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

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

[a]*Department of Machine Learning, H. Lee Moffitt Cancer Center & Research Institute, Tampa, FL, USA*
[b]*Department of Pathology, H. Lee Moffitt Cancer Center & Research Institute, FL, USA*
[c]*Affiliated Pathologists Medical Group, Inc, Rancho Dominguez, CA, USA*
[d]*Center of Excellence for Digital and AI-Empowered Pathology, Quest Diagnostics, Tampa, FL, USA*
[e]*University of South Florida, Morsani College of Medicine, Tampa, FL, USA*
[f]*National Artificial Intelligence Institute, US Department of Veteran Affairs, USA*
[g]*James A. Haley Veterans' Hospital, Tampa, FL, USA*
[h]*Department of Neuro-Oncology, H. Lee Moffitt Cancer Center & Research Institute, Tampa, FL, USA*

Abstract

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, which also enables computational pathology. By leveraging artificial intelligence (AI) and machine learning, computational pathology has emerged as a promising field in recent years. Recently, task-specific AI (e.g., 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 models depends on the availability of many annotated training datasets, which presents as a rate-limiting factor for AI development in pathology. Tasks-specific AI models cannot benefit from multimodal data and lack generalization, e.g., the AI models often struggle to generalize to new datasets or unseen variations in image quality, staining techniques, or tissue types. The 2020s are witnessing the rise of foundation models and generative AI. A foundation model is a large

*Corresponding author: Asim.Waqas@moffitt.org

AI model trained using sizeable data, which is later adapted (or finetuned) to perform different tasks using a modest amount of task-specific annotated data. These AI models provide in-context learning, can self-correct their 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.

Keywords: artificial intelligence, machine learning, digital pathology, histopathology, computational pathology, whole slide imaging, large language models, vision-language models, multimodal data, foundation models

---

1. 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 has emerged as a new approach for conducting precise quantitative assessments. Digital pathology 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 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: (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 morphological 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]. Machine learning 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 sub-field of AI that deals with image processing) and natural language processing (a sub-field of AI that deals with text and speech) [3]. Although the initial adoption of these technologies in medicine and healthcare 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, 5, 6, 7, 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, e.g., 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 machine learning 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, e.g., 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 require 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 (e.g., 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, e.g., patient demographics, laboratory data, or their prior disease history cannot help the model improve its prediction accuracy [5, 4]. 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 as these models use self-, semi- or unsupervised learning techniques. Foundation models can consume data from various modalities, including images (e.g., WSIs), text (e.g., pathology reports), and tabular data (e.g., medical records). The well-known generative AI model, ChatGPT, is based on a foundation model called Generative Pre-training Transformer (GPT) [15, 16, 17, 18, 19, 20]. Foundation models hold much promise for quantitative image analysis, diagnosis and prognosis, pathology report generation, and question/answering with conversational use in pathology lab workflow[13, 14].

Section 2 provides a brief overview of AI and machine learning models and advances enabled by these task-specific AI models in computational pathology. We introduce various foundation models, their structure, characteristics, and limitation in Section 3. Section 4 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. We also outline the potential role that foundation models and generative AI can play in pathology education and training before concluding the review in Section 5.

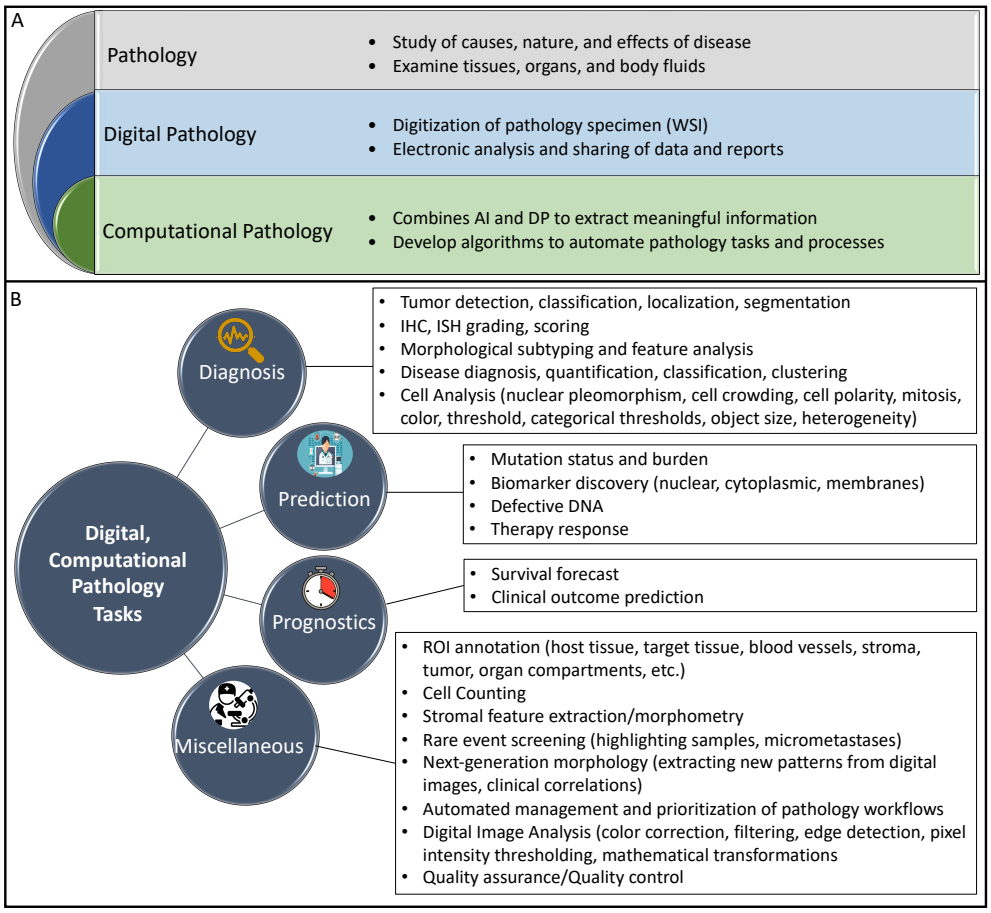

Figure 1: Pathology, digital pathology, and computational pathology - definitions and tasks are presented.

Box 1 — Definitions of key terminologies

Digital pathology	A comprehensive term that includes various tools and systems to digitize pathology slides and associated meta-data, as well as their storage, review, analysis, and supporting infrastructure.
Computational pathology	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	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	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, i.e., 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 to unlabeled samples, it is weakly-supervised learning.
Self-supervised learning	This technique of training AI does not require explicit data annotations. AI learns to solve a pre-designed auxiliary task or objective using the data's inherent structure, allowing the model to learn meaningful representations through self-created supervision signals.
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	A field of AI that enables computers and systems to derive meaningful information from digital images, videos, and other visual data.
Natural language processing	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 scalded dot-product) to capture relationships between input elements, especially in long sequences. They can process and learn from all data types, including images, text, speech, etc.
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, GANs, and foundation models such as ChatGPT, GPT-4, Stable Diffusion, and Dall-E 2.

2. AI 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, 4, 22, 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].

2.1. Digital Pathology

Digital pathology involves digitizing tissue specimens, allowing them to be analyzed and shared electronically. Digital pathology uses complex imaging systems to capture high-resolution images of tissue specimens, which can then be viewed and analyzed on a computer screen [25]. Digital pathology 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 digital pathology, 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 (TCGA) [27], Clinical Proteomic Tumor Analysis Consortium (CPTAC) [28], and The Cancer Imaging Archive (TCIA) [29], host pathology image data for multiple cancer sites. This is possible only because of digital pathology and other advancements.

Whole Slide Imaging (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].

2.2. Computational Pathology

Computational pathology combines digital pathology with AI, machine learning, and other computational techniques to extract meaningful information [30]. Often interchangeably called “histomics,” “pathomics,” or “tissue

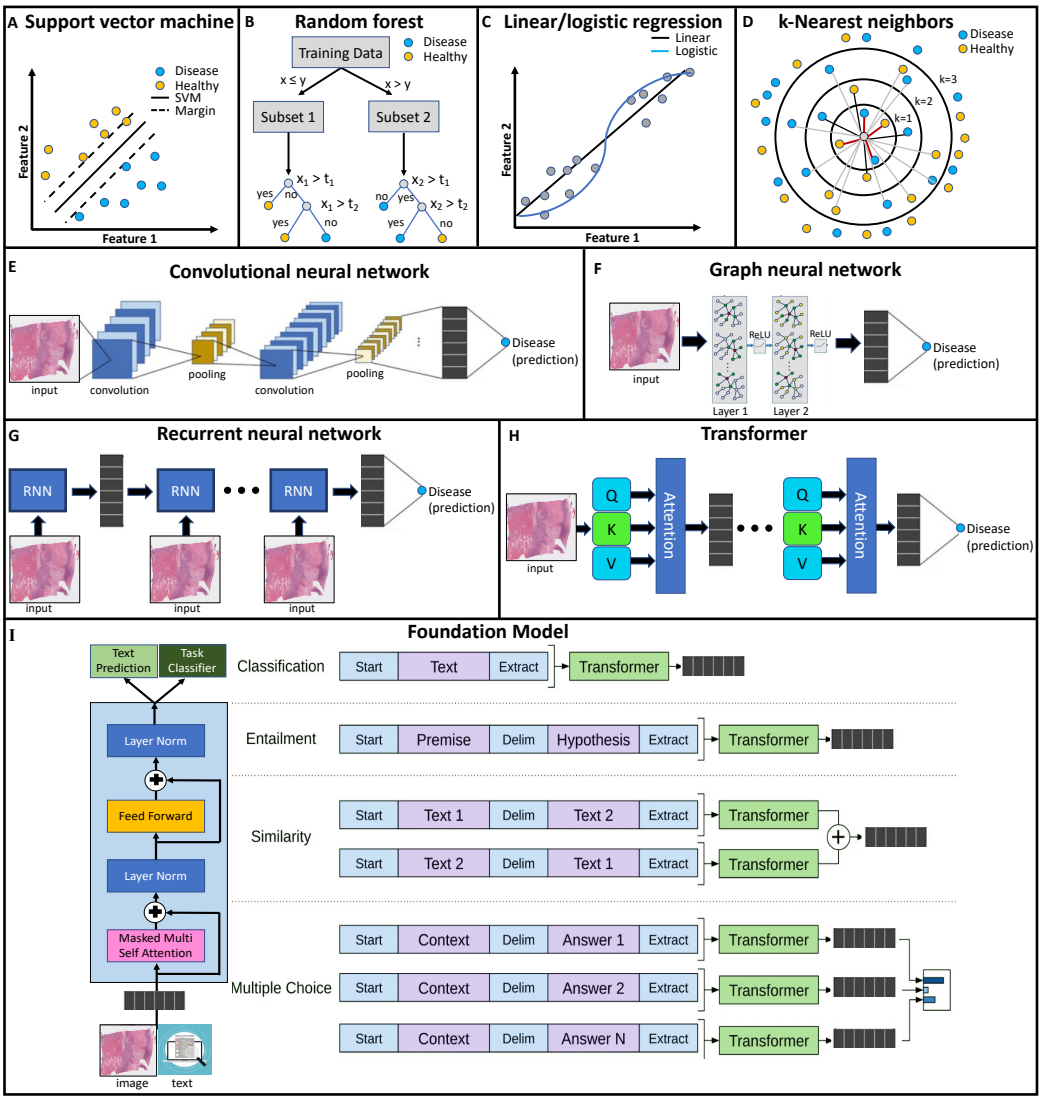

Figure 2: A schematic layout of various machine learning algorithms and AI models used in digital pathology. The top row (sub-figures A to D) highlights classical machine learning algorithms. Rows 2 and 3 (sub-figures E to H) present task-specific AI models). The last row (sub-figure I) refers to foundation models, the brain behind generative AI, such as ChatGPT. We argue that the role of classical machine learning 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 Section 4.

phenomics," computational pathology aims to develop 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 digital pathology [32], computational pathology methods can perform well on tasks such as classification, segmentation, and analysis of digital pathology images, at times surpassing human-level performance [33, 34]. The definitions of pathology, digital pathology, and computational pathology are illustrated in Fig. 1A, and their key tasks are illustrated in Fig. 1B.

2.3. Classical Machine Learning in Digital Pathology

Classical machine learning 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 machine learning has been extensively used in digital pathology for image segmentation and classification [2] using Support Vector Machines (SVMs), Random Forests, k-Nearest Neighbor (k-NN), Decision Trees, and others [2, 35]. A detailed review of the classical machine learning techniques in digital pathology is presented in [36] and [37]. Owing to the manual selection of usable and informative features, the applicability of classical machine learning methods is limited [36, 37].

2.4. Task-specific AI in Digital Pathology

More recently, task-specific AI models based on artificial neural networks have been gaining popularity [38, 39, 40, 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 Convolutional Neural Networks (CNNs), variants of Recurrent Neural Networks (RNNs), Graph Neural Networks (GNNs), and Transformers, as illustrated in Fig. 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 machine learning workloads), artificial neural networks-based task-specific AI models have found a strong foothold in digital pathology [4, 34, 43]. In the following discussion, we briefly introduce task-specific AI models and their essential components. In Table 1, we present a non-exhaustive list of various categories of task-specific AI models used in digital pathology. Interested readers are encouraged to explore the relevant works of interest.

2.4.1. Convolutional Neural Networks (CNNs)

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 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.

2.4.2. Recurrent Neural Networks (RNNs)

RNNs process sequential data like 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 (LSTMs) and Gated Recurrent Units (GRUs) are RNNs that help the model better capture long-term dependencies and avoid the vanishing gradient problem of RNNs using sequential processing of data.

2.4.3. Graph Neural Networks (GNNs)

GNNs process graph-structured data, such as social networks, molecular data, and knowledge graphs [44]. 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 [45, 46, 47].

2.4.4. Transformers

Transformers were initially introduced for language translation [48]. However, they have shown remarkable performance in various AI tasks, including computer vision and time series analysis [13, 49]. 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 [48]. Transformers have achieved state-of-the-art results in various natural language computer vision and graph processing tasks [48, 49].

2.5. AI-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, four of which are for pathology [75]. Among them, two were introduced earlier, and the other two more recently. "PAPNET Testing System" was approved in 1995, and it was designed for re-screening 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 fifteen 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 [76, 7]. The PaigeAI prostate algorithm and the Pathwork digital and AI platform are the pio-

Table 1: Summary of AI models used in digital pathology.

AI Model	**Task**	**Ref**
CNN	MIDOG: Mitosis domain generalization challenge.	[50]
CNN	Gleason grading and diagnosis of prostate cancer.	[51]
CNN	Feature extraction to classify brain tumor grade.	[52]
CNN	Disease outcome prediction in colorectal cancer.	[53]
CNN	Prediction of OS using Glioma multimodal data.	[54]
CNN	Mitosis detection in breast cancer.	[55]
CNN	Correlations between true hypoxia fraction in histological and the approximated fraction in MRI scans.	[56]
CNN+GAN	Similarity between virtually stained images (generated by AI model) & histochemically stained images.	[56]
GAN	Nuclei segmentation on histopathology images.	[57]
CNN	Segment nuclei in histology images using weakly-supervised training	[58]
CNN-Review	Deep learning in digital pathology for breast cancer.	[59]
LSTM	Predicting sentiment, text categorization in records.	[60]
LSTM	Medical image denoising.	[61]
LSTM	Medical event prediction using a multi-channel fusion of EHR data.	[62]
LSTM	De-identification of medical text.	[63]
LSTM	4D medical image segmentation.	[64]
GNN	Learn micro- and macro-structural features in H&E slides of breast cancer.	[65]
GNN	Grading colorectal cancer in histology images.	[66]
GNN	Classify healthy tissue from dysplastic gland areas in the colorectal cancer histology slides.	[67]
GNN	Classify infiltrating ductal carcinoma (IDC) and ductal carcinoma in situ (DCIS) breast cancer and grade Gleason 3 and 4 prostate cancer.	[68]
GNN	Stratify prostate cancer using tissue microarrays.	[46]
GNN-Review	GNN-based methods in cancer pathology.	[44, 69]
Transformers	Predicting RNAseq expressions from kidney WSIs using multiple instance learning.	[22]
Transformers	Predicting biomarkers from histopathology slides in colorectal cancer.	[70]
Transformers	WSI representations using unsupervised learning.	[71]
Transformers	Reviews on use of Transformers in Medical field.	[44, 72, 73, 74]

neering algorithms that have significantly impacted pathology practices by aiding in diagnosing and characterizing various diseases.

Pathology has a branch of anatomic pathology (AP) and clinical pathol-

ogy (CP). The four 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 have the capability to create new data or images, potentially aiding in generating synthetic samples for training and research purposes. While such algorithms have seen success in other domains, their application in pathology, encompassing both anatomic 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.

2.6. Limitations of Task-Specific AI

Task-specific AI models have many limitations restricting their widespread use in digital pathology.

1. Task-specificity: Task specificity refers to the fact that the trained AI performs well on a single task only, e.g., grading cancer sub-types 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: These AI models require the input data to have similar characteristics and follow the same probability distribution function of the input data (the mean and 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 [40, 39, 79]. These AI models cannot generalize to changes in data resulting from various common reasons, e.g., hardware, software, firmware upgrades in scanners, changes in the staining quality or the protocol, shifts in population demographics (e.g., a different geographical region), and changes in data

patterns due to new diseases such as COVID [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 tasks-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 (i.i.d) 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 digital pathology. These AI models cannot directly benefit from large amounts of unannotated datasets, e.g., WSIs, pathology reports, clinical notes, etc., and may require techniques such as weakly supervised learning, unsupervised 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, e.g., the patient's medical data from medical records, omics data, or radiographs, into the AI decision-making is generally not straightforward [44]. Recently, some research efforts have focused on creating AI models that can process multimodal data to improve their predictive accuracy with moderate success [47, 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 570GB 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 digital pathology, medical imaging, or any area of medical data processing. Task-specific, narrow versions of AI models are built by individual academic labs or industries that do not contribute to reusable knowledge accumulation [44].

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]. While 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].

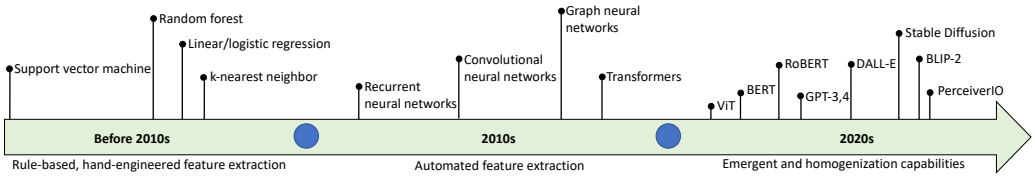

Figure 3: The evolution of machine learning models used in digital pathology. The 2020s are witnessing the rise of generative AI based on foundation models. In the near future, foundation models and generative AI will become the preferred computational approaches in digital pathology.

3. Foundation Models and Generative AI

The 2020s are witnessing the rise of foundation models - large AI models pre-trained using unannotated multimodal datasets (please refer to 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 tasks-specific AI. In the following, we present our perspective

on how foundation models and generative AI that use these models can transform the digital pathology 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.

3.1. Foundation Models

The term "foundation models" was initially coined by Bommasani *et al.* to describe recently proposed models that have led to a paradigm shift in AI model design, development, and deployment processes [13]. Foundation models are huge models trained at scale using comprehensive unannotated data (possibly multimodal). Foundation models generally have billions or trillions of learnable parameters and thousands of petaflops (floating point operations) [16, 17, 90, 91, 92]. 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 pre-training, transfer learning, and unsupervised and self-supervised learning. However, their essence lies in *scaling* because of the following three factors: (i) the introduction of Transformer architecture [48] that supports training models with the number of learnable parameters in billions or trillions, (ii) the availability of thousands of GPUs, and (iii) 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, 48].

Recently, a host of foundation models have been trained for language, vision, and joint language-vision (multimodal) tasks and shared via GitHub[1] and Hugging Face[2]. 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, 97, 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].

[1] https://github.com/trending
[2] https://huggingface.co

3.2. 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 [48]. 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], and CoCa [108].
- *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, 109].
- *Emergence* is the characteristic introduced by scaling the Transformer architecture with large datasets and computational resources [48, 13]. Emergence means that the behavior of the trained AI model is implicitly induced rather than explicitly constructed [13, 109]. In-context learning is an example of emergence in foundation models [109, 110].
- *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, e.g., BERT [93], GPT [15, 17, 18, 19], T5 [90], or many others [111].

- *Transfer learning, adaptation, and fine-tuning* are the defining characteristics of foundation models [13, 112, 110]. 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 pre-trained using unannotated datasets and later adapted using small annotated datasets for specific downstream tasks. A recent survey reviews the various pre-training methods used in deep learning and foundation models on medical data [113].
- *In-context learning* is the ability of a trained foundation model to learn a new task or correct itself using demonstration and without updating model's parameters which is usually done via gradient descent algorithm [17, 110, 114]. In-context learning is a scale-enabled emergent ability that allows foundation models to generalize to new tasks without having to re-train 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*, i.e., finding the most appropriate prompt to allow AI to solve the task at hand [114, 110]. A prompt is a piece of text, image, or symbols inserted in the input of AI so that the given task can be re-formulated as the original task for which the model was trained [17, 109, 110].

3.3. Types of Foundation Models

Foundation models are an emerging area of AI that has shown great promise, e.g., 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 [16, 15, 115, 116]. 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.

3.3.1. Large Language AI 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 [117], 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, 118, 119]. The Transformer architecture enabled self-supervised learning at scale resulting in models like BERT, GPT, GPT-2, GPT-3, GPT-4, LLaMA, T5, and BART [93, 19, 18, 17, 15, 94, 90, 120, 121]. Most of these industry-sponsored models are not open-source for researchers [15, 16]. Recently, BLOOM, a 176B-parameter 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 [111].

3.3.2. Vision-Language AI Models

Vision-language AI can learn to perform various tasks involving images (or videos) and corresponding natural language text [92, 111]. The vision-language models are one step closer to how humans perceive the world, learn about it, and execute various tasks in it [92, 122]. In the following, we describe two 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, 123]. 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, 124]. Visual question-answering tasks typically require a more detailed understanding of the image and complex reasoning than a system producing image captions [124]. The recent foundation models in image-text tasks include Contrastive Language-Image Pre-Training (CLIP), A Large-scale ImaGe and Noisy-Text Embedding (ALIGN), SimVLM, Florence, Flamingo, CoCa, and Clinical-BERT [104, 105, 108, 106, 107, 125].

- *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 pre-training 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 pre-training of vision encoders. The state-of-the-art foundation models include: (1) image classification - UniCL, CLIP, and ALIGN [126, 104, 105], (2) object detection in a given image - ViLD, RegionCLIP, GLIP, Detic, PromptDet, OWL-ViT, OV-DETR, and X-DERT [127, 128, 129, 130, 131, 132, 133], and (3) segmentation of different objects in a given image - LSeg, OpenSeg, CLIPSeg, MaskCLIP, DenseCLIP, and GroupViT [134, 135, 98, 136, 137, 138].

3.4. Training Foundation Models and Generative AI

Foundation models employ two 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 (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, e.g., the BERT language model is trained to predict randomly removed words from sentences or fill in the blank [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 one of two techniques, (1) de-noising or (2) auto-regressive. During the training of the de-noising models, the input is corrupted with noise, and the model is expected to produce noise-free input patterns [90]. The auto-regressive models, after training, can generate the input data piece by piece, iteratively predicting the next element in a sequence given the previous elements [139].

3.5. Challenges and Limitations of Foundation Models

Developing foundation models require 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 [120]. 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 [140].

"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 [141, 142, 15]. The accuracy and integrity of the generated text and images may be challenging to establish using factual data from verified sources [142]. One possible solution is to use an engineered system like *Bing Chat* that also generates links to the actual websites, articles, and reference material[1]. In some cases, the generative AI models can identify their own mistakes [141]. Furthermore, the generative models are sensitive to the form and choice of words, referred to as the "prompt." A prompt may

[1] `https://www.bing.com/`

consist of text, image(s), or symbol(s) inserted in the input of generative AI so that the given task can be re-formulated as the original task for which the model was trained [114, 141]. 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 [114, 141]. 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 tasks-specific AI models, including explainability, robustness, and trustworthiness [88, 143, 13, 78, 144, 39, 79, 41, 142].

4. 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 digital pathology. 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 digital pathology. Because of foundation models' strong adaptation and scalability properties, they can be effectively trained once and modified infinite times to suit various digital pathology tasks. 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 [145]. The authors used a self-supervised hierarchical learning mechanism on 33 cancer site data having approximately 105 million pathology images to predict nine slide-level tasks, including cancer subtyping, survival, and unique morphological phenotypes [145]. Although molecular procedures and analysis 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]. Similarly, attention-based multiple instance learning has accurately predicted biomarkers from cancer pathology slides in a self-supervised learning setting [70]. The authors showed the performance of attention-based multiple instance learning framework for predicting microsatellite instability and mutations in BRAF, KRAS, NRAS, and PIK3CA in colorectal cancer pathology slides [70]. 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 [146]. Such pre-foundation models can be scaled to predict biomarkers directly from the histopathology slides belonging to pan-cancer sites [147]. The Transformer model pre-trained on a large publicly available pathology dataset can be fine-tuned under a weakly-supervised contrastive learning scheme on smaller datasets. Wang *et al.* have shown that such a training framework can outperform the state-of-the-art WSI classification on three different tasks [148]. For the multimodal medical data analysis, modality co-attention Transformers have been shown to outperform other methods in survival predictions by fused learning on WSI data and genomic sequences [149]. Moreover, Transformers are far more robust to adversarial attacks and perturbations in digital pathology than CNNs because of the more robust latent representation of clinically relevant information [79]. The performance and robustness of Transformers-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 (e.g., 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 finding 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 created using data available in the public domain, such as the National Cancer Institute's The Cancer Genome Atlas (TCGA) for genetic data, Clinical Proteomic Tumor Analysis Consortium (CPTAC) for proteomics data, and The Cancer Imaging Archive (TCIA) for imaging data [27, 28, 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 [150]. 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 [151]. 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 digital pathology.

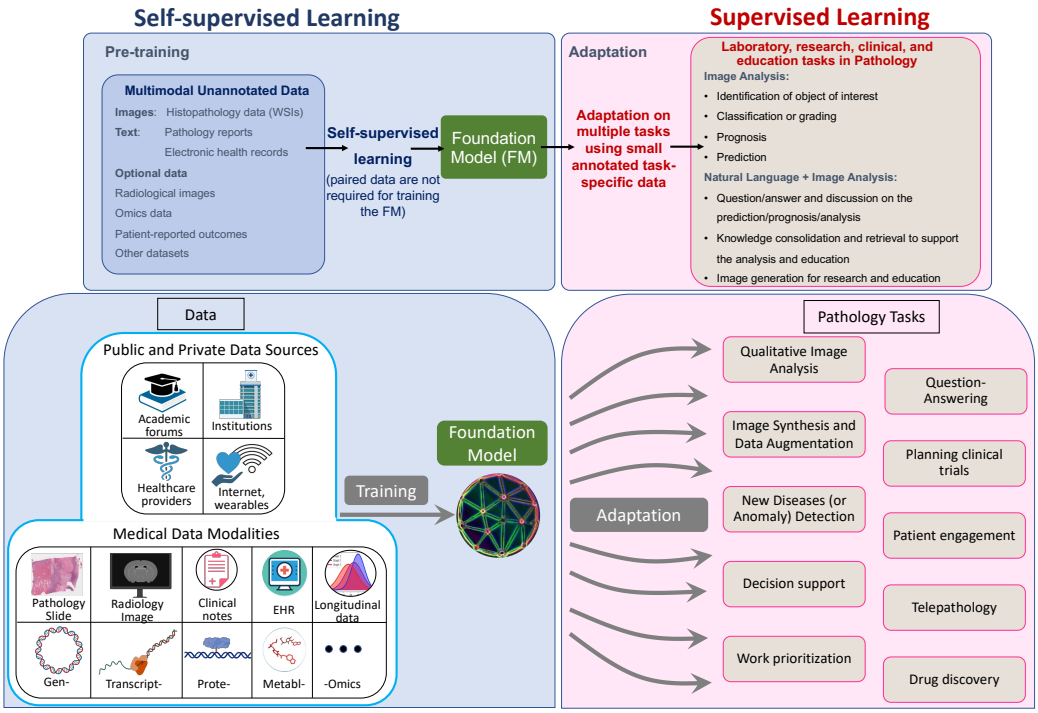

Figure 4: A prospective schematic layout of using foundation models and generative AI for various digital pathology tasks is presented. In our view, other data modalities, e.g., diagnostic radiology or molecular data, if available, can be combined with pathology data in the future to improve model performance.

4.1. 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 (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 multi-site cytology data (cervix, kidney, breast, lung, thyroid, bladder, bone marrow, skin, etc.) to perform various downstream tasks such as malignant cells classification, slide-level stratification, cells location in cytological smears, and cell components identification [152].

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 [153, 154]. Some sample pathology reports generated using ChatGPT (March 14 update) are presented in Fig. 5A, B, and 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 inter-rater variability while improving the quality and consistency of pathology reports [147]. The image analysis and report generation pipeline can serve as the "first pair of eyes" and potentially help pathology labs with workload and 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.

4.2. Image Synthesis, De-noising, 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 [155]. However, we argue that there are enough pathology image data in

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

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 (1) de-noising 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 pathological conditions, thus providing pathologists with a better understanding of disease evolution and enabling more informed treatment planning, (5) education and training - create diverse and realistic examples for educational purposes, thus help trainee pathologists gain experience in diagnosing a wide range of conditions and improve their diagnostic skills without relying on actual patient samples [156], and (6) synthesizing images to study the effects of various factors on disease presentation, such as genetic mutations, environmental factors, or treatment options, thus contribute to a better understanding of disease mechanisms and the development of more effective therapies.

4.3. 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 sub-type 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.

4.4. 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 Fig. 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 non-medical person. In addition, they can educate the patient about the disease entity just diagnosed by the pathologist and possible treatment options.

4.5. 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 [156]. 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 pedagogical 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 the specialists, such as pathologists, for their efficacy and acceptability.

4.6. AI-Driven Standardization in Digital Pathology Workflow

Foundation models and generative AI can help standardize digital pathology 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, morphological measurements, and biomarker quantification, reducing the variability that may arise from manual or semi-automated 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 digital pathology 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 non-structured 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 [157]. 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 digital pathology ecosystem.

5. Conclusion

Foundation models and generative AI have the potential to transform digital pathology, 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 lab workflow, and contributing to the education and training. This review presented an overview of generative AI and foundation models and their potential role in digital pathology. 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 present our perspective on the future role of generative AI and foundation models in digital pathology and future use cases where generative and conversational AI and foundation models can have a transformative impact in digital pathology. 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

GR and MB conceived the initial idea. AW and GR reviewed the literature and wrote the initial draft. All authors reviewed and contributed to the draft.

Funding

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

Declaration of Competing Interest

The authors declare no competing interests at this time.

Data Availability Statement

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

Ethics Approval / Consent to Participate

Not applicable.

References

[1] Dick Stephanie. Artificial Intelligence *Harvard Data Science Review.* 2019;1.

[2] Swanson Kyle, Wu Eric, Zhang Angela, Alizadeh Ash A, Zou James. From patterns to patients: Advances in clinical machine learning for cancer diagnosis, prognosis, and treatment *Cell.* 2023. In Press, Corrected Proof available at: `https://doi.org/10.1016/j.cell.2023.01.035`.

[3] Shanahan Murray. Talking About Large Language Models *arXiv preprint arXiv:2212.03551.* 2023. `https://arxiv.org/abs/2212.03551`.

[4] Tizhoosh Hamid Reza, Pantanowitz Liron. Artificial intelligence and digital pathology: challenges and opportunities *Journal of Pathology Informatics.* 2018;9:38.

[5] Falahkheirkhah Kianoush, Tiwari Saumya, Yeh Kevin, et al. Deepfake Histologic Images for Enhancing Digital Pathology *Laboratory Investigation.* 2023;103.

[6] Scopio Team . Scopio Labs 2023. available at: https://scopiolabs.com/. Last accessed on Mar 31, 2023.

[7] Paige Team . Pathology Artificial Intelligence Guidance Engine (PAIGE) 2023. available at: https://paige.ai/. Last accessed on Mar 31, 2023.

[8] Aiforia Team . AI for Image Analysis (AIFORIA) 2023. available at: https://www.aiforia.com/. Last accessed on Mar 31, 2023.

[9] Gupta Rajarsi, Kurc Tahsin, Sharma Ashish, Almeida Jonas S, Saltz Joel. The emergence of pathomics *Current Pathobiology Reports.* 2019;7:73–84.

[10] Drogt Jojanneke, Milota Megan, Vos Shoko, Bredenoord Annelien, Jongsma Karin. Integrating artificial intelligence in pathology: a qualitative interview study of users' experiences and expectations *Modern Pathology.* 2022;35:1540–1550.

[11] Kim Inho, Kang Kyungmin, Song Youngjae, Kim Tae-Jung. Application of Artificial Intelligence in Pathology: Trends and Challenges *Diagnostics.* 2022;12:2794.

[12] Patel Ankush U, Shaker Nada, Mohanty Sambit, et al. Cultivating Clinical Clarity through Computer Vision: A Current Perspective on Whole Slide Imaging and Artificial Intelligence *Diagnostics.* 2022;12:1778.

[13] Bommasani Rishi, Hudson Drew A, Adeli Ehsan, et al. On the opportunities and risks of foundation models *arXiv preprint arXiv:2108.07258.* 2021. https://arxiv.org/abs/2108.07258.

[14] Moor Michael, Banerjee Oishi, Abad Zahra Shakeri Hossein, et al. Foundation Models for Generalist Medical Artificial Intelligence *Nature.* 2023;616:259-265.

[15] OpenAI . GPT-4 2023. Available at: https://openai.com/research/gpt-4. Last accessed on April 5, 2023.

[16] OpenAI . Introducing ChatGPT 2022. https://openai.com/blog/chatgpt. Last accessed on: March 10, 2023.

[17] Brown Tom, Mann Benjamin, Ryder Nick, et al. Language Models are Few-Shot Learners in *Advances in Neural Information Processing Systems*;33:1877–1901 2020.

[18] Radford Alec, Wu Jeffrey, Child Rewon, Luan David, Amodei Dario, Sutskever Ilya. Better language models and their implications 2019. available at: `https://openai.com/research/better-language-models`. Last accessed on April 5, 2023.

[19] Radford Alec, Narasimhan Karthik, Salimans Tim, Sutskever Ilya, others . Improving language understanding by generative pre-training 2018. Available at `https://openai.com/research/language-unsupervised`. Last accessed: April 5, 2023.

[20] Ouyang Long, Wu Jeffrey, Jiang Xu, et al. Training language models to follow instructions with human feedback in *Advances in Neural Information Processing Systems* (Oh Alice H., Agarwal Alekh, Belgrave Danielle, Cho Kyunghyun. , eds.) 2022.

[21] Bera Kaustav, Schalper Kurt A, Rimm David L, Velcheti Vamsidhar, Madabhushi Anant. Artificial intelligence in digital pathology—new tools for diagnosis and precision oncology *Nature reviews Clinical oncology.* 2019;16:703–715.

[22] Alsaafin Areej, Safarpoor Amir, Sikaroudi Milad, Hipp Jason D, Tizhoosh HR. Learning to predict RNA sequence expressions from whole slide images with applications for search and classification *Communications Biology.* 2023;6:304.

[23] Cifci Didem, Veldhuizen Gregory P, Foersch Sebastian, Kather Jakob Nikolas. AI in Computational Pathology of Cancer: Improving Diagnostic Workflows and Clinical Outcomes? *Annual Review of Cancer Biology.* 2023;7.

[24] Adam George, Rampášek Ladislav, Safikhani Zhaleh, Smirnov Petr, Haibe-Kains Benjamin, Goldenberg Anna. Machine learning approaches to drug response prediction: challenges and recent progress *NPJ Precision Oncology.* 2020;4:19.

[25] Demetriou Demetra, Hull Rodney, Kgoebane-Maseko Mmamoletla, Lockhat Zarina, Dlamini Zodwa. AI-Enhanced Digital Pathology and

Radiogenomics in Precision Oncology in *Artificial Intelligence and Precision Oncology: Bridging Cancer Research and Clinical Decision Support*:93–113Springer 2023.

[26] Pantanowitz Liron, Hartman Douglas, Qi Yan, et al. Accuracy and efficiency of an artificial intelligence tool when counting breast mitoses *Diagnostic Pathology.* 2020;15:1–10.

[27] Tomczak Katarzyna, Czerwińska Patrycja, Wiznerowicz Maciej. Review The Cancer Genome Atlas (TCGA): An immeasurable source of knowledge *Contemporary Oncology.* 2015;2015:68–77.

[28] Ellis Matthew J., Gillette Michael, Carr Steven A., et al. Connecting Genomic Alterations to Cancer Biology with Proteomics: The NCI Clinical Proteomic Tumor Analysis Consortium *Cancer Discovery.* 2013;3:1108-1112.

[29] Clark Kenneth, Vendt Bruce, Smith Kirk, et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository *Journal of digital imaging.* 2013;26:1045–1057.

[30] Abels Esther, Pantanowitz Liron, Aeffner Famke, et al. Computational pathology definitions, best practices, and recommendations for regulatory guidance: a white paper from the Digital Pathology Association *The Journal of pathology.* 2019;249:286–294.

[31] Kiehl Tim-Rasmus. Digital and Computational Pathology: A Specialty Reimagined in *The Future Circle of Healthcare: AI, 3D Printing, Longevity, Ethics, and Uncertainty Mitigation*:227–250Springer 2022.

[32] Bai Bijie, Yang Xilin, Li Yuzhu, Zhang Yijie, Pillar Nir, Ozcan Aydogan. Deep learning-enabled virtual histological staining of biological samples *Light: Science & Applications.* 2023;12:57.

[33] Cifci Didem, Foersch Sebastian, Kather Jakob Nikolas. Artificial intelligence to identify genetic alterations in conventional histopathology *The Journal of Pathology.* 2022;257:430–444.

[34] Echle Amelie, Rindtorff Niklas Timon, Brinker Titus Josef, Luedde Tom, Pearson Alexander Thomas, Kather Jakob Nikolas. Deep learning

in cancer pathology: a new generation of clinical biomarkers *British Journal of Cancer.* 2021;124:686–696.

[35] Cui Miao, Zhang David Y. Artificial intelligence and computational pathology *Laboratory Investigation.* 2021;101:412–422.

[36] Gurcan Metin N, Boucheron Laura E, Can Ali, Madabhushi Anant, Rajpoot Nasir M, Yener Bulent. Histopathological image analysis: A review *IEEE reviews in biomedical engineering.* 2009;2:147–171.

[37] Irshad Humayun, Veillard Antoine, Roux Ludovic, Racoceanu Daniel. Methods for nuclei detection, segmentation, and classification in digital histopathology: a review—current status and future potential *IEEE Reviews in Biomedical Engineering.* 2013;7:97–114.

[38] Waqas Asim, Dera Dimah, Rasool Ghulam, Bouaynaya Nidhal Carla, Fathallah-Shaykh Hassan M. Brain Tumor Segmentation and Surveillance with Deep Artificial Neural Networks *Deep Learning for Biomedical Data Analysis.* 2021:311–350.

[39] Dera Dimah, Bouaynaya Nidhal, Rasool Ghulam, Shterenberg R, Fathallah-Shaykh Hassan. PremiUm-CNN: Propagating Uncertainty Towards Robust Convolutional Neural Networks *IEEE Transactions on Signal Processing.* 2021.

[40] Waqas Asim, Farooq Hamza, Bouaynaya Nidhal C, Rasool Ghulam. Exploring Robust Architectures for Deep Artificial Neural Networks *Communications Engineering.* 2022;1:46.

[41] Ahmed Sabeen, Dera Dimah, Hassan Saud Ul, Bouaynaya Nidhal, Rasool Ghulam. Failure detection in deep neural networks for medical imaging *Frontiers in Medical Technology.* 2022;4.

[42] Albahra Samer, Gorbett Tom, Robertson Scott, et al. Artificial Intelligence and Machine Learning Overview in Pathology & Laboratory Medicine: A General Review of Data Preprocessing and Basic Supervised Concepts in *Seminars in Diagnostic Pathology*Elsevier 2023.

[43] Shen Dinggang, Wu Guorong, Suk Heung-Il. Deep learning in medical image analysis *Annual Review of Biomedical Engineering.* 2017;19:221–248.

[44] Waqas Asim, Tripathi Aakash, Ramachandran Ravi P, Stewart Paul, Rasool Ghulam. Multimodal Data Integration for Oncology in the Era of Deep Neural Networks: A Review *arXiv preprint arXiv:2303.06471.* 2023. `https://arxiv.org/abs/2303.06471`.

[45] Adnan Mohammed, Kalra Shivam, Tizhoosh Hamid R. Representation learning of histopathology images using graph neural networks in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*:988–989 2020.

[46] Wang Jingwen, Chen Richard J, Lu Ming Y, Baras Alexander, Mahmood Faisal. Weakly supervised prostate TMA classification via graph convolutional networks in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*:239–243IEEE 2020.

[47] Chen Richard J, Lu Ming Y, Wang Jingwen, et al. Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis *IEEE Transactions on Medical Imaging.* 2020.

[48] Vaswani Ashish, Shazeer Noam, Parmar Niki, et al. Attention is All you Need in *Advances in Neural Information Processing Systems* (Guyon I., Luxburg U. Von, Bengio S., et al. , eds.);30Curran Associates, Inc. 2017.

[49] Ahmed Sabeen, Nielsen Ian E, Tripathi Aakash, Siddiqui Shamoon, Rasool Ghulam, Ramachandran Ravi P. Transformers in Time-series Analysis: A Tutorial *arXiv preprint arXiv:2205.01138.* 2022. `https://arxiv.org/abs/2205.01138`.

[50] Aubreville Marc, Stathonikos Nikolas, Bertram Christof A, et al. Mitosis domain generalization in histopathology images—The MIDOG challenge *Medical Image Analysis.* 2023;84:102699.

[51] Bulten Wouter, Kartasalo Kimmo, Chen Po-Hsuan Cameron, et al. Artificial intelligence for diagnosis and Gleason grading of prostate cancer: the PANDA challenge *Nature Medicine.* 2022;28:154–163.

[52] Ma Xiao, Jia Fucang. Brain tumor classification with multimodal MR and pathology images in *Brainlesion: Glioma, Multiple Sclero-*

sis, Stroke and Traumatic Brain Injuries: 5th International Workshop, BrainLes 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 17, 2019, Revised Selected Papers, Part II 5:343–352Springer 2020.

[53] Bychkov Dmitrii, Turkki Riku, Haglund Caj, Linder Nina, Lundin Johan. Deep learning for tissue microarray image-based outcome prediction in patients with colorectal cancer in *Medical Imaging 2016: Digital Pathology*;9791:298–303SPIE 2016.

[54] Braman Nathaniel, Gordon Jacob WH, Goossens Emery T, Willis Caleb, Stumpe Martin C, Venkataraman Jagadish. Deep orthogonal fusion: multimodal prognostic biomarker discovery integrating radiology, pathology, genomic, and clinical data in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part V 24*:667–677Springer 2021.

[55] Veta Mitko, Van Diest Paul J, Willems Stefan M, et al. Assessment of algorithms for mitosis detection in breast cancer histopathology images *Medical Image Analysis.* 2015;20:237–248.

[56] Jardim-Perassi Bruna V, Mu Wei, Huang Suning, et al. Deep-learning and MR images to target hypoxic habitats with evofosfamide in preclinical models of sarcoma *Theranostics.* 2021;11:5313.

[57] Mahmood Faisal, Borders Daniel, Chen Richard J, et al. Deep adversarial training for multi-organ nuclei segmentation in histopathology images *IEEE Transactions on Medical Imaging.* 2019;39:3257–3267.

[58] Guo Ruoyu, Xie Kunzi, Pagnucco Maurice, Song Yang. SAC-Net: Learning with weak and noisy labels in histopathology image segmentation *Medical Image Analysis.* 2023:102790.

[59] Ibrahim Asmaa, Gamble Paul, Jaroensri Ronnachai, et al. Artificial intelligence in digital breast pathology: techniques and applications *The Breast.* 2020;49:267–273.

[60] Edara Deepak Chowdary, Vanukuri Lakshmi Prasanna, Sistla Venkatramaphanikumar, Kolli Venkata Krishna Kishore. Sentiment analysis

and text categorization of cancer medical records with LSTM *Journal of Ambient Intelligence and Humanized Computing.* 2019:1–17.

[61] Rajeev R, Samath J Abdul, Karthikeyan NK. An intelligent recurrent neural network with long short-term memory (LSTM) BASED batch normalization for medical image denoising *Journal of Medical Systems.* 2019;43:1–10.

[62] Liu Sicen, Wang Xiaolong, Xiang Yang, Xu Hui, Wang Hui, Tang Buzhou. Multi-channel fusion LSTM for medical event prediction using EHRs *Journal of Biomedical Informatics.* 2022;127:104011.

[63] Leevy Joffrey L, Khoshgoftaar Taghi M. A short survey of LSTM models for de-identification of medical free text in *2020 IEEE 6th International Conference on Collaboration and Internet Computing (CIC)*:117–124IEEE 2020.

[64] Gao Yang, Phillips Jeff M, Zheng Yan, Min Renqiang, Fletcher P Thomas, Gerig Guido. Fully convolutional structured LSTM networks for joint 4D medical image segmentation in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI)*:1104–1108IEEE 2018.

[65] Anand Deepak, Gadiya Shrey, Sethi Amit. Histographs: graphs in histopathology in *Medical Imaging 2020: Digital Pathology*;11320:150–155SPIE 2020.

[66] Zhou Yanning, Graham Simon, Alemi Koohbanani Navid, Shaban Muhammad, Heng Pheng-Ann, Rajpoot Nasir. CGC-Net: Cell Graph Convolutional Network for Grading of Colorectal Cancer Histology Images in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*:0–0 2019.

[67] Studer Linda, Wallau Jannis, Dawson Heather, Zlobec Inti, Fischer Andreas. Classification of intestinal gland cell-graphs using graph neural networks in *2020 25th International conference on pattern recognition (ICPR)*:3636–3643IEEE 2021.

[68] Sureka Mookund, Patil Abhijeet, Anand Deepak, Sethi Amit. Visualization for histopathology images using graph convolutional neural net-

works in *2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE)*:331–335IEEE 2020.

[69] Ahmedt-Aristizabal David, Armin Mohammad Ali, Denman Simon, Fookes Clinton, Petersson Lars. A survey on graph-based deep learning for computational histopathology *Computerized Medical Imaging and Graphics.* 2022;95:102027.

[70] Niehues Jan Moritz, Quirke Philip, West Nicholas P, et al. Generalizable biomarker prediction from cancer pathology slides with self-supervised deep learning: A retrospective multi-centric study *Cell Reports Medicine.* 2023.

[71] Vu Quoc Dang, Rajpoot Kashif, Raza Shan E Ahmed, Rajpoot Nasir. Handcrafted Histological Transformer (H2T): Unsupervised representation of whole slide images *Medical Image Analysis.* 2023:102743.

[72] Huang Shih-Cheng, Pareek Anuj, Jensen Malte, Lungren Matthew P, Yeung Serena, Chaudhari Akshay S. Self-supervised learning for medical image classification: a systematic review and implementation guidelines *NPJ Digital Medicine.* 2023;6:74.

[73] Azad Reza, Kazerouni Amirhossein, Heidari Moein, et al. Advances in medical image analysis with vision transformers: A comprehensive review *arXiv preprint arXiv:2301.03505.* 2023.

[74] Xia Kun, Wang Jinzhuo. Recent advances of transformers in medical image analysis: a comprehensive review *MedComm–Future Medicine.* 2023;2:e38.

[75] Computational Pathology & AI . FDA Has Cleared Four Pathology AI Algorithms 2023. Available at: `https://www.pathologynews.com/computational-pathology-ai/fda-has-now-cleared-more-than-500-healthcare-ai-algorithms-four-of-which-are-for-pathology/`. Last accessed on Jul 20, 2023.

[76] FDA . FDA Approved (AI/ML)-Enabled Medical Devices 2022. Available at: `https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices`. Last accessed on Jul 20, 2023.

[77] Dera Dimah, Rasool Ghulam, Bouaynaya Nidhal. Extended Variational Inference for Propagating Uncertainty in Convolutional Neural Networks in *2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP)*:1–6IEEE 2019.

[78] Carannante Giuseppina, Dera Dimah, Bouaynaya Nidhal C, Fathallah-Shaykh Hassan M, Rasool Ghulam. Trustworthy Medical Segmentation with Uncertainty Estimation *arXiv preprint arXiv:2111.05978.* 2021. `https://arxiv.org/abs/2111.05978`.

[79] Ghaffari Laleh Narmin, Truhn Daniel, Veldhuizen Gregory Patrick, et al. Adversarial attacks and adversarial robustness in computational pathology *Nature Communications.* 2022;13:5711.

[80] Alwosheel Ahmad, Cranenburgh Sander, Chorus Caspar G. Is your dataset big enough? Sample size requirements when using artificial neural networks for discrete choice analysis *Journal of Choice Modelling.* 2018;28:167–182.

[81] Khan Hikmat, Bouaynaya Nidhal Carla, Rasool Ghulam. Adversarially Robust Continual Learning in *2022 International Joint Conference on Neural Networks (IJCNN)*:1–8IEEE 2022.

[82] Ahn Euijoon, Kumar Ashnil, Feng Dagan, Fulham Michael, Kim Jinman. Unsupervised deep transfer feature learning for medical image classification in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI)*:1915–1918IEEE 2019.

[83] Boehm Kevin M, Khosravi Pegah, Vanguri Rami, Gao Jianjiong, Shah Sohrab P. Harnessing multimodal data integration to advance precision oncology *Nature Reviews Cancer.* 2021:1–13.

[84] Vanguri Rami S, Luo Jia, Aukerman Andrew T, 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 *Nature Cancer.* 2022;3:1151–1164.

[85] McKinney Scott Mayer, Sieniek Marcin, Godbole Varun, et al. International evaluation of an AI system for breast cancer screening *Nature.* 2020;577:89–94.

[86] Haibe-Kains Benjamin, Adam George Alexandru, Hosny Ahmed, et al. Transparency and reproducibility in artificial intelligence *Nature.* 2020;586:E14–E16.

[87] McKinney Scott Mayer, Karthikesalingam Alan, Tse Daniel, et al. Reply to: Transparency and reproducibility in artificial intelligence *Nature.* 2020;586:E17–E18.

[88] Nielsen Ian E, Dera Dimah, Rasool Ghulam, Ramachandran Ravi P, Bouaynaya Nidhal Carla. Robust explainability: A tutorial on gradient-based attribution methods for deep neural networks *IEEE Signal Processing Magazine.* 2022;39:73–84.

[89] Nielsen Ian E, Ramachandran Ravi P, Bouaynaya Nidhal, Fathallah-Shaykh Hassan M, Rasool Ghulam. EvalAttAI: A Holistic Approach to Evaluating Attribution Maps in Robust and Non-Robust Models *arXiv preprint arXiv:2303.08866.* 2023. `https://arxiv.org/abs/2303.08866`.

[90] Raffel Colin, Shazeer Noam, Roberts Adam, et al. Exploring the limits of transfer learning with a unified text-to-text Transformer *The Journal of Machine Learning Research.* 2020;21:5485–5551.

[91] Scao Teven Le, Fan Angela, Akiki Christopher, et al. BLOOM: A 176b-parameter open-access multilingual language model *arXiv preprint arXiv:2211.05100.* 2022.

[92] Gan Zhe, Li Linjie, Li Chunyuan, et al. Vision-language pre-training: Basics, recent advances, and future trends *Foundations and Trends in Computer Graphics and Vision.* 2022;14:163–352.

[93] Devlin Jacob, Chang Ming-Wei, Lee Kenton, Toutanova Kristina. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*:4171–4186Association for Computational Linguistics 2019.

[94] Liu Yinhan, Ott Myle, Goyal Naman, et al. RoBERTa: A Robustly Optimized BERT Pretraining Approach 2020.

[95] Dosovitskiy Alexey, Beyer Lucas, Kolesnikov Alexander, et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale *arXiv preprint arXiv:2010.11929.* 2021.

[96] Cheng Bowen, Misra Ishan, Schwing Alexander G, Kirillov Alexander, Girdhar Rohit. Masked-attention mask Transformer for universal image segmentation in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*:1290–1299 2022.

[97] Jain Jitesh, Li Jiachen, Chiu Mang Tik, Hassani Ali, Orlov Nikita, Shi Humphrey. Oneformer: One transformer to rule universal image segmentation in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*:2989–2998 2023.

[98] Lüddecke Timo, Ecker Alexander. Image Segmentation Using Text and Image Prompts in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*:7076–7086IEEE 2022.

[99] Jaegle Andrew, Borgeaud Sebastian, Alayrac Jean-Baptiste, et al. Perceiver IO: A General Architecture for Structured Inputs & Outputs *International Conference on Learning Representations.* 2022.

[100] Surís Dídac, Menon Sachit, Vondrick Carl. Vipergpt: Visual inference via python execution for reasoning *arXiv preprint arXiv:2303.08128.* 2023. `https://arxiv.org/abs/2303.08128`.

[101] Liu Haotian, Li Chunyuan, Wu Qingyang, Lee Yong Jae. Visual instruction tuning *arXiv preprint arXiv:2304.08485.* 2023. `https://arxiv.org/abs/2304.08485`.

[102] Li Junnan, Li Dongxu, Savarese Silvio, Hoi Steven. BLIP-2: Bootstrapping Language-Image Pre-training with Frozen Image Encoders and Large Language Models *arXiv preprint arXiv:2301.12597.* 2023. `https://arxiv.org/abs/2301.12597`.

[103] Hudson Drew A, Zitnick Larry. Generative adversarial transformers in *International conference on machine learning*:4487–4499PMLR 2021.

[104] Radford Alec, Kim Jong Wook, Hallacy Chris, et al. Learning Transferable Visual Models From Natural Language Supervision in *Proceedings of the 38th International Conference on Machine Learning* (Meila

Marina, Zhang Tong. , eds.);139 of *Proceedings of Machine Learning Research*:8748–8763PMLR 2021.

[105] Jia Chao, Yang Yinfei, Xia Ye, et al. Scaling Up Visual and Vision-Language Representation Learning With Noisy Text Supervision in *Proceedings of the 38th International Conference on Machine Learning* (Meila Marina, Zhang Tong. , eds.);139 of *Proceedings of Machine Learning Research*:4904–4916PMLR 2021.

[106] Wang Zirui, Yu Jiahui, Yu Adams Wei, Dai Zihang, Tsvetkov Yulia, Cao Yuan. SimVLM: Simple Visual Language Model Pretraining with Weak Supervision in *International Conference on Learning Representations* 2022.

[107] Alayrac Jean-Baptiste, Donahue Jeff, Luc Pauline, et al. Flamingo: a Visual Language Model for Few-Shot Learning in *Advances in Neural Information Processing Systems* (Oh Alice H., Agarwal Alekh, Belgrave Danielle, Cho Kyunghyun. , eds.) 2022.

[108] Yu Jiahui, Wang Zirui, Vasudevan Vijay, Yeung Legg, Seyedhosseini Mojtaba, Wu Yonghui. CoCa: Contrastive Captioners are Image-Text Foundation Models *Transactions on Machine Learning Research.* 2022;Aug 2022.

[109] Lester Brian, Al-Rfou Rami, Constant Noah. The Power of Scale for Parameter-Efficient Prompt Tuning in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*:3045–3059Association for Computational Linguistics 2021.

[110] Wei Jason, Tay Yi, Bommasani Rishi, et al. Emergent Abilities of Large Language Models *Transactions on Machine Learning Research.* 2022. Survey Certification.

[111] Zhou Ce, Li Qian, Li Chen, et al. A Comprehensive Survey on Pretrained Foundation Models: A History from BERT to ChatGPT 2023. `https://arxiv.org/abs/2302.09419`.

[112] Willemink Martin J, Roth Holger R, Sandfort Veit. Toward Foundational Deep Learning Models for Medical Imaging in the New Era of Transformer Networks *Radiology: Artificial Intelligence.* 2022;4:e210284.

[113] Qiu Yixuan, Lin Feng, Chen Weitong, Xu Miao. Pre-training in Medical Data: A Survey *Machine Intelligence Research.* 2023:1–33.

[114] Liu Pengfei, Yuan Weizhe, Fu Jinlan, Jiang Zhengbao, Hayashi Hiroaki, Neubig Graham. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing *ACM Computing Surveys.* 2023;55:1–35.

[115] Rombach Robin, Blattmann Andreas, Lorenz Dominik, Esser Patrick, Ommer Björn. High-Resolution Image Synthesis With Latent Diffusion Models in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*:10684-10695 2022.

[116] Ramesh Aditya, Dhariwal Prafulla, Nichol Alex, Chu Casey, Chen Mark. Hierarchical Text-Conditional Image Generation with CLIP Latents 2022. https://arxiv.org/abs/2204.06125.

[117] Turian Joseph, Ratinov Lev, Bengio Yoshua. Word representations: a simple and general method for semi-supervised learning in *Proceedings of the 48th annual meeting of the association for computational linguistics*:384–394 2010.

[118] Peters Matthew E., Neumann Mark, Iyyer Mohit, et al. Deep contextualized word representations 2018.

[119] Howard Jeremy, Ruder Sebastian. Universal Language Model Fine-tuning for Text Classification 2018.

[120] Touvron Hugo, Lavril Thibaut, Izacard Gautier, et al. LLaMA: Open and Efficient Foundation Language Models 2023. https://arxiv.org/abs/2302.13971.

[121] Lewis Mike, Liu Yinhan, Goyal Naman, et al. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*:7871–7880 2020.

[122] Lu Haoyu, Zhou Qiongyi, Fei Nanyi, et al. Multimodal foundation models are better simulators of the human brain *arXiv preprint arXiv:2208.08263.* 2022. `https://arxiv.org/abs/2208.08263`.

[123] Dirik Alara, Paul Sayak. A Dive into Vision-Language Models 2023.

[124] Hudson Drew A., Manning Christopher D.. GQA: A New Dataset for Real-World Visual Reasoning and Compositional Question Answering in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 2019.

[125] Yan Bin, Pei Mingtao. Clinical-BERT: Vision-Language Pre-training for Radiograph Diagnosis and Reports Generation in *Proceedings of the AAAI Conference on Artificial Intelligence*;36:2982–2990 2022.

[126] Yang Jianwei, Li Chunyuan, Zhang Pengchuan, et al. Unified Contrastive Learning in Image-Text-Label Space in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*:19163-19173 2022.

[127] Gu Xiuye, Lin Tsung-Yi, Kuo Weicheng, Cui Yin. Open-vocabulary Object Detection via Vision and Language Knowledge Distillation in *International Conference on Learning Representations* 2022.

[128] Zhong Y., Yang J., Zhang P., et al. RegionCLIP: Region-based Language-Image Pretraining in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*(Los Alamitos, CA, USA):16772-16782IEEE Computer Society 2022.

[129] Li L., Zhang P., Zhang H., et al. Grounded Language-Image Pre-training in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*(Los Alamitos, CA, USA):10955-10965IEEE Computer Society 2022.

[130] Zhou Xingyi, Girdhar Rohit, Joulin Armand, Krähenbühl Philipp, Misra Ishan. 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):350–368Springer-Verlag 2022.

[131] Minderer Matthias, Gritsenko Alexey, Stone Austin, 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):728–755Springer-Verlag 2022.

[132] Zang Yuhang, Li Wei, Zhou Kaiyang, Huang Chen, Loy Chen Change. 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):106–122Springer-Verlag 2022.

[133] Cai Zhaowei, Kwon Gukyeong, Ravichandran Avinash, 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):290–308Springer-Verlag 2022.

[134] Li Boyi, Weinberger Kilian Q, Belongie Serge, Koltun Vladlen, Ranftl Rene. Language-driven Semantic Segmentation in *International Conference on Learning Representations* 2022.

[135] Ghiasi Golnaz, Gu Xiuye, Cui Yin, Lin Tsung-Yi. Scaling Open-Vocabulary Image Segmentation With Image-Level Labels in *Computer Vision – ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXVI*(Berlin, Heidelberg):540–557Springer-Verlag 2022.

[136] Zhou Chong, Loy Chen Change, Dai Bo. 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):696–712Springer-Verlag 2022.

[137] Rao Y., Zhao W., Chen G., et al. DenseCLIP: Language-Guided Dense Prediction with Context-Aware Prompting in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*(Los Alamitos, CA, USA):18061-18070IEEE Computer Society 2022.

[138] Xu Jiarui, De Mello Shalini, Liu Sifei, et al. GroupViT: Semantic Segmentation Emerges from Text Supervision in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*:18113-18123 2022.

[139] Ramesh Aditya, Pavlov Mikhail, Goh Gabriel, et al. Zero-shot text-to-image generation in *International Conference on Machine Learning*:8821–8831PMLR 2021.

[140] Smith Shaden, Patwary Mostofa, Norick Brandon, et al. Using DeepSpeed and Megatron to Train Megatron-Turing NLG 530B, A Large-Scale Generative Language Model *arXiv preprint arXiv:2201.11990.* 2022. `https://arxiv.org/abs/2201.11990`.

[141] Lee Peter, Bubeck Sebastien, Petro Joseph. Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine *New England Journal of Medicine.* 2023;388:1233–1239.

[142] Alkaissi Hussam, McFarlane Samy I. Artificial Hallucinations in ChatGPT: Implications in Scientific Writing *Cureus.* 2023;15:e35179.

[143] Jin Weina, Li Xiaoxiao, Fatehi Mostafa, Hamarneh Ghassan. Guidelines and evaluation of clinical explainable AI in medical image analysis *Medical Image Analysis.* 2023;84:102684.

[144] Carannante Giuseppina, Dera Dimah, Rasool Ghulam, Bouaynaya Nidhal C. Self-Compression in Bayesian Neural Networks in *2020 IEEE 30th International Workshop on Machine Learning for Signal Processing (MLSP)*:1–6IEEE 2020.

[145] Chen Richard J, Chen Chengkuan, Li Yicong, et al. Scaling vision transformers to gigapixel images via hierarchical self-supervised learning in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*:16144–16155 2022.

[146] CUI Yufei, Liu Ziquan, Liu Xiangyu, et al. Bayes-MIL: A New Probabilistic Perspective on Attention-based Multiple Instance Learning for Whole Slide Images in *The Eleventh International Conference on Learning Representations* 2023.

[147] Shmatko Artem, Ghaffari Laleh Narmin, Gerstung Moritz, Kather Jakob Nikolas. Artificial intelligence in histopathology: enhancing cancer research and clinical oncology *Nature Cancer.* 2022;3:1026–1038.

[148] Wang Xiyue, Xiang Jinxi, Zhang Jun, et al. SCL-WC: Cross-slide contrastive learning for weakly-supervised whole-slide image classification *Advances in Neural Information Processing Systems.* 2022;35:18009–18021.

[149] Chen Richard J, Lu Ming Y, Weng Wei-Hung, et al. Multimodal co-attention transformer for survival prediction in gigapixel whole slide images in *Proceedings of the IEEE/CVF International Conference on Computer Vision*:4015–4025 2021.

[150] Lipkova Jana, Chen Richard J, Chen Bowen, et al. Artificial intelligence for multimodal data integration in oncology *Cancer Cell.* 2022;40:1095–1110.

[151] Fu Yu, Jung Alexander W, Torne Ramon Viñas, et al. Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis *Nature Cancer.* 2020;1:800–810.

[152] Jiang Hao, Zhou Yanning, Lin Yi, Chan Ronald C.K., Liu Jiang, Chen Hao. Deep learning for computational cytology: A survey *Medical Image Analysis.* 2023;84:102691.

[153] Ma Ruibin, Chen Po-Hsuan Cameron, Li Gang, et al. Human-centric metric for accelerating pathology reports annotation *arXiv preprint arXiv:1911.01226.* 2019. `https://arxiv.org/abs/1911.01226`.

[154] Sinha Ranwir K, Roy Asitava Deb, Kumar Nikhil, Mondal Himel, Sinha Ranwir. Applicability of ChatGPT in Assisting to Solve Higher Order Problems in Pathology *Cureus.* 2023;15.

[155] Kather Jakob Nikolas, Ghaffari Laleh Narmin, Foersch Sebastian, Truhn Daniel. Medical domain knowledge in domain-agnostic generative AI *NPJ Digital Medicine.* 2022;5:90.

[156] Hassell Lewis A, Absar Syeda Fatima, Chauhan Chhavi, et al. Pathology education powered by virtual and digital transformation *Arch Pathol Lab Med.* 2022.

[157] Gibson Blake A, McKinnon Elizabeth, Bentley Rex C, et al. Communicating Certainty in Pathology ReportsInterpretation Differences Among Staff Pathologists, Clinicians, and Residents in a Multicenter Study *Archives of Pathology & Laboratory Medicine.* 2022;146:886–893.