze directly, so they are tessellated into smaller tiles or patches, which serve as input for pathology AI workflows.[30]

Computational Pathology

Computational pathology combines DP with AI, ML, and other computational techniques to extract meaningful information.[30] Often interchangeably called “histomics,” “pathomics,” or “tissue phenomics,” computational pathology aims to develop

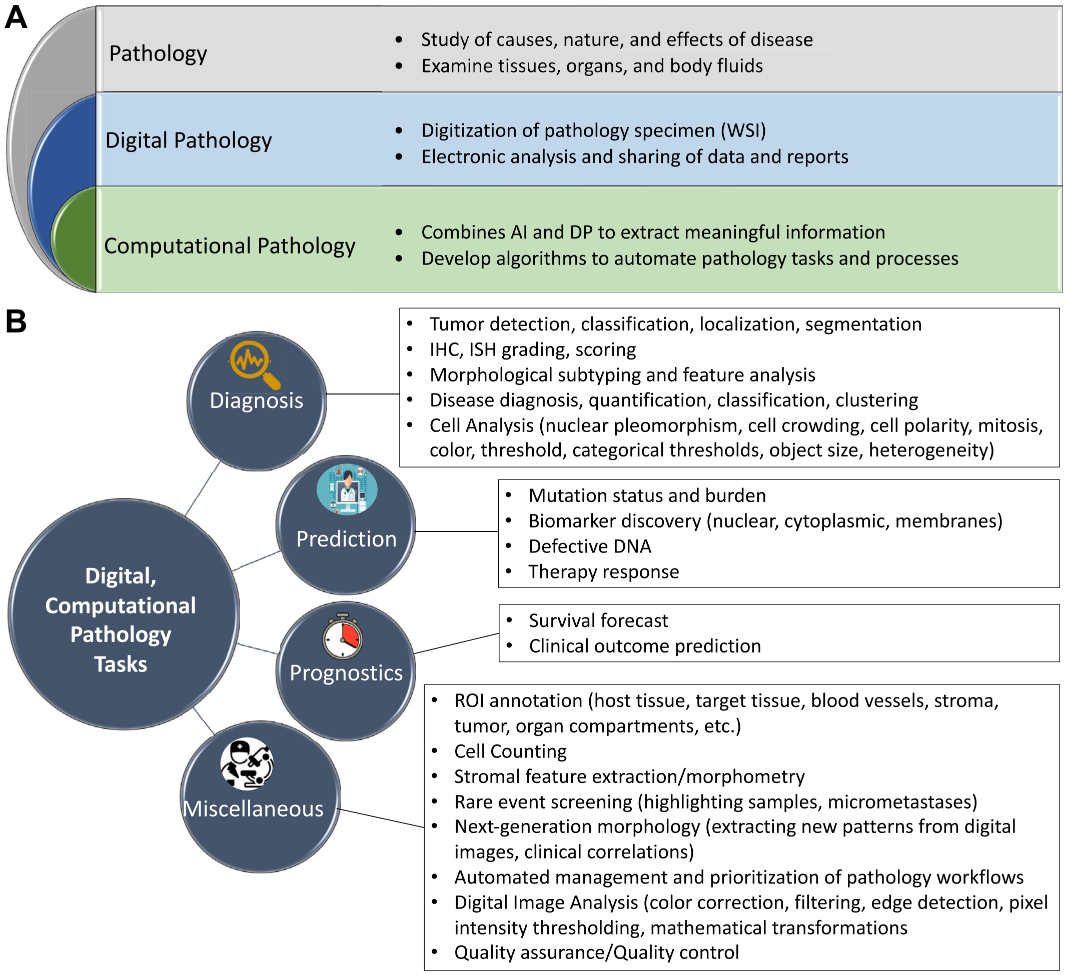

Figure 1.
Pathology, digital pathology, and computational pathology. Definitions and tasks are presented.

algorithms that can automatically detect and classify pathology images, predict disease outcomes, and identify new biomarkers for disease.[31] Computational pathology involves extracting many features from histopathology slides (called histomics) or pathology slides (called pathomics) and analyzing these features to relate to biological and clinical endpoints. Computational pathology also aims to standardize pathology diagnoses and reduce variability between pathologists.[30] Notwithstanding quality issues in DP,[32] computational pathology methods can perform well on tasks, such as classification, segmentation, and analysis of DP images, at times surpassing human-level performance.[33,34] The definitions of pathology, DP, and computational pathology are illustrated in Figure 1A, and their key tasks are illustrated in Figure 1B.

Classical Machine Learning in Digital Pathology

Classical ML consists of manually selecting informative features from the data by domain experts and then using these features for prediction, classification, or regression. The manual extraction and selection of features is also referred to as *feature engineering* using computer vision techniques based on morphology and texture, for instance. Classical ML has been extensively used in DP for image segmentation and classification[2] using Support Vector Machines, Random Forests, k-Nearest Neighbor, Decision Trees, and others.[2,35] A detailed review of the classical ML techniques in DP is presented in Gurcan et al[36] and Irshad et al.[37] Owing to the manual selection of usable and informative features, the applicability of classical ML methods is limited.[36,37]

Task-specific Artificial Intelligence in Digital Pathology

More recently, task-specific AI models based on artificial neural networks have been gaining popularity.[38-41] Artificial neural networks use stacked layers of artificial neurons to process large amounts of data and identify underlying patterns. The model selects a set of useful and informative features based on the assigned task without any human intervention. These models include CNNs, variants of RNNs, graph neural networks (GNNs), and Transformers, as illustrated in Figure 2.[2] We refer to these approaches as task-specific or narrow AI because of their limited scope. Also known as weak AI, they are incapable of general intelligence or human-like reasoning.[42] Developing a task-specific AI starts with selecting a particular task, followed by data collection and annotation. Finally, AI is supervised to learn patterns in the data by minimizing its prediction error. With the availability of digital slides and large computational power fueled by graphical-processing units (GPUs; electronic circuits responsible for graphics manipulation and output) and tensor-processing units (TPUs; Google's custom-integrated circuits used to accelerate ML

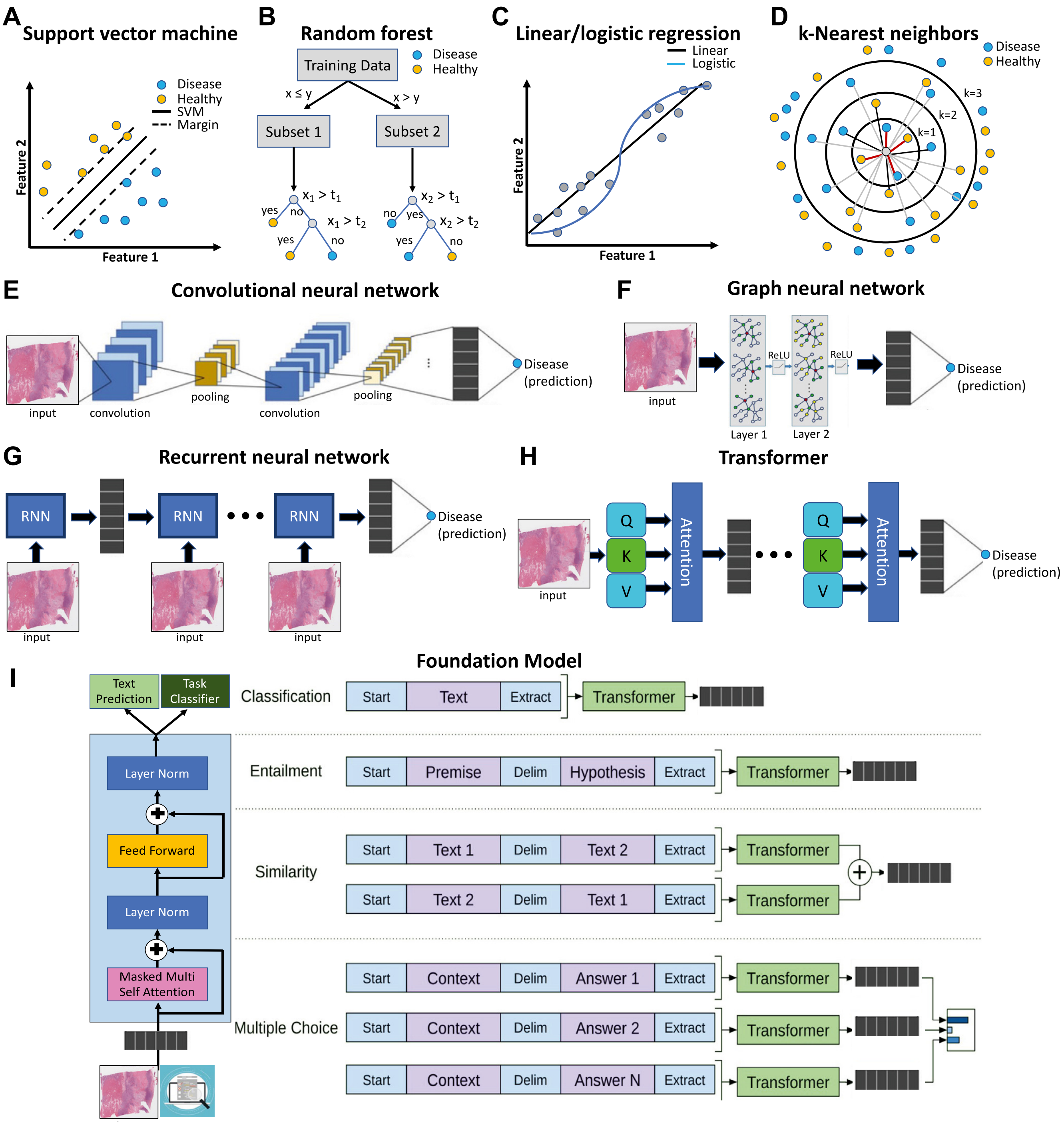

Figure 2.
A schematic layout of various machine learning (ML) algorithms and AI models used in digital pathology. The top row (A–D) highlights classical ML algorithms. Rows 2 and 3 (E–H) present task-specific AI models. The last row (sub-figure I) refers to foundation models, the brain behind generative AI, such as ChatGPT. We argue that the role of classical ML and task-specific AI is diminishing, being taken over by foundation models and generative AI. A set of large pathology-specific foundation models will sufficiently cover all digital pathology tasks as outlined in "Transformers, Foundation Models, and Digital Pathology."

workloads), artificial neural networks–based task-specific AI models have found a strong foothold in DP.[4,34,43] In the following discussion, we briefly introduce task-specific AI models and their essential components. In Table 2,[22,44-70] we present a non-exhaustive list of various categories of task-specific AI models used in DP. Interested readers are encouraged to explore the relevant works of interest.

Convolutional Neural Networks

CNNs are specialized artificial neural networks for processing image data. CNNs are designed to automatically learn and extract features from images, such as lines, edges, corners, and textures, through the convolution operation. Convolution involves sliding a filter over an input image and computing dot products between

Table 2
Summary of AI models used in digital pathology

DP Task	AI Model	Details	Ref
Diagnosis	CNN	MIDOG: Mitosis domain generalization challenge.	44
		Gleason grading and diagnosis of prostate cancer.	45
		Feature extraction to classify brain tumor grade.	46
		Mitosis detection in breast cancer.	47
		Segment nuclei in histology images using weakly supervised training.	48
	GAN	Nuclei segmentation on histopathology images.	49
	LSTM	4D medical image segmentation.	50
	GNN	Learn micro- and macrostructural features in HandE slides of breast cancer.	51
		Grading colorectal cancer in histology images.	52
		Classify healthy tissue from dysplastic gland areas in the colorectal cancer histology slides.	53
		Classify infiltrating ductal carcinoma (IDC) and ductal carcinoma in situ (DCIS) breast cancer and grade Gleason 3 and 4 prostate cancer.	54
Prediction	CNN	Disease outcome prediction in colorectal cancer.	55
	LSTM	Predicting sentiment, text categorization in records.	56
		Medical event prediction using a multichannel fusion of EHR data.	57
	GNN	Stratify prostate cancer using tissue microarrays.	58
	Transformers	Predicting RNAseq expressions from kidney WSIs using multiple-instance learning.	22
		Predicting biomarkers from histopathology slides in colorectal cancer.	59
Prognostics	CNN	Prediction of OS using Glioma multimodal data.	60
Misc	CNN	Correlations between true hypoxia fraction in histologic and the approximated fraction in MRI scans.	61
	GAN	Similarity between virtually stained images (generated by AI model) & histochemically stained images.	61
	LSTM	Medical image denoising.	62
		De-identification of medical text.	63
	Transformers	WSI representations using unsupervised learning.	64
Review	CNN	Deep learning in digital pathology for breast cancer.	65
	GNN	GNN-based methods in cancer pathology.	66,67
	Transformers	Transformers in the medical field.	66,68-70

AI, artificial intelligence; CNN, convolutional neural network; GAN, generative adversarial network; GNN, graph neural network; LSTM, Long short-term memory network.

the filter and the image pixels. The resulting features are used to classify or detect objects in the image. Based on the filter type, shape, size, and arrangement, various architectures of CNNs have been proposed.

Recurrent Neural Networks

RNNs process sequential data, such as speech, text, or time series. RNNs are designed to capture temporal dependencies in the data by maintaining a hidden state that is updated at each time step. The hidden state encodes information from previous time steps and provides context for the current time step. Long short-term memory networks and gated recurrent units are RNNs that help the model better capture long-term dependencies and avoid the vanishing gradient problem of RNNs using sequential data processing.

Graph Neural Networks

GNNs process graph-structured data, such as social networks, molecular data, and knowledge graphs.[66] GNNs are designed to capture local and global graph structures by aggregating information from neighboring nodes and edges. GNNs typically operate on a fixed-size local neighborhood around each node, allowing them to scale to large graphs. GNNs have shown promising results in various applications, including node classification, link prediction, and graph generation. GNNs have been used to analyze complex biological networks, drug discovery models, cell classification, tumor structures, and protein structures.[58,71,72]

Transformers

Transformers were initially introduced for language translation.[73] However, they have performed remarkably in various AI tasks, including computer vision and time series analysis.[13,74] Unlike RNNs, Transformers do not require that the sequential data be processed in order. Instead, they are designed to process variable-length input sequences (such as words in a sentence) without recurrent connections. Transformers use a self-attention mechanism that allows each piece of input (or token) to attend to other tokens in the sequence, capturing long-range dependencies.[73] Transformers have achieved state-of-the-art results in various natural language processing, computer vision, and graph-processing tasks.[73,74]

Artificial Intelligence-based Algorithms Used in Pathology

Interest in AI/ML-enabled medical devices has increased in recent years. The US Food and Drug Administration (FDA) has cleared more than 500 healthcare–related AI algorithms, 4 of which are for pathology.[75] Among them, 2 were introduced earlier, and the other 2 more recently. "PAPNET Testing System" was approved in 1995, and it was designed for rescreening negative Pap tests or as a primary screener.[76] "Pathwork Tissue of Origin Test" was approved in 2008, and it is a molecular diagnostic test developed to assist in diagnosing metastatic, poorly differentiated, and undifferentiated cancer.[76] "Tissue of Origin Test Kit FFPE" was approved in 2012, and it is an in vitro diagnostic to measure the degree of similarity between the RNA expression patterns in a patient's formalin-fixed, paraffin-embedded (FFPE) tumor and the RNA expression patterns in a database of 15 tumor types.[76] "Paige Prostate" was recently approved in 2021, and it is a software device to assist pathologists in the detection of foci that are suspicious for cancer during the review of scanned WSI from prostate needle biopsies prepared from H&E-stained FFPE tissue.[7,76] The PaigeAI prostate algorithm and the Pathwork digital and AI platform are the pioneering algorithms that have significantly impacted pathology practices by aiding in diagnosing and characterizing various diseases.

Pathology has a branch of anatomical pathology (AP) and clinical pathology (CP). The 4 algorithms mentioned above exclusively pertain to AP, where the focus lies on examining tissue samples for diagnosing diseases, such as cancer. It is worth mentioning that listing AI/ML-related algorithms in CP, which concentrates on the analysis of bodily fluids and laboratory tests, is beyond the purview of this specific review. Moreover, despite the noteworthy progress in pathology AI, to the best of our knowledge, there has not yet been a generative AI algorithm developed for pathology, AP, or CP. Generative algorithms can create new data or images, potentially aiding in generating synthetic samples for training and research purposes. Although such algorithms have seen success in other domains, their application in pathology, encompassing both anatomical and clinical aspects, has yet to be realized. The absence of generative AI in pathology presents a promising avenue for future research and exploration to unlock new possibilities and enhance the field's diagnostic and prognostic capabilities.

Limitations of Task-Specific Artificial Intelligence

Task-specific AI models have many limitations restricting their widespread use in DP:

1. Task specificity: task specificity refers to the fact that the trained AI performs well on a single task only, eg, grading cancer subtypes in an organ using H&E slides. A change in the number of grades, organ type, or cancer type (same organ) will render the model useless (significantly reducing its accuracy with low reliability) and will require model retraining.[39,77]
2. Distribution of the input data: task-specific AI models require the input data to have similar characteristics and follow the same probability distribution function of the input data (the mean, standard deviation and range of the pixel values of WSI pixels).[39,78] Adding natural or adversarial noise may significantly reduce AI's performance.[79] AI models are known to be fragile in the presence of noisy inputs, subtle changes in the data, or adversarial attacks.[39,40,79] These AI models cannot generalize to changes in data resulting from various common reasons, eg, hardware, software, firmware upgrades in scanners, changes in the staining quality or the protocol, shifts in population demographics (eg, a different geographic region), and changes in data patterns due to new diseases, such as COVID-19.[41,77] New representative data must be collected and annotated for each changing scenario to retrain (fine tune) the AI models to be current and accurate.
3. Requirement of large annotated task-specific datasets: the task-specific AI requires large annotated datasets for training. The success of these models depends mainly on the availability of large, task-specific annotated datasets. This requirement stems from the data-driven nature of these models, which learn to identify informative features from data without needing domain experts to engineer data features.[80] By leveraging vast amounts of annotated, independent, and identically distributed data, models can uncover hidden patterns and subvisual features that may be difficult for humans to detect. However, obtaining a large annotated dataset remains a critical challenge for AI models in DP. These AI models cannot directly benefit from large amounts of unannotated datasets, eg, WSIs, pathology reports, and clinical notes, and may require techniques, such as weakly supervised learning, unsupervised learning, self-supervised learning, transfer learning, and continual learning.[81,82]
4. Single data modality: the task-specific AI models are generally restricted to processing one data modality only. Incorporating information from other modalities, eg, the patient's medical data from medical records, omics data, or radiographs, into AI decision making is generally not straightforward.[66] Recently, some research efforts have focused on creating AI models that can process multimodal data to improve their predictive accuracy with moderate success.[72,83,84]
5. Knowledge accumulation: the recent success of ChatGPT has shown that creating an internal general-purpose knowledge base is essential for successful and robust AI models.[16,17] ChatGPT has a central repository of information created during model training using 570 GB of data from books, web-based text, Wikipedia, articles, and other online writings.[15,16,18] There is no precedence for creating such models in DP, medical imaging, or any area of medical data processing. Task-specific, narrow versions of AI models are built by individual academic laboratories or industries that do not contribute to reusable knowledge accumulation.[66]
6. Transparency and reproducibility: transparency and reproducibility of AI models are a challenge that undermines the enormous potential of applying such methods to complex tasks. The lack of sufficient details regarding Methods and the unavailability of algorithm/code in a published work by the Google Health team on breast cancer screening[85] was recently raised.[86,87] The research community is gradually transitioning to open-access, reproducible, and transparent methodologies.
7. Explainability: the explainability of AI refers to the challenge of understanding how and why an AI makes a particular decision or prediction.[88] Although AI can make accurate predictions or decisions, they often do so in ways that are opaque or difficult to understand for human beings. This lack of transparency can be problematic in scenarios where decisions made by AI have significant real-world consequences. Interpretable or explainable narrow AI models with attribution maps produce results humans can easily understand and interpret. However, these approaches come at the cost of reduced accuracy or increased model complexity.[89]

Foundation Models and Generative Artificial Intelligence

The 2020s are witnessing the rise of foundation models – large AI models pretrained using unannotated multimodal datasets

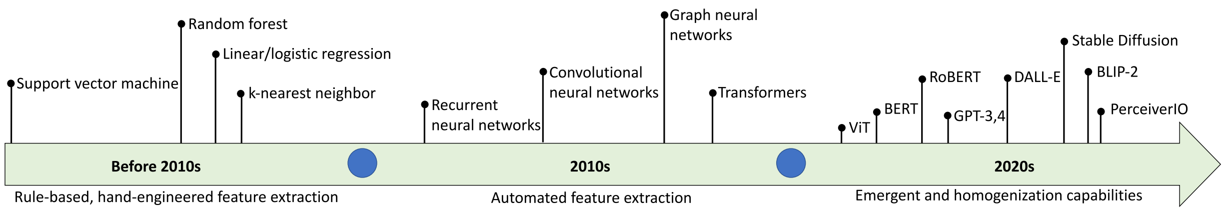

Figure 3.
The evolution of machine learning models used in digital pathology. The 2020s are witnessing the rise of generative AI based on foundation models. Soon, foundation models and generative AI will become the preferred computational approaches in digital pathology.

(Figs. 2 and 3).[13] A trained foundation model can be adapted (or fine-tuned) to perform different tasks using limited annotated examples, much less than required to train task-specific AI. In the following, we present our perspective on how foundation models and generative AI that use these models can transform the DP laboratory workflow. The pathology-specific foundation models can be created and fine-tuned to serve as a pathologist's expert assistant by performing quantitative image analysis for diagnosis, prognosis, disease grading, and prediction. It can then generate pathology reports based on the presented imaging data and converse with the pathologist to justify the findings presented in the generated reports.

Foundation Models

The term "foundation models" was initially coined by Bommasani et al[13] to describe recently proposed models that have led to a paradigm shift in AI model design, development, and deployment processes. Foundation models are huge models trained at scale using comprehensive unannotated data (possibly multimodal). Two relevant attributes of foundation models are their size, which refers to the number of learnable parameters or weights, and the number of compute operations quantified using floating point operations at the model testing or inference stage. Foundation models generally have billions or trillions of learnable parameters and billions of floating point operations.[16,17,90-94] The unannotated datasets may consist of billions of words (or tokens) and images from the Internet without any labels assigned by human operators.[92] Foundation models leverage the existing concepts of pretraining, transfer learning, and unsupervised and self-supervised learning. However, their essence lies in *scaling* because of the following 3 factors: (1) the introduction of Transformer architecture[73] that supports training models with the number of learnable parameters in billions or trillions, (2) the availability of thousands of GPUs, and (3) availability of massive training datasets that can reach billions of tokens for natural language processing and hundreds of millions of images for computer vision tasks.[13,73]

Recently, a host of foundation models have been trained for language, vision, and joint language-vision (multimodal) tasks and shared via GitHub (https://github.com/trending) and Hugging Face (https://huggingface.co). Some of the remarkable works include BERT and RoBERTa in language processing,[93,94] Vision Transformers for image-processing tasks,[95] Mask2Former, OneFormer, and ClipSeg for image segmentation,[96-98] Perceiver IO for multimodal (text, images, audio, and video) problems,[99] ViperGPT for answering visual queries using code generation,[100] LLaVA for visual instruction tuning,[101] and BLIP-2 for image captioning, visual question-answering, and chat-based prompting.[102]

Characteristics of Foundation Models

Some distinguishing characteristics of foundation models are summarized as follows:

- *Expressivity* is the ability of foundation models to learn, capture, and represent the relevant information from data.[13] Foundation models are more expressive than their task-specific AI models as they exclusively use the Transformers architecture, which learns long-range relationships and higher-order interactions in the data using a self-attention mechanism.[73] There exists a trade-off between the model's expressivity and its efficiency. Increasing the model size may increase its expressivity at the cost of reduced efficiency.[13] Recently proposed foundation models, such as Perceiver IO and GANformer attempt to offer a balance between efficiency and expressivity.[99,103]
- *Scalability* refers to the ability of a foundation model to efficiently consume large amounts of data.[13] With the ever-growing availability of data from diverse sources, the foundation model needs to be capable of further scaling while overcoming the challenges of failure and catastrophic forgetting.[41,81]
- *Multimodality* is the ability of the foundation model to learn relations among various modalities of the data.[13] Humans perceive knowledge through processing multimodal data. GPT-4 is a multimodal foundation model.[15] Other multimodal models include CLIP,[104] ALIGN,[105] SimVLM,[106] Flamingo,[107] CoCa,[108] and CONCH.[109]
- *Compositionality* is the ability of a foundation model to generalize to new tasks and contexts.[13] Compositionality helps foundation models achieve out-of-distribution generalization and perform in-context learning.[13,110]
- *Emergence* is the characteristic introduced by scaling the Transformer architecture with large datasets and computational resources.[13,73] Emergence means that the behavior of the trained AI model is implicitly induced rather than explicitly constructed.[13,110] In-context learning is an example of emergence in foundation models[110,111]
- *Homogenization* is also introduced by scaling and refers to the consolidation of methodologies for building AI models across a wide range of applications.[13] For example, almost all language-processing tasks can be performed by a single large language model, eg, BERT,[93] GPT,[15,17-19] T5,[90] or many others.[112]
- *Transfer learning, adaptation, and fine tuning* are the defining characteristics of foundation models.[13,111,113] These characteristics imply that the skills that AI may learn from one task will often transfer to new tasks. A foundation model may adapt to the new tasks without the need for any annotated examples, referred to as zero-shot learning. When a

few examples are used to fine tune the AI, we call this few-shot learning.[13] Generally, all foundational models are pretrained using unannotated datasets and later adapted using small annotated datasets for specific downstream tasks. A recent survey reviews the various pretraining methods used in deep learning and foundation models on medical data.[114]

- *In-context learning* is the ability of a trained foundation model to learn a new task or correct itself using demonstration and without updating the model's parameters, which is usually done via gradient descent algorithm.[17,111,115] In-context learning is a scale-enabled emergent ability that allows foundation models to generalize to new tasks without having to retrain the AI model again. GPT-2, a relatively small model having 1.5 billion parameters, did not permit in-context learning.[18] It was GPT-3 with its 175 billion parameters that exhibited in-context learning.[17] However, in-context learning introduces the necessity for *prompt engineering*, ie, finding the most appropriate prompt to allow AI to solve the task at hand.[111,115] A prompt is a piece of text, image, or symbols inserted in the input of AI so that the given task can be reformulated as the original task for which the model was trained.[17,110,111]

Types of Foundation Models

Foundation models are an emerging area of AI that has shown great promise, eg, ChatGPT, GPT-4, DALL-E 2, and Stable Diffusion are foundation models that can generate impressive text and images, provide concise summaries of large datasets, and help analyze unstructured data efficiently.[15,16,116,117] These models can be further divided into large language models that tackle natural language-processing tasks and vision-language models that handle multimodal learning jointly from images, text, and other data sources.

Large Language Artificial Intelligence Models

Large language models can handle various natural language-processing tasks, including text generation, natural language understating, sentiment analysis, question answering, information retrieval, reading comprehension, commonsense reasoning, natural language inferences, word sense disambiguation, and others.[13] With the introduction of *word embeddings*, where each word in a sentence was associated with a context-independent vector of real numbers,[118] the natural language-processing field has seen considerable progress.[93] Following the success of word embeddings, autoregressive language models were proposed to employ self-supervised or weakly supervised learning to predict the next word in a sentence given the previous words.[93] Autoregressive models such as GPT, ELMo, and ULMFiT use the context of the words in representation embeddings.[19,119,120] The Transformer architecture enabled self-supervised learning at scale resulting in models, such as BERT, GPT, GPT-2, GPT-3, GPT-4, LLaMA, T5, and BART[15,17-19,90,93,94,121,122] Most of these industry-sponsored models are not open-source for researchers.[15,16] Recently, BLOOM, a 176Bparameter open-access language model, was developed with the collaboration of hundreds of researchers.[91] BLOOM is a decoder-only Transformer language model trained on the ROOTS corpus, a dataset comprising hundreds of sources in 46 natural and 13 programming languages (59 in total).[91] A comprehensive review of large language models is out of the scope of this article. For a comprehensive review of the large language models, please refer to review papers and blogs,[112]

Vision-Language Artificial Intelligence Models

Vision-language AI can learn to perform various tasks involving images (or videos) and corresponding natural language text.[92,112] The vision-language models are one step closer to how humans perceive the world, learn about it, and execute various tasks in it.[92,123] In the following, we describe 2 types of imaging analysis tasks that vision-language models can perform:

- *Image-text mixed tasks:* These tasks reside at the intersection of natural language processing and computer vision fields and consist of extracting information from images and natural language text and finding the relationships and patterns to link text and images.[92,124] Image captioning, visual question answering, visual dialog, image or text retrieval given text or image, visual grounding, and image generation are a few image-text tasks undertaken by these foundation AI models.[92,125] Visual question-answering tasks typically require a more detailed understanding of the image and complex reasoning than a system producing image captions.[125] The recent foundation models in image-text tasks include Contrastive Language-Image Pretraining (CLIP), A Large-scale ImaGe and Noisy-Text Embedding (ALIGN), SimVLM, Florence, Flamingo, CoCa, Clinical-BERT, and Contrastive learning from Captions for Histopathology (CONCH).[104-109,126]
- *Image-processing tasks:* Image classification, object detection, and segmentation are the core visual recognition tasks in the field of computer vision. Traditionally, these tasks were considered pure vision problems without needing to include language information while learning these tasks. However, CLIP and ALIGN models showed that language supervision could play an essential role in pretraining vision-language that can do various visual recognition tasks with zero-shot learning.[104,105] CLIP and ALIGN use noisy image-text data from the Internet to enable large-scale pretraining of vision encoders. The state-of-the-art foundation models include the following: (1) image classification − UniCL, CLIP, and ALIGN,[104,105,127]; (2) object detection in a given image − ViLD, RegionCLIP, GLIP, Detic, PromptDet, OWL-ViT, OV-DETR, and XDERT,[128-134]; and (3) segmentation of different objects in a given image − LSeg, OpenSeg, CLIPSeg, MaskCLIP, DenseCLIP, and GroupViT.[98,135-139]

Training Foundation Models and Generative Artificial Intelligence

Foundation models employ 2 key techniques in training: self-supervised learning and generative training. The true potential of the enormous quantity of unannotated data is only possible with supervised learning, without the need to create annotations or labels using human effort. Examples of such data include the following: (1) text, images, and videos available online or (2) medical records, diagnostic imaging, molecular data, and histopathology WSIs available in hospital databases. During the training of foundation models, the supervision signal is determined by the context of the input data, eg, the BERT language model is trained to predict randomly removed words from sentences or fill in the blanks.[93] Sometimes, the models are shown plausible and implausible pairs of images and corresponding texts. Thus, the model learns to associate image features with their correct text description.[104] This perspective generalizes the traditional close-set classification AI models to recognize unseen concepts in real-world applications, such as open-vocabulary object detection.[92]

The generative training methods help foundation models learn the joint or conditional probability distributions over training input data.[13] That is, the trained foundation model will be able to accurately generate the input data pattern similar to the ones used for training it. Generative training is performed using 1 of the 2 techniques: (1) denoising or (2) autoregressive. During the training of the denoising models, the input is corrupted with noise, and the model is expected to produce noise-free input patterns.[90] The autoregressive models, after training, can generate the input data piece by piece, iteratively predicting the next element in a sequence given the previous elements.[140]

Challenges and Limitations of Foundation Models

Developing foundation models requires massive datasets, computational resources, and technical expertise.[13] Owing to their massive size, it may not be possible to fit the parameters of a foundation model in the memory of the largest GPU or a single computer. For example, a recent large language model shared by Meta AI, LLaMA, has 65 billion parameters and was trained using 1.4 trillion tokens.[121]. The enormous computational operations inside foundation models can result in unrealistically long training and inference times. Foundation models require specialized software, hardware, and inference algorithms to train and use.[141]

"Hallucination" is a known limitation of generative AI, which refers to mistakes in the generated text or images that are semantically, syntactically, or visually plausible but are, in fact, incorrect, nonsensical, and do not refer to any real-world concepts.[15,142,143] The accuracy and integrity of the generated text and images may be challenging to establish using factual data from verified sources.[143] One possible solution is to use an engineered system, such as *Bing Chat*, which also generates links to the actual websites, articles, and reference material (https://www.bing.com/). In some cases, the generative AI models can identify their own mistakes.[142] Furthermore, the generative models are sensitive to the form and choice of words, referred to as the "prompt." A prompt may consist of text, image(s), or symbol(s) inserted in the input of generative AI so that the given task can be reformulated as the original task for which the model was trained.[115,142] The future generative AI models may be less sensitive to the precise prompt. However, the current models need "prompt engineering" to produce the best results[115,142] Therefore, effectively using a generative AI may require engineering an appropriate prompt by the human user. Foundation models and generative AI also face other challenges similar to task-specific AI models, including explainability, robustness, and trustworthiness.[13,39,41,78,79,88,143-145]

Transformers, Foundation Models, and Digital Pathology

This section presents recent work from the literature focused on using Transformers (the core component of foundation models) in DP. We focus on the work where a single AI model based on Transformer architecture is trained using large, diverse datasets to perform multiple tasks. Later, we present our perspective on the potentially transformative role of foundation model-based AI in DP. Because of foundation models' strong adaptation and scalability properties, they can be effectively trained once and modified infinite times to suit various DP tasks. We also present our perspective on the trustworthiness and acceptability of generative AI and foundation models by pathologists. Figure 4 presents a prospective framework for utilizing foundation models and generative AI for various pathology tasks.

Transformers architecture has recently been modified to consume high-resolution gigapixel WSI data.[146] The authors used a self-supervised hierarchical learning mechanism on 33 cancer site data having approximately 105 million pathology images to predict 9 slide-level tasks, including cancer subtyping, survival, and unique morphologic phenotypes.[146] Although molecular procedures and analyses have led to remarkable discoveries, they are usually time consuming, expensive, and require multiple tissue samples. Transformer-based foundation models can address these challenges by predicting the bulk RNA-seq directly from the whole-slide images.[22] Transformers-based foundation model, CONCH, has shown state-of-the-art performance on multiple tasks including histology image classification, segmentation, captioning, and text-to-image and image-to-text retrieval using task-agnostic pretraining on 1.17 million image-caption pairs.[109] Similarly, attention-based multiple-instance learning has accurately predicted biomarkers from cancer pathology slides in a self-supervised learning setting.[59] The authors showed the performance of an attention-based multiple-instance learning framework for predicting microsatellite instability and mutations in BRAF, KRAS, NRAS, and PIK3CA in colorectal cancer pathology slides.[59] To address the interpretability challenge of the AI model's decisions, a probabilistic perspective on attention-based multiple-instance learning on WSI data has outperformed previous methods in matching the pathologists' annotations.[147] Such prefoundation models can be scaled to predict biomarkers directly from the histopathology slides belonging to pan-cancer sites.[148] The Transformer model pretrained on a large publicly available pathology dataset can be fine-tuned under a weakly supervised contrastive learning scheme on smaller datasets. Wang et al[149] have shown that such a training framework can outperform the state-of-the-art WSI classification on 3 different tasks. For the multimodal medical data analysis, modality coattention Transformers have been shown to outperform other methods in survival predictions by fused learning on WSI data and genomic sequences.[150] Moreover, Transformers are far more robust to adversarial attacks and perturbations in DP than CNNs because of the more robust latent representation of clinically relevant information.[79] The performance and robustness of Transformer-based models in various tasks and modality settings have shown the prospective utilization of a single foundation model for large-scale rollout involving multiple tasks.

Given the strong support for compositionality and multimodality and the modular nature of the foundation models, image and language models can be combined to share their learned representations as a larger foundation model. Thus, a Transformer trained to interpret WSIs can be combined with a trained language-generation model (eg, GPT) to create a vision-language model. Such a model will interpret and analyze WSIs and generate text reports based on the analysis. The same model can be augmented to annotate relevant areas on the input image to support its findings in the generated report. Finally, a conversational component can be added to allow the model to interact with the pathologist to answer their question about the model's output.

The authors believe that a multimodal pathology foundation model capable of processing WSIs and natural language can be

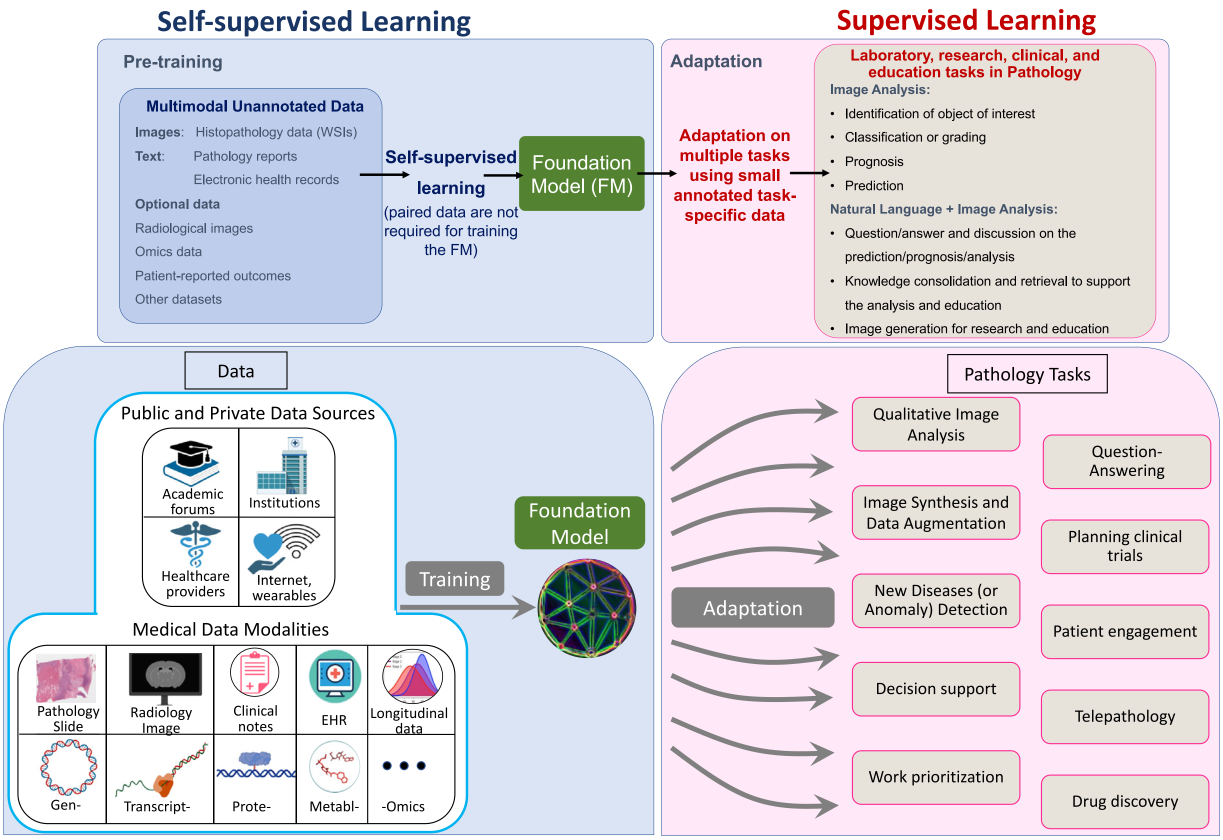

Figure 4.

A prospective schematic layout of using foundation models and generative AI for various digital pathology tasks is presented. In our view, other data modalities, eg, diagnostic radiology or molecular data, if available, can be combined with pathology data in the future to improve model performance.

created using data available in the public domain, such as the National Cancer Institute's The Cancer Genome Atlas for genetic data, Clinical Proteomic Tumor Analysis Consortium for proteomics data, and The Cancer Imaging Archive for imaging data.[27-29] The base model can be trained with pan-cancer datasets and later fine-tuned for various organs, cancer types, and use cases with only a few task-specific annotated examples. The base pathology foundation model can be shared with the community, eliminating the need to collect data, annotate, and train AI models from scratch for each use case. A recent synopsis explores AI techniques for multimodal data fusion and disease association discovery in oncology data.[151] Quantifying patterns across 17,355 H&E-stained slides from 28 cancer types through deep learning accurately classified cancer types and correlated learned features with numerous recurrent genetic aberrations across considered cancer types.[152] In the following, we build on the idea of training and sharing a base pathology foundation model that can be adapted for research, clinical, laboratory, and educational use cases in DP.

Qualitative Image Analysis

A trained foundation AI model can be adapted for various pathology image analysis tasks. The adaptation may not require any annotated data (zero-shot learning) or may require only a handful of samples (few-shot learning). Examples include the following: (1) separating the different types of cells in an image and identifying the regions of interest, (2) identifying and counting the number of cells in a given image, (3) categorizing cells into different types based on their appearance and features, (4) identifying the presence and extent of cancerous tissue in an image, (5) assessing the severity and extent of a disease by grading and staging tissue samples, (6) measuring the number of specific proteins or molecules in a tissue sample to determine their potential as biomarkers for disease, (7) predicting the likelihood of disease progression or patient outcome based on the analysis of tissue samples, or (8) immunohistochemistry scoring.

Apart from adapting the base pathology foundation model to various imaging tasks, we can use the same model for analyzing different types of stains, images from different scanners, and noisy slides containing different artifacts. Foundation models can leverage the multisite cytology data (cervix, kidney, breast, lung, thyroid, bladder, bone marrow, skin, and others) to perform various downstream tasks such as malignant cell classification, slide-level stratification, cell location in cytologic smears, and cell component identification.[153]

Going one step further, the image analysis performed by the AI can be internally fed to the generative AI, allowing pathology report generation directly from the image.[154,155] Some sample pathology reports generated using ChatGPT (March 14 update) are presented in Figure 5A–C. These reports were generated by text prompt only without providing any image to ChatGPT as it cannot process image data.

AI supported by large models can reduce pathologists' workload and interrater variability while improving the quality and consistency of pathology reports.[148] The image analysis and report generation pipeline can serve as the "first pair of eyes" and potentially help pathology laboratories with workload and

A

Brain Surgery Report

Patient Name: John Doe
Date of Birth: 01/01/1960
Date of Procedure: 03/21/2023
Procedure: Craniotomy for Resection of Brain Tumor

Clinical History:
The patient presented with a several-month history of progressive headaches, nausea, and dizziness. Imaging studies revealed a 4 cm mass in the left frontal lobe of the brain, which was suspicious for a malignant tumor. The patient underwent a craniotomy for resection of the mass.

Gross Description:
Received in formalin is a 4 x 3.5 x 3.0 cm piece of tan-pink soft tissue. The specimen is labeled with a suture indicating the superior aspect of the tumor. The tumor is well-circumscribed and has a soft, friable consistency. Multiple sections are submitted for histologic evaluation.

Microscopic Description:
The histologic sections show a high-grade astrocytoma, consistent with a grace IV glioblastoma. The tumor is characterized by pleomorphic, hyperchromatic nuclei with frequent mitotic figures and areas of necrosis. The tumor cells infiltrate the adjacent brain parenchyma and extend into the subarachnoid space. There is no evidence of vascular invasion.

Immunohistochemical Stains:
Immunohistochemical staining shows positive expression of glial fibrillary acidic protein (GFAP) and negative expression of IDH1 R132H mutation.

Final Diagnosis:
- Left frontal lobe brain tumor: Grade IV glioblastoma (high-grade astrocytoma).

Comments:
The findings are consistent with a diagnosis of grade IV glioblastoma, which is a high-grade astrocytoma. The tumor is well-circumscribed and has a soft, friable consistency. Multiple sections were submitted for histologic evaluation, and there is no evidence of vascular invasion. Further treatment planning, including radiation and chemotherapy, is recommended.

Pathologist:
Dr. Sarah Lee
MD, Board Certified in Anatomic and Clinical Pathology

B

Breast Biopsy Report

Patient Name: Jane Smith
Date of Birth: 05/10/1975
Date of Procedure: 03/15/2023
Procedure: Excision of Left Breast Mass

Clinical History:
The patient presented with a palpable mass in her left breast. Imaging studies revealed a suspicious mass, and a biopsy was performed, which showed invasive ductal carcinoma. The patient underwent a left breast lumpectomy for definitive treatment.

Gross Description:
Received in formalin is a left breast lumpectomy specimen measuring 3.5 x 2.5 x 2.0 cm. The specimen is oriented with sutures indicating the superior, inferior, medial, and lateral margins. The inked margins are negative for tumor. The mass is well-circumscribed and measures 1.5 cm in greatest dimension. The tumor is firm, white, and tan.

Microscopic Description:
The lumpectomy specimen shows a well-circumscribed, invasive ductal carcinoma measuring 1.5 cm in greatest dimension. The tumor is predominantly composed of small, uniform cells arranged in glandular and solid patterns. There is no evidence of lymphovascular invasion. The surgical margins are free of tumor.

Immunohistochemical Stains:
Immunohistochemical staining is positive for estrogen receptor (ER), progesterone receptor (PR), and HER2/neu.

Final Diagnosis:
- Left breast lumpectomy: Invasive ductal carcinoma, well-circumscribed, measuring 1.5 cm in greatest dimension
- Tumor markers: Positive for ER, PR, and HER2/neu.

Comments:
The findings are consistent with a diagnosis of invasive ductal carcinoma of the left breast. The tumor is well-circumscribed and and measures 1.5 cm in greatest dimension. The surgical margins are free of tumor. Further staging workup and treatment planning are recommended.

Pathologist:
Dr. John Smith
MD, Board Certified in Anatomic and Clinical Pathology

C

Pancreatic Biopsy Report

Patient Name: John Doe
Date of Birth: 01/01/1960
Date of Procedure: 03/01/2023
Procedure: Pancreatic Biopsy

Clinical History:
The patient presented with persistent abdominal pain, loss of appetite, and weight loss. Imaging studies revealed a pancreatic mass suspicious of malignancy. A biopsy was performed for further evaluation.

Gross Description:
Received in formalin is a core biopsy measuring 1.2 cm in length and 0.3 cm in diameter. The specimen is tan-white and firm.

Microscopic Description:
The biopsy shows infiltration of pancreatic tissue by poorly differentiated adenocarcinoma. The tumor cells are arranged in glandular and solid patterns and have large, irregular nuclei with prominent nucleoli. Mitotic activity is high. No normal pancreatic tissue is identified in the biopsy specimen.

Immunohistochemical Stains:
Immunohistochemical staining is positive for CK7 and CA19-9 and negative for CK20, CDX2, and TTF-1.

Final Diagnosis:
- Pancreatic Biopsy: Poorly differentiated adenocarcinoma
- Tumor markers: Positive for CK7 and CA19-9 and negative for CK20, CDX2, and TTF-1.

Comments:
The findings are consistent with a diagnosis of poorly differentiated adenocarcinoma of the pancreas. Further staging workup and treatment planning are recommended.

D

Pancreatic Biopsy Description

Biopsy Site: Pancreas
The biopsy was taken from the pancreas, which is an organ in the abdomen involved in digestion and producing hormones like insulin.

Clinical Information:
The patient came to the doctor with belly pain and yellowing of the skin and eyes (jaundice). Imaging tests found a mass (abnormal growth) in the pancreas. A biopsy (taking a small tissue sample) was performed for further examination.

Macroscopic Description:
The biopsy sample was preserved in a chemical called formalin. It's a small, firm, tan-white piece of tissue about 1.5 cm long and 0.5 cm wide.

Microscopic Description:
Looking at the tissue under a microscope, it contains normal pancreas components, including acinar cells (which produce digestive enzymes), ductal cells (forming ducts that transport the enzymes), and fibrous stroma (connective tissue). However, there's also a well-developed cancer called adenocarcinoma invading the nearby pancreatic tissue. The cancer cells form gland-like structures and can be seen moving into the neighboring ducts. The cancer cells are also found invading nerves (perineural invasion). The tumor is moderately differentiated, meaning it somewhat resembles normal tissue. There are 10 cell divisions (mitoses) seen in each square millimeter of the tumor, which helps determine how fast the tumor is growing.

Immunohistochemical Analysis:
Special tests called immunohistochemistry were performed, showing the cancer cells produce specific proteins (cytokeratin 19 and carcinoembryonic antigen), confirming the diagnosis of pancreatic adenocarcinoma.

Diagnosis:
1. Pancreatic adenocarcinoma, moderately differentiated: This is a type of pancreatic cancer that is moderately similar to normal tissue
2. Mitotic count of 10 per square millimeter: This is a measure of how fast the tumor cells are dividing and growing.
3. Perineural invasion present: Cancer cells are invading the nerves, which is an important finding that may affect treatment and prognosis.

Figure 5.
Three different pathology biopsy reports generated by ChatGPT are presented in (A–C). The AI was prompted using the following text: "Generate a sample pathology report for [organ name]." The right bottom image (D) represents a lay-person description of the pathology report generated by ChatGPT using the pathology report presented in (C).

staffing shortages. AI model adaptation and fine tuning allow it to learn from its mistakes and update itself under the guidance of the pathologist(s). Thus, promoting the AI from just a data processing pipeline to an assistant who will, over time, learn to help the users perform their tasks efficiently.

Image Synthesis, Denoising, and Virtual Staining

Publicly available generative AI has yet to show plausible pathology image-generation capabilities. It has been recently shown that, despite being state of the art at the time of assessment, the text-guided diffusion model (GUIDE) lacked a good depiction of the style and contents of medical images.[156] However, we argue that there is enough pathology image data in the public domain to train pathology image-generation models using GUIDE, Stable Diffusion, or Dall-E 2, as the starting point. A well-trained pathology image-generation AI can address various research and clinical challenges including the following: (1) denoising digitized slides to remove noise and artifacts and normalize the image to a standard color and tone, effectively making the task of image analysis pipeline easy and less prone to error; (2) virtual staining – generating images with different staining techniques without requiring additional physical samples, helping pathologists compare and contrast the effects of various stains and facilitate more accurate diagnoses[32]; (3) super-resolution imaging – using AI synthesis techniques to generate super-resolution images from low-quality and noisy digital slides, aiding pathologists in examining fine details and structures that may not be visible in the original images due to noisy or erroneous digitization process[5]; (4) simulating disease progression – generating images simulating the progression or regression of pathologic conditions, thus providing pathologists with a better understanding of disease evolution and enabling more informed treatment planning; (5) education and training – creating diverse and realistic examples for educational purposes, thus helping trainee pathologists gain experience in diagnosing a wide range of conditions and improving their diagnostic skills without relying on actual patient samples[157]; and (6) synthesizing images to study the effects of various factors on disease presentation, such as genetic mutations, environmental factors, or treatment options, thus contributing to a better understanding of disease mechanisms and the development of more effective therapies.

Detecting Zebras (New Disease Identification)

Foundation models can be adapted to identify deviations from the norm, which may indicate potential anomalies, such as abnormal cell structures, lesions, or other abnormalities. Anomaly detection of finding zebras goes beyond regular tasks of identifying disease subtypes or grading. This use case aims to identify and report patterns never seen in the training data to improve the accuracy and efficiency of identifying unusual or unexpected events. Transformer-based models can learn to directly predict the bulk RNA-seq from WSI and simultaneously output the WSI representation.[22] Such models can augment pathologists' expertise and provide more accurate and timely diagnoses.

Patient Engagement

Generative AI can help pathologists, who are the "doctor's doctor," engage directly with the patients by bringing them to the front line without additional time or resource commitment. Language models can generate more approachable and accurate descriptions and explanations of the pathologist's findings for the patients. Image-generation models can create annotated images to depict the disease visually. In Figure 5D, we present the description of a biopsy report generated by GPT-4. The text is aimed to explain the pathology biopsy report (presented in Fig. 5C) to a nonmedical person. In addition, they can educate the patient about the disease entity just diagnosed by the pathologist and possible treatment options.

Education and Training

Pathology education is currently powered and driven by virtual and digital transformations and is swiftly adapting to the advancements offered by AI.[157] Generative AI can retrieve and integrate knowledge from various sources, such as textbooks, and scientific articles, providing a comprehensive view of the state of knowledge. Pathology-focused ChatGPT-like models can answer pedagogic questions quickly, such as the definition of terms or recent advancements reported in the literature.

Conversational AI, such as ChatGPT and its variants, can solve higher-order reasoning questions. ChatGPT has the comparative relational level of accuracy in pathology, as noted by the responses shown in Figure 5. Hence, students and academicians have the opportunity to adapt to this emerging technology and use it for solving reasoning-type questions. Further evolution of such conversational tools needs to be critically analyzed by specialists, such as pathologists, for their efficacy and acceptability.

Artificial Intelligence-Driven Standardization in Digital Pathology Workflow

Foundation models and generative AI can help standardize DP by addressing various aspects of the diagnostic process, such as image acquisition, analysis, interpretation, and reporting.

1. Image preprocessing and normalization: AI models can correct for inconsistencies in image acquisition, such as variations in lighting, staining, and scanning parameters. By automatically adjusting for these factors, AI can ensure that images are more consistent and comparable across laboratories and scanners.
2. Automated feature extraction and quantification: AI-based tools can extract and quantify relevant features in images in a standardized and reproducible manner. This can include cell counting, morphologic measurements, and biomarker quantification, reducing the variability that may arise from manual or semiautomated methods. A human operator will need to approve AI-generated features.
3. Computer-aided diagnosis: AI-driven algorithms can provide a second reader opinion or decision support for pathologists, reducing diagnostic variability and errors. By learning from large datasets and incorporating best practices, AI can help standardize the diagnostic process and improve the overall quality of diagnoses.
4. Quality control: AI can help identify inconsistencies in staining techniques, equipment, and reporting protocols, enabling better standardization and quality assurance across laboratories. By monitoring and benchmarking these factors, AI can improve the overall quality of DP services.
5. Patient timeline and synoptic reporting: AI models can process and summarize the patient visits and interventions spread over multiple time points in the form of patient timeline and EMR summary of care. These models may also generate synoptic reports culled from the nonstructured data in the pathology reports.
6. Reporting and data integration: AI-driven language models can assist in the standardized extraction of information from pathology reports and facilitate the integration of this information with other clinical and research data. The AI model can also provide a degree of certainty to the diagnostic information extracted from pathology reports.[158] This can help improve the consistency, certainty, and comprehensiveness of data available for decision making and research purposes.
7. Education and training: AI can create standardized training materials and assessment tools for pathologists, ensuring that they are educated and evaluated based on best practices and the latest advancements in the field.
8. Interoperability and data sharing: AI can facilitate better communication and collaboration among laboratories and healthcare providers by providing a common platform for data analysis, visualization, and decision support. The AI language models can provide the translation of a pathology report between English and other languages for communication and collaboration among pathologists in different regions of the

world. This can contribute to standardizing workflows and practices across the DP ecosystem.

Trustworthiness and Acceptability of Artificial Intelligence by Pathologists

AI will not replace Pathologists but will help them in augmenting their tasks. Pathologists' trust in AI-based technologies is pivotal for successfully incorporating these tools into laboratory practice. The main ingredients for developing such trust are the validation of AI models before deployment in pathology laboratory workflow, performance monitoring after deployment, and continuous interaction between pathologists and the developers of the AI systems. The question of whether pathologists should trust AI is complex. We have demonstrated AI's potential to revolutionize pathology in the near future through innovative tools for helping in diagnosis, prognosis, and making accurate predictions swiftly. Nevertheless, concerns persist about the reliability and trustworthiness of AI, in part owing to the biases and missingness in training datasets and the inability of algorithms to tackle such limitations.

Additionally, AI models are vulnerable to adversarial examples that can deceive the AI model to drastically change its output with only a subtle change in the input.[79,159] Groups, such as Trustworthy Software Foundation[160] in the United Kingdom and Coalition for Health AI[161] in the United States are addressing trust issues by advocating for credible, fair, and transparent AI systems in healthcare. Trustworthy Software Foundation defined 5 facets of software trustworthiness: safety, reliability, availability, resilience, and security.[160] Coalition for Health AI has recently published a blueprint for trustworthy AI implementation in healthcare and defined 7 key elements of trustworthy AI in health care: useful, safe, accountable and transparent, explainable and interpretable, fair, secure and resilient, and privacy enhanced.[162] Furthermore, Dorr et al[163] proposed to create a "Code of Conduct for AI in Health Care" that aims to harmonize standards and ensure responsible AI usage. The potential benefits, limitations, and risks of generative AI systems and foundation models, such as GPT-4, have recently been highlighted, emphasizing their cautious use in clinical settings.[142] Rajpurkar and Lungren[164] presented the generalization checks for AI systems as transparency, clinician-AI collaboration, and post-deployment monitoring. Nakagawa et al[165] introduced various challenges in digitizing medical workflows, including data biases, privacy concerns, and algorithm fragility, while emphasizing the need to carefully consider AI's impact on pathologists. AI model cards are an important documentation framework for understanding, sharing, and improving ML models.[166,167] We summarize the strategies that may contribute to building the trust of pathologists in AI systems as follows:

1. Developing transparent and explainable AI models: users, such as practicing pathologists, should understand the basic functionality of AI models and their reasoning or logic for a certain decision, just like radiologists understand the core principles that are used by CT scanners, MRI machines, or X-rays to generate a certain type of images. Therefore, AI developers need to prioritize the use of transparent and explainable models that provide insights into their functionality and decision-making process.
2. Fostering a collaborative AI development approach: equally important is the engagement of pathologists in AI model development and validation processes, encouraging valuable feedback from both pathology and AI domain experts in AI development and deployment. Collaborative efforts and continuous feedback allow for performance improvement, ensuring the long-term utility of the models.
3. Education and training: there is value in educating and training medical professionals (practicing pathologists in this case) on the topics related to AI model development and the potential use of these technologies for solving various problems in health care. We have developed a hands-on ML course for medical professionals that aims to provide specialists such as pathologists with a practical acquaintance of AI concepts.[168]
4. Validating AI models in clinical settings: before deploying AI models in a clinical workflow, it is essential to perform rigorous evaluation and validation in the clinical settings where the model is intended to be used in a sandbox environment. This rigorous assessment will help identify and fix modeling issues, software bugs, and data and algorithm interoperability challenges. Such validation and evaluation processes will instill users' trust in AI models, assuring their suitability for real-world implementation and sustained use.
5. Incorporating ethical considerations: during AI model development, validation, and deployment, it is crucial to address ethical considerations to ensure fairness and reduce bias in the models' decisions. Validating and reporting the model's output for various groups based on age, race, and gender should be considered an essential part of model development and deployment.[166]

Foundation models and generative AI have the potential to transform DP, leading to faster and more accurate diagnoses, improved patient outcomes, and a better understanding of disease mechanisms, along with reducing workload for pathologists, helping standardize laboratory workflow, and contributing to the education and training. These powerful technologies will not replace pathologists but augment and streamline their skill sets. This review presented an overview of generative AI and foundation models and their potential role in DP. We demonstrated how AI as a field has grown from a narrow problem-solving technique to a comprehensive tool for language understanding, image analysis, data generation, question-answering, and conversation. Finally, we presented our perspective on the future role of generative AI and foundation models in DP and future use cases where generative and conversational AI and foundation models can have a transformative impact in DP. Adapting and integrating generative foundation models in traditional diagnostic methods can provide a more comprehensive and accurate assessment of pathology specimens while enabling the development of personalized treatments for patients. However, generative AI and foundation models have associated challenges and limitations. Further research and development efforts are needed to fully realize the current AI wave's potential to ensure their safe and effective implementation in clinical practice.

Author Contributions

G.R. and M.B. conceived the initial idea. A.W. and G.R. reviewed the literature and wrote the initial draft. All authors reviewed and contributed to the draft.

Data Availability

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Funding

This work was partly supported by the United States National Science Foundation awards ECCS-1903466, OAC-2008690, and OAC-2234836.

Declaration of Competing Interest

The authors declare no competing interests.

Ethics Approval and Consent to Participate

Not applicable.

Declaration of Generative AI and AI-assisted Technologies in the Writing Process

During the preparation of this work, the author(s) used ChatGPT[16] to generate the different pathology biopsy reports as depicted in Figure 5. The authors take full responsibility for the content of the publication.

References

1. Dick S. Artificial intelligence. *Harv Data Sci Rev*. 2019;1(1). https://doi.org/10.1162/99608f92.92fe150c
2. Swanson K, Wu E, Zhang A, Alizadeh AA, Zou J. From patterns to patients: advances in clinical machine learning for cancer diagnosis, prognosis, and treatment. *Cell.* 2023;186(8):1772–1791. https://doi.org/10.1016/j.cell.2023.01.035
3. Shanahan M. Talking about large language models. Preprint. Posted online February 25, 2019. arXiv:2212.03551. https://arxiv.org/abs/2212.03551
4. Tizhoosh HR, Pantanowitz L. Artificial intelligence and digital pathology: challenges and opportunities. *J Pathol Inform*. 2018;9(1):38. https://doi.org/10.4103/jpi.jpi_53_18
5. Falahkheirkhah K, Tiwari S, Yeh K, et al. Deepfake histologic images for enhancing digital pathology. *Lab Invest*. 2023;103(1):100006. https://doi.org/10.1016/j.labinv.2022.100006
6. Scopio Team. Scopio Labs; 2023. Accessed March 31, 2023. https://scopiolabs.com/
7. Paige Team. Pathology Artificial Intelligence Guidance Engine (PAIGE); 2023. Accessed March 31, 2023. https://paige.ai/
8. Aiforia Team. AI for Image Analysis (AIFORIA). 2023. Available from: Accessed March 31, 2023. https://www.aiforia.com/
9. Gupta R, Kurc T, Sharma A, Almeida JS, Saltz J. The emergence of pathomics. *Curr Pathobiol Rep*. 2019;7(3):73–84. https://doi.org/10.1007/s40139-019-00200-x
10. Drogt J, Milota M, Vos S, Bredenoord A, Jongsma K. Integrating artificial intelligence in pathology: a qualitative interview study of users' experiences and expectations. *Mod Pathol*. 2022;35(11):1540–1550. https://doi.org/10.1038/s41379-022-01123-6
11. Kim I, Kang K, Song Y, Kim TJ. Application of artificial intelligence in pathology: trends and challenges. *Diagnostics (Basel)*. 2022;12(11):2794. https://doi.org/10.3390/diagnostics12112794
12. Patel AU, Shaker N, Mohanty S, et al. Cultivating clinical clarity through computer vision: a current perspective on whole slide imaging and artificial intelligence. *Diagnostics (Basel)*. 2022;12(8):1778. https://doi.org/10.3390/diagnostics12081778
13. Bommasani R, Hudson DA, Adeli E, et al. On the opportunities and risks of foundation models. Preprint. Published online August 16, 2021. arXiv: 2108.07258. https://arxiv.org/abs/2108.07258
14. Moor M, Banerjee O, Abad ZSH, et al. Foundation models for generalist medical artificial intelligence. *Nature*. 2023;616(7956):259–265. https://doi.org/10.1038/s41586-023-05881-4
15. OpenAI. GPT-4; 2023. Accessed April 5, 2023. https://openai.com/researc h/gpt-4
16. OpenAI. Introducing ChatGPT; 2022. Accessed March 10, 2023. https://openai.com/blog/ chatgpt
17. Brown T, Mann B, Ryder N, et al. Language models are few-shot learners. In: Advances in Neural Information Processing Systems. *34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.* 2020;33:1877–1901.
18. Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I. Better language models and their implications; 2019. Accessed April 5, 2023. https://openai.com/research/better-language-models
19. Radford A, Narasimhan K, Salimans T, et al. Improving language understanding by generative pre-training; 2018. Accessed April 5, 2023. https://openai.com/research/language-unsupervised
20. Ouyang L, Wu J, Jiang X, et al. Training language models to follow instructions with human feedback. In: Oh AH, Agarwal A, Belgrave D, Cho K, eds. *Advances in Neural Information Processing Systems*. 2022.
21. Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A. Artificial intelligence in digital pathology—new tools for diagnosis and precision oncology. *Nat Rev Clin Oncol*. 2019;16(11):703–715. https://doi.org/10.1038/s41571-019-0252-y
22. Alsaafin A, Safarpoor A, Sikaroudi M, Hipp JD, Tizhoosh HR. Learning to predict RNA sequence expressions from whole slide images with applications for search and classification. *Commun Biol*. 2023;6(1):304. https://doi.org/10.1038/s42003-023-04583-x
23. Cifci D, Veldhuizen GP, Foersch S, Kather JN. AI in computational pathology of cancer: improving diagnostic workflows and clinical outcomes? *Annu Rev Cancer Biol*. 2023;7(1):57–71. https://doi.org/10.1146/annurev-cancerbio-061521-092038
24. Adam G, Rampášek L, Safikhani Z, Smirnov P, Haibe-Kains B, Goldenberg A. Machine learning approaches to drug response prediction: challenges and recent progress. *NPJ Precis Oncol*. 2020;4(1):19. https://doi.org/10.1038/s41698-020-0122-1
25. Demetriou D, Hull R, Kgoebane-Maseko M, Lockhat Z, Dlamini Z. AI-Enhanced Digital Pathology and Radiogenomics in Precision Oncology. In: *Artificial Intelligence and Precision Oncology: Bridging Cancer Research and Clinical Decision Support*. Springer; 2023: 93–113.
26. Pantanowitz L, Hartman D, Qi Y, et al. Accuracy and efficiency of an artificial intelligence tool when counting breast mitoses. *Diagn Pathol*. 2020;15(1):80. https://doi.org/10.1186/s13000-020-00995-z
27. Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol (Pozn)*. 2015;19(1A):A68–A77. https://doi.org/10.5114/wo.2014.47136
28. Ellis MJ, Gillette M, Carr SA, et al. Connecting genomic alterations to cancer biology with proteomics: the NCI Clinical Proteomic Tumor Analysis Consortium. *Cancer Discov*. 2013;3(10):1108–1112. https://doi.org/10.1158/2159-8290.CD-13-0219
29. Clark K, Vendt B, Smith K, et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging*. 2013;26(6):1045–1057. https://doi.org/10.1007/s10278-013-9622-7
30. Abels E, Pantanowitz L, Aeffner F, et al. Computational pathology definitions, best practices, and recommendations for regulatory guidance: a white paper from the Digital Pathology Association. *J Pathol*. 2019;249(3):286–294. https://doi.org/10.1002/path.5331
31. Kiehl TR. Digital and computational pathology: a specialty reimagined. In: *The Future Circle of Healthcare: AI, 3D Printing, Longevity, Ethics, and Uncertainty Mitigation*. Springer; 2022:227–250.
32. Bai B, Yang X, Li Y, Zhang Y, Pillar N, Ozcan A. Deep learning-enabled virtual histological staining of biological samples. *Light Sci Appl*. 2023;12(1):57. https://doi.org/10.1038/s41377-023-01104-7
33. Cifci D, Foersch S, Kather JN. Artificial intelligence to identify genetic alterations in conventional histopathology. *J Pathol*. 2022;257(4):430–444. https://doi.org/10.1002/path.5898
34. Echle A, Rindtorff NT, Brinker TJ, Luedde T, Pearson AT, Kather JN. Deep learning in cancer pathology: a new generation of clinical biomarkers. *Br J Cancer*. 2021;124(4):686–696. https://doi.org/10.1038/s41416-020-01122-x
35. Cui M, Zhang DY. Artificial intelligence and computational pathology. *Lab Invest*. 2021;101(4):412–422. https://doi.org/10.1038/s41374-020-00514-0
36. Gurcan MN, Boucheron LE, Can A, Madabhushi A, Rajpoot NM, Yener B. Histopathological image analysis: a review. *IEEE Rev Biomed Eng*. 2009;2: 147–171. https://doi.org/10.1109/RBME.2009.2034865
37. Irshad H, Veillard A, Roux L, Racoceanu D. Methods for nuclei detection, segmentation, and classification in digital histopathology: a review—current status and future potential. *IEEE Rev Biomed Eng*. 2014;7: 97–114. https://doi.org/10.1109/RBME.2013.2295804
38. Waqas A, Dera D, Rasool G, Bouaynaya NC, Fathallah-Shaykh HM. Brain Tumor Segmentation and Surveillance with Deep Artificial Neural Networks. In: Elloumi M, ed. *Deep Learning for Biomedical Data Analysis*. Cham: Springer; 2021. https://doi.org/10.1007/978-3-030-71676-9_13
39. Dera D, Bouaynaya NC, Rasool G, Shterenberg R, Fathallah-Shaykh HM. PremiUm-CNN: Propagating uncertainty towards robust convolutional neural networks. *IEEE Trans Signal Process*. 2021;69:4669–4684. https://doi.org/10.1109/TSP.2021.3096804
40. Waqas A, Farooq H, Bouaynaya NC, Rasool G. Exploring robust architectures for deep artificial neural networks. *Commun Eng*. 2022;1(1):46. https://doi.org/10.1038/s44172-022-00043-2
41. Ahmed S, Dera D, Hassan SU, Bouaynaya N, Rasool G. Failure detection in deep neural networks for medical imaging. *Front Med Technol*. 2022;4: 919046. https://doi.org/10.3389/fmedt.2022.919046

42. Albahra S, Gorbett T, Robertson S, et al. Artificial intelligence and machine learning overview in pathology & laboratory medicine: a general review of data preprocessing and basic supervised concepts. *Semin Diagn Pathol*. 2023;40(2):71–87. https://doi.org/10.1053/j.semdp.2023.02.002
43. Shen D, Wu G, Suk HI. Deep learning in medical image analysis. *Annu Rev Biomed Eng*. 2017;19:221–248. https://doi.org/10.1146/annurev-bioeng-071516-044442
44. Aubreville M, Stathonikos N, Bertram CA, et al. Mitosis domain generalization in histopathology images—The MIDOG challenge. *Med Image Anal*. 2023;84, 102699.
45. Bulten W, Kartasalo K, Chen PHC, et al. Artificial intelligence for diagnosis and Gleason grading of prostate cancer: the PANDA challenge. *Nat Med*. 2022;28(1):154–163.
46. Ma X, Jia F. Brain tumor classification with multimodal MR and pathology images. Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Springer; 2020:343–352, 5th International Workshop, BrainLes 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 17, 2019, Revised Selected Papers, Part II 5.
47. Veta M, Van Diest PJ, Willems SM, et al. Assessment of algorithms for mitosis detection in breast cancer histopathology images. *Med Image Anal*. 2015;20(1):237–248.
48. Guo R, Xie K, Pagnucco M, Song Y. SAC-Net: Learning with weak and noisy labels in histopathology image segmentation. *Med Image Anal*. 2023;86, 102790.
49. Mahmood F, Borders D, Chen RJ, et al. Deep adversarial training for multi-organ nuclei segmentation in histopathology images. IEEE Trans Med Imaging. 2019;39(11):3257–3267.
50. Gao Y, Phillips JM, Zheng Y, Min R, Fletcher PT, Gerig G. Fully convolutional structured LSTM networks for joint 4D medical image segmentation. 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI): IEEE; 2018:1104–1108.
51. Anand D, Gadiya S, Sethi A. Histographs: graphs in histopathology. SPIE; 2020:150–155. Medical Imaging 2020: Digital Pathology;11320.
52. Zhou Y, Graham S, Alemi Koohbanani N, Shaban M, Heng PA, Rajpoot N. CGC-Net: Cell Graph Convolutional Network for Grading of Colorectal Cancer Histology Images. Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops; 2019.
53. Studer L, Wallau J, Dawson H, Zlobec I, Fischer A. Classification of intestinal gland cell-graphs using graph neural networks. 2020 25th International conference on pattern recognition (ICPR). IEEE; 2021:3636–3643.
54. Sureka M, Patil A, Anand D, Sethi A. Visualization for histopathology images using graph convolutional neural networks. 2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE). IEEE; 2020: 331–335.
55. Bychkov D, Turkki R, Haglund C, Linder N, Lundin J. Deep learning for tissue microarray image-based outcome prediction in patients with colorectal cancer. SPIE; 2016:298–303. Medical Imaging 2016: Digital Pathology; 9791.
56. Edara DC, Vanukuri LP, Sistla V, Kolli VKK. Sentiment analysis and text categorization of cancer medical records with LSTM. *J Ambient Intell Humaniz Comput*. 2019;14:1–17.
57. Liu S, Wang X, Xiang Y, Xu H, Wang H, Tang B. Multi-channel fusion LSTM for medical event prediction using EHRs. *J Biomed Inform*. 2022;127, 104011.
58. Wang J, Chen RJ, Lu MY, Baras A, Mahmood F. Weakly supervised prostate TMA classification via graph convolutional networks. In: *2020 IEEE 17th ISBI*. IEEE, Iowa City, IA; 2020:239–243. https://doi.org/10.1109/ISBI45749.2020.9098534
59. 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(4):100980. https://doi.org/10.1016/j.xcrm.2023.100980
60. Braman N, Gordon JW, Goossens ET, Willis C, Stumpe MC, Venkataraman J. Deep orthogonal fusion: multimodal prognostic biomarker discovery integrating radiology, pathology, genomic, and clinical data. Medical Image Computing and Computer Assisted Intervention—MICCAI2021: 24th International Conference, Strasbourg, France, September 27—October 1, 2021, Proceedings, Part V24: Springer.; 2021:667–677.
61. Jardim-Perassi BV, Mu W, Huang S, et al. Deep-learning and MR images to target hypoxic habitats with evofosfamide in preclinical models of sarcoma. *Theranostics*. 2021;11(11):5313.
62. Rajeev R, Samath JA, Karthikeyan N. An intelligent recurrent neural network with long short-term memory (LSTM) BASED batch normalization for medical image denoising. *J Med Syst*. 2019;43:1–10.
63. Leevy JL, Khoshgoftaar TM. A short survey of LSTM models for deidentification of medical free text. 2020 IEEE 6th International Conference on Collaboration and Internet Computing (CIC). IEEE; 2020:117–124.
64. Vu QD, Rajpoot K, Raza SEA, Rajpoot N. Handcrafted Histological Transformer (H2T): Unsupervised representation of whole slide images. *Med Image Anal*. 2023;85, 102743.
65. Ibrahim A, Gamble P, Jaroensri R, et al. Artificial intelligence in digital breast pathology: techniques and applications. *Breast*. 2020;49:267–273.
66. Waqas A, Tripathi A, Ramachandran RP, Stewart P, Rasool G. Multimodal data integration for oncology in the era of deep neural networks: a review. Preprint. Posted online March 11, 2023. arXiv:2303.06471; 2023. https://arxiv.org/abs/2303.06471.
67. Ahmedt-Aristizabal D, Armin MA, Denman S, Fookes C, Petersson L. A survey on graph-based deep learning for computational histopathology. *Comput Med Imaging Graph*. 2022;95, 102027.
68. Huang SC, Pareek A, Jensen M, Lungren MP, Yeung S, Chaudhari AS. Self-supervised learning for medical image classification: a systematic review and implementation guidelines. *NPJ Digi Med*. 2023;6(1):74.
69. Azad R, Kazerouni A, Heidari M, et al. Advances in medical image analysis with vision transformers: A comprehensive review. *arXiv*. 2023. Preprint; arXiv:2301.03505.
70. Xia K, Wang J. Recent advances of transformers in medical image analysis: a comprehensive review. *MedComm–Future Medicine*. 2023; 2(1):e38.
71. Adnan M, Kalra S, Tizhoosh HR. Representation learning of histopathology images using graph neural networks. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). 2020: 4254–4261. https://doi.org/10.1109/CVPRW50498.2020.00502
72. Chen RJ, Lu MY, Wang J, et al. Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. *IEEE Trans Med Imaging*. 2022;41(4):757–770. https://doi.org/10.1109/TMI.2020.3021387
73. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. In: Guyon I, Luxburg UV, Bengio S, et al., eds. Advances in Neural Information Processing Systems. 2017.
74. Ahmed S, Nielsen IE, Tripathi A, Siddiqui S, Rasool G, Ramachandran RP. Transformers in time-series analysis: a tutorial. Preprint. Posted online. April 28, 2022. arXiv:2205.01138;2022. https://arxiv.org/abs/2205.01138
75. Computational Pathology & AI. FDA has cleared four pathology AI algorithms; 2023. Accessed July 20, 2023. https://www.pathologynews.com/computational-pathology-ai/fda-has-now-cleared-more-than-500-healthcare-ai-algorithms-four-of-which-are-for-pathology
76. FDA. FDA approved (AI/ML)-enabled medical devices; 2022. Accessed July 20, 2023. https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices
77. Dera D, Rasool G, Bouaynaya N. Extended variational inference for propagating uncertainty in convolutional neural networks. In: *2019 IEEE 29th International Workshop on MLSP*. IEEE; 2019:1–6.
78. Carannante G, Dera D, Bouaynaya NC, Fathallah-Shaykh HM, Rasool G. Trustworthy medical segmentation with uncertainty estimation. Preprint. Posted online November 10, 2021. arXiv:2111.05978. https://arxiv.org/abs/21 11.05978
79. Ghaffari Laleh N, Truhn D, Veldhuizen GP, et al. Adversarial attacks and adversarial robustness in computational pathology. *Nat Commun*. 2022;13(1):5711. https://doi.org/10.1038/s41467-022-33266-0
80. Alwosheel A, Cranenburgh S, Chorus CG. Is your dataset big enough? Sample size requirements when using artificial neural networks for discrete choice analysis. *J Choice Model*. 2018;28:167–182. https://doi.org/10.1016/j.jocm.2018.07.002
81. Khan H, Bouaynaya NC, Rasool G. Adversarially robust continual learning. In: *2022 IJCNN:1-8IEEE*. 2022. https://doi.org/10.1109/IJCNN55064.2022.9892970
82. Ahn E, Kumar A, Feng D, Fulham M, Kim J. Unsupervised deep transfer feature learning for medical image classification. In: *2019 IEEE 16th ISBI: IEEE*. 2019:1915–1918.
83. Boehm KM, Khosravi P, Vanguri R, Gao J, Shah SP. Harnessing multimodal data integration to advance precision oncology. *Nat Rev Cancer*. 2022;22(2): 114–126. https://doi.org/10.1038/s41568-021-00408-3
84. Vanguri RS, Luo J, Aukerman AT, et al. Multimodal integration of radiology, pathology and genomics for prediction of response to PD-(L)1 blockade in patients with non-small cell lung cancer. *Nat Cancer*. 2022;3(10): 1151–1164. https://doi.org/10.1038/s43018-022-00416-8
85. McKinney SM, Sieniek M, Godbole V, et al. International evaluation of an AI system for breast cancer screening. *Nature*. 2020;577(7788):89–94. https://doi.org/10.1038/s41586-019-1799-6
86. Haibe-Kains B, Adam GA, Hosny A, et al. Transparency and reproducibility in artificial intelligence. *Nature*. 2020;586(7829):E14–E16. https://doi.org/10.1038/s41586-020-2766-y
87. McKinney SM, Karthikesalingam A, Tse D, et al. Reply to: transparency and reproducibility in artificial intelligence. *Nature*. 2020;586(7829):E17–E18. https://doi.org/10.1038/s41586-020-2767-x
88. Nielsen IE, Dera D, Rasool G, Ramachandran RP, Bouaynaya NC. Robust explainability: a tutorial on gradient-based attribution methods for deep neural networks. *IEEE Signal Process Mag*. 2022;39(4):73–84. https://doi.org/10.1109/MSP.2022.3142719
89. Nielsen IE, Ramachandran RP, Bouaynaya N, Fathallah-Shaykh HM, Rasool G. EvalAttAI: a holistic approach to evaluating attribution maps in robust and non-robust models. *IEEE Access*. 2023;11:82556–82569. https://doi.org/10.1109/ACCESS.2023.3300242
90. Raffel C, Shazeer N, Roberts A, et al. Exploring the limits of transfer learning with a unified text-to-text transformer. *JMLR*. 2020;21(1):5485–5551.

91. Scao TL, Fan A, Akiki C, et al. BLOOM: a 176b-param*eter o*pen-access multilingual language model. Preprint. Posted online November 9, 2022. arXiv:2211.05100.
92. Gan Z, Li L, Li C, Wang L, Liu Z, Gao J. Vision-language pre-training: basics, recent advances, and future trends. *Found Trends Comput Graph Vis*. 2022;14(3-4):163−352. https://doi.org/10.1561/0600000105
93. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pretraining of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the ACL: Human Language Technologies, Volume 1 *(Long and Short Papers)*. ACL. 2019:4171- 4186.
94. Liu Y, Ott M, Goyal N, et al. RoBERTa: a robustly optimized BERT pretraining approach. Preprint. Posted online July 26, 2019. https://arxiv.org/abs/1907.11692
95. Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16×16 words: transformers for image recognition at scale. Preprint. Posted online October 22, 2020. arXiv preprint arXiv:2010.11929.
96. Cheng B, Misra I, Schwing AG, Kirillov A, Girdhar R. Masked-attention mask transformer for universal image segmentation. In: *Proceedings of the IEEE/CVF Conference on CVPR*. 2022:1290−1299.
97. Jain J, Li J, Chiu MT, Hassani A, Orlov N, Shi H. Oneformer: one transformer to rule universal image segmentation. In: *Proceedings of the IEEE/CVF Conference on CVPR*. 2023:2989−2998.
98. Lüddecke T, Ecker A. Image segmentation using text and image prompts. In: *2022 IEEE/CVF Conference on CVPR*. IEEE; 2022:7076−7708.
99. Jaegle A, Borgeaud S, Alayrac JB, et al. Perceiver IO: a general architecture for structured inputs & outputs. Preprint. Posted online July 30, 2021. arXiv: 2107.14795.
100. Surís D, Menon S, Vondrick C. Vipergpt: visual inference via python execution for reasoning. Preprint. Posted online March 14, 2023. arXiv: 2303.08128. https://arxiv.org/abs/2303.08128.
101. Liu H, Li C, Wu Q, Lee YJ. Visual instruction tuning. Preprint. Posted online April 17, 2023.arXiv:2304.08485; 2023. https://arxiv.org/abs/2304.08485
102. Li J, Li D, Savarese S, Hoi S. BLIP-2: bootstrapping language-image pretraining with frozen image encoders and large language models. Preprint. Posted online January 30, 2023 arXiv:2301.12597. https://arxiv.org/abs/2301.12597
103. Hudson DA, Zitnick L. Generative adversarial transformers. *ICML:PMLR*. 2021;139:4487−4499. https://proceedings.mlr.press/v139/hudson21a.html
104. Radford A, Kim JW, Hallacy C, et al. Learning transferable visual models from natural language supervision. In: Meila M, Zhang T, eds. *Proceedings of the 38th ICML*. 139. PMLR; 2021:8748−8763.
105. Jia C, Yang Y, Xia Y, et al. Scaling up visual and vision-language representation learning with noisy text supervision. In: Meila M, Zhang T, eds. *Proceedings of the 38th ICML 139 of PMLR*. 2021:4904−4916.
106. Wang Z, Yu J, Yu AW, Dai Z, Tsvetkov Y, Cao Y. SimVLM: simple visual language model pretraining with weak supervision. In: *ICLR*. 2022.
107. Alayrac JB, Donahue J, Luc P, et al. Flamingo: A visual language model for few-shot learning. In: Oh AH, Agarwal A, Belgrave D, Cho K, eds. *Advances in Neural Information Processing Systems*. 2022.
108. Yu J, Wang Z, Vasudevan V, Yeung L, Seyedhosseini M, Wu Y. *CoCa: contrastive captioners are image-text foundation models*. TMLR; 2022.
109. Lu MY, Chen B, Williamson DF, et al. Towards a visual-language foundation model for computational pathology. Preprint. Posted online July 24, 2023. arXiv:2307.12914. https://arxiv.org/abs/2307.12914
110. Lester B, Al-Rfou R, Constant N. The power of scale for parameter-efficient prompt tuning. In: *Proceedings of the 2021 EMNLP*. Association for Computational Linguistics; 2021:3045−3059.
111. Wei J, Tay Y, Bommasani R, et al. Emergent abilities of large language models. [Survey Certification]. *TMLR* 2022. Preprint. Posted online October 26, 2022. arXiv:2206.07682v2**.** https://doi.org/10.48550/arXiv.2206.07682
112. Zhou C, Li Q, Li C, et al. A comprehensive survey on pretrained foundation models: a history from BERT to ChatGPT; Preprint. Posted online February 18, 2023. https: //arxiv.org/abs/2302.09419
113. Willemink MJ, Roth HR, Sandfort V. Toward foundational deep learning models for medical imaging in the new era of transformer networks. *Radiol: Artif Intell*. 2022;4(6), e210284. https://doi.org/10.1148/ryai.210284
114. Qiu Y, Lin F, Chen W, Xu M. Pre-training in medical data: a survey. *Mach Intell Res*. 2023;20(2):147−179. https://doi.org/10.1007/s11633-022-1382-8
115. Liu P, Yuan W, Fu J, Jiang Z, Hayashi H, Neubig G. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Comput Surv*. 2023;55(9):1−35. https://doi.org/10.1145/3560815
116. Rombach R, Blattmann A, Lorenz D, Esser P, Ommer B. High-resolution image synthesis with latent diffusion models. In: *Proceedings of the IEEE/CVF Conference on CVPR*. 2022:10684−10695.
117. Ramesh A, Dhariwal P, Nichol A, Chu C, Chen M. Hierarchical text-conditional image generation with CLIP latents. Preprint. Posted online April 13, 2022. https://arxiv.org/abs/2204.06125
118. Turian J, Ratinov L, Bengio Y. Word representations: a simple and general method for semi-supervised learning. In: *Proceedings of the 48th Annual Meeting of the ACL*. 2010:384−394.
119. Peters ME, Neumann M, Iyyer M, et al. Deep contextualized word representations. In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers);* 2018. https://doi.org/10.18653/v1/N18-1202
120. Howard J, Ruder S. Universal language model fine-tuning for text classification. In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers);* 2018. https://doi.org/10.18653/v1/P18-1031
121. Touvron H, Lavril T, Izacard G, et al. LLaMA: open and efficient foundation language models. Preprint. Posted online February 27, 2023. https://arxiv.org/abs/2302.13971.
122. Lewis M, Liu Y, Goyal N, et al. BART: Denoising sequence-to-sequence pretraining for natural language generation, translation, and comprehension. In: Proceedings of the 58th Annual Meeting of the ACL; ACL. 20:7871-7880. https://aclanthology.org/2020.acl-main.703
123. Lu H, Zhou Q, Fei N, et al. Multimodal foundation models are better simulators of the human brain. Preprint. Posted online August 17, 2022. arXiv: 2208.08263;. https://arxiv.org/abs/2208.08263.
124. Dirik A, Paul S. A dive into vision-language models. Published online February 3, 2023 https://huggingface.co/blog/vision_language_pretraining
125. Hudson DA, Manning CD. GQA: a new dataset for real-world visual reasoning and compositional question answering. In: *Proceedings of the IEEE/CVF Conference on CVPR*. 2019:6693−6702. https://doi.org/10.1109/CVPR.2019.00686
126. Yan B, Pei M. Clinical-BERT: Vision-language pre-training for radiograph diagnosis and reports generation. In: Proceedings of the AAAI Conference; 2022;36(3):2982-2990.
127. Yang J, Li C, Zhang P, et al. Unified contrastive learning in image-text-label space. In: *Proceedings of the IEEE/CVF Conference on CVPR*. 2022: 19163−19173.
128. Gu X, Lin TY, Kuo W, Cui Y. Open-vocabulary object detection via vision and language knowledge distillation. ICLR; 2022.
129. Zhong Y, Yang J, Zhang P, et al. RegionCLIP: region-based language-image pretraining. In: *2022 IEEE/CVF Conference on CVPR; Los Alamitos, CA*. IEEE-CS; 2022:16772−16782.
130. Li L, Zhang P, Zhang H, et al. Grounded Language-Image Pretraining. In: *2022 IEEE/CVF Conference on CVPR; Los Alamitos, CA*. IEEE-CS; 2022: 10955−10965.
131. Zhou X, Girdhar R, Joulin A, Krähenbühl P, Misra I. Detecting twenty-thousand classes using image-level supervision. In: *Computer Vision - ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23-27, 2022, Proceedings, Part IX*. Berlin, Heidelberg: Springer-Verlag; 2022: 350−368.
132. Minderer M, Gritsenko A, Stone A, et al. Simple open-vocabulary object detection with vision transformers. In: *Computer Vision - ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23-27, 2022, Proceedings, Part X*. Berlin, Heidelberg: Springer-Verlag; 2022:728−755.
133. Zang Y, Li W, Zhou K, Huang C, Loy CC. Open-vocabulary DETR with conditional matching. In: *Computer Vision - ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23-27, 2022, Proceedings, Part IX*. Berlin, Heidelberg: Springer-Verlag; 2022:106−122.
134. Cai Z, Kwon G, Ravichandran A, et al. X-DETR: a versatile architecture for instance-wise vision-language tasks. In: *Computer Vision - ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23-27, 2022, Proceedings, Part XXXVI*. Berlin, Heidelberg: Springer-Verlag; 2022:290−308.
135. Li B, Weinberger KQ, Belongie S, Koltun V, Ranftl R. Language-driven Semantic Segmentation. *ICLR*; 2022.
136. Ghiasi G, Gu X, Cui Y, Lin TY. Scaling Open-Vocabulary Image Segmentation With Image-Level Labels. In: *Computer Vision − ECCV 2022: 17th European Conference, October 23-27, 2022, Proceedings, Part XXXVI*. Berlin, Heidelberg: Springer-Verlag; 2022:540−557.
137. Zhou C, Loy CC, Dai B. Extract free dense labels from CLIP. In: *Computer Vision − ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23−27, 2022, Proceedings, Part XXVIII*. Berlin, Heidelberg: Springer-Verlag; 2022:696−712.
138. Rao Y, Zhao W, Chen G, et al. DenseCLIP: language-guided dense prediction with context-aware prompting. In: *2022 IEEE/CVF Conference on CVPR. Los Alamitos, CA*. IEEE-CS; 2022:18061−18070.
139. Xu J, De Mello S, Liu S, et al. GroupViT: Semantic segmentation emerges from text supervision. In: *2022 IEEE/CVF Conference on CVPR*. 2022:18113−18123.
140. Ramesh A, Pavlov M, Goh G, et al. Zero-shot text-to-image generation. In: *ICML: PMLR*. 2021:8821−8831.
141. Smith S, Patwary M, Norick B, et al. Using deepSpeed and megatron to train megatron-turing NLG 530B, a large-scale generative language model. Preprint. Posted online January 28, 2022. arXiv:2201.11990. https://arxiv.org/abs/2201.11990
142. Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI Chatbot for medicine. *NEJM*. 2023;388(13):1233−1239. https://doi.org/10.1056/NEJMsr2214184
143. Alkaissi H, McFarlane SI. Artificial hallucinations in ChatGPT: implications in scientific writing. *Cureus*. 2023;15(2):e35179. https://doi.org/10.7759/cureus.35179

144. Jin W, Li X, Fatehi M, Hamarneh G. Guidelines and evaluation of clinical explainable AI in medical image analysis. *Med Image Anal*. 2023;84, 102684. https://doi.org/10.1016/j.media.2022.102684
145. Carannante G, Dera D, Rasool G, Bouaynaya NC. Self-compression in Bayesian neural networks. In: *2020 IEEE 30th International Workshop on MLSP*. IEEE; 2020:1−6.
146. Chen RJ, Chen C, Li Y, et al. Scaling vision transformers to gigapixel images via hierarchical self-supervised learning. In: *Proceedings of the IEEE/CVF Conference on CVPR*. 2022:16123−16134.
147. Cui Y, Liu Z, Liu X, et al. Bayes-MIL: a new probabilistic perspective on attention-based multiple instance learning for whole slide images. In: *The Eleventh ICLR*. 2023.
148. Shmatko A, Ghaffari Laleh N, Gerstung M, Kather JN. Artificial intelligence in histopathology: enhancing cancer research and clinical oncology. *Nat Cancer*. 2022;3(9):1026−1038. https://doi.org/10.1038/s43018-022-00436-4
149. Wang X, Xiang J, Zhang J, et al. SCL-WC: cross-slide contrastive learning for weakly-supervised whole-slide image classification. *Adv Neural Inf Process Syst*. 2022;35:18009−18021.
150. Chen RJ, Lu MY, Weng WH, et al. Multimodal co-attention transformer for survival prediction in gigapixel whole slide images. In: *Proceedings of the IEEE/CVF ICCV*. 2021:4015−4025.
151. Lipkova J, Chen RJ, Chen B, et al. Artificial intelligence for multimodal data integration in oncology. *Cancer Cell*. 2022;40(10):1095−1110. https://doi.org/10.1016/j.ccell.2022.09.012
152. Fu Y, Jung AW, Torne RV, et al. Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. *Nat Cancer*. 2020;1(8):800−810. https://doi.org/10.1038/s43018-020-0085-8
153. Jiang H, Zhou Y, Lin Y, Chan RCK, Liu J, Chen H. Deep learning for computational cytology: a survey. *Med. Image Anal*. 2023;84:102691. https://doi.org/10.1016/j.media.2022.102691
154. Ma R, Chen PHC, Li G, et al. Human-centric metric for accelerating pathology reports annotation. Preprint. Posted online October 31, 2019. arXiv preprint arXiv:1911.01226; 2019. https://arxiv.org/abs/1911.01226.
155. Sinha RK, Roy AD, Kumar N, Mondal H. Applicability of ChatGPT in assisting to solve higher order problems in pathology. *Cureus*. 2023;15(2):e35237. https://doi.org/10.7759/cureus.35237
156. Kather JN, Ghaffari Laleh N, Foersch S, Truhn D. Medical domain knowledge in domain-agnostic generative AI. *NPJ Digit Med*. 2022;5(1):90. https://doi.org/10.1038/s41746-022-00634-5
157. Hassell LA, Absar SF, Chauhan C, et al. Pathology education powered by virtual and digital transformation: now and the future. *Arch Pathol Lab Med*. 2023;147(4):474−491. https://doi.org/10.5858/arpa.2021-0473-RA
158. Gibson BA, McKinnon E, Bentley RC, et al. Communicating certainty in pathology reports. *Arch Pathol Lab Med*. 2022;146(7):886−893. https://doi.org/10.5858/arpa.2020-0761-OA
159. Finlayson SG, Bowers JD, Ito J, Zittrain JL, Beam AL, Kohane IS. Adversarial attacks on medical machine learning. *Science*. 2019;363(6433):1287−1289. https://doi.org/10.1126/science.aaw4399
160. *Institution of Analysts & Programmers*. Trustworthy Software Foundation; 2023. Accessed July 31, 2023. https://www.tsfdn.org/
161. Berkeley and Change Healthcare and Duke Health and Google and JHU and Mayo Clinic and Microsoft and MITRE and SAS and Stanford Medicine and UCSF and Vanderbilt University Medical Center. Coalition for Health AI; 2023. Accessed July 31, 2023. https://coalitionforhealthai.org/
162. Coalition for Health AI. *Blueprint for Trustworthy AI Implementation Guidance and Assurance for Healthcare*. 2023. Accessed July 31, 2023. https://coalitionforhealthai.org/papers/blueprint-for-trustworthy-ai_V1.0.pdf
163. Dorr DA, Adams L, Embí P. Harnessing the promise of artificial intelligence responsibly. *JAMA*. 2023;329(16):1347−1348. https://doi.org/10.1001/jama.2023.2771
164. Rajpurkar P, Lungren MP. The current and future state of AI interpretation of medical images. *N Engl J Med*. 2023;388(21):1981−1990. https://doi.org/10.1056/NEJMra2301725
165. Nakagawa K, Moukheiber L, Celi LA, et al. AI in pathology: what could possibly go wrong? *Semin Diagn Pathol*. 2023;40(2):100−108. https://doi.org/10.1053/j.semdp.2023.02.006
166. Mitchell M, Wu S, Zaldivar A, et al. *Model cards for model reporting. ACM FAccT*; 2019:220−229.
167. Ozoani E, Gerchick M, Mitchell M. *Model Cards*. 2022. https://huggingface.co/blog/model-cards
168. Rasool G. *HOME: hands-on machine learning course for medical professionals*. 2023. https://lab.moffitt.org/rasool/home/