Content Generation Models in Computational Pathology: A Comprehensive Survey on Methods, Applications, and Challenges

Yuan Zhang, Xinfeng Zhang, Xiaoming Qi, Xinyu Wu, Feng Chen, Guanyu Yang, *Senior Member, IEEE*, and Huazhu Fu, *Senior Member, IEEE*

***Abstract*— Content generation modeling has emerged as a promising direction in computational pathology, offering capabilities such as data-efficient learning, synthetic data augmentation, and task-oriented generation across diverse diagnostic tasks. This review provides a comprehensive synthesis of recent progress in the field, organized into four key domains: image generation, text generation, molecular profile–morphology generation, and other specialized generation applications. By analyzing over 150 representative studies, we trace the evolution of content generation architectures—from early generative adversarial networks to recent advances in diffusion models and generative vision–language models. We further examine the datasets and evaluation protocols commonly used in this domain and highlight ongoing limitations, including challenges in generating high-fidelity whole slide images, clinical interpretability, and concerns related to the ethical and legal implications of synthetic data. The review concludes with a discussion of open challenges and prospective research directions, with an emphasis on developing integrated and clinically deployable generation systems. This work aims to provide a foundational reference for researchers and practitioners developing content generation models in computational pathology.**

***Index Terms*— Computational Pathology, Generative Models, Diffusion Models, Generative Adversarial Networks, Synthetic Medical Images**

I. Introduction

Content generation models are machine learning models that learn underlying data distributions and generate new biomedical instances that preserve the essential characteristics of the training data. They have become a transformative force in medical image analysis, offering capabilities that extend beyond traditional discriminative learning [1], [2]. By modeling complex data distributions to synthesize high-fidelity content, these models enable a wide range of downstream tasks, such as data augmentation, cross-modal translation, and representation learning [3]. The diversity of these generation targets in computational pathology is illustrated in Fig. 1. Their utility is especially pronounced in computational pathology [4], a domain characterized by high-resolution whole-slide images (WSIs), heterogeneous tissue morphology, and significant annotation burdens.

Yuan Zhang and Xinfeng Zhang contributed equally to this work. Corresponding author: Guanyu Yang and Huazhu Fu.

Y. Zhang and G. Yang are with the Key Laboratory of New Generation Artificial Intelligence Technology and Its Interdisciplinary Applications (Southeast University), Ministry of Education, Nanjing, China (e-mail: yuanzhang_@seu.edu.cn, yang.list@seu.edu.cn).

X. Zhang is with the School of Biomedical Engineering, Tsinghua University, Beijing, China (e-mail: zhang-xf22@mails.tsinghua.edu.cn).

X. Qi is with the Department of Biomedical Engineering and Department of Electrical and Computer Engineering, National University of Singapore, Singapore (e-mail: qixiaoming12138@163.com).

X. Wu is with the School of Information Science and Engineering, Southeast University, Nanjing, China (e-mail: xinyuwu@seu.edu.cn).

F. Chen is with the Department of Biostatistics, Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing, China (e-mail: fengchen@njmu.edu.cn).

H. Fu is with the Institute of High-Performance Computing, Agency for Science, Technology and Research, Singapore (e-mail: hzfu@ieee.org).

Fig. 1. Diverse generation targets in computational pathology. This figure illustrates the spectrum of generative tasks, covering image and mask synthesis, spatial layout generation, textual report generation, and visual question answering.

Computational pathology has rapidly evolved into a data-intensive field driven by advances in deep learning and whole slide imaging [5]. However, this progress also reveals core limitations in data availability, generalization, and robustness across clinical domains [6]. Generative models such as generative adversarial networks (GAN) [7]–[10], variational autoencoders (VAE) [11]–[14], diffusion models [15]–[17], and foundation models like large language models (LLM) and vision-language models (VLM) [18]–[20] have proven effective in addressing these challenges. These models support a wide range of applications, including synthetic data generation, stain normalization, style transfer, tissue synthesis, and rare pathology simulation, thereby enhancing learning-based approaches under data-scarce or domain-shifted conditions. Research on content generation models in pathology has grown rapidly, as illustrated in Fig. 2. After a period of sparse, exploratory studies, publication volume increased steadily, with a sharp rise from 2024 onward. This inflection point marks a broader shift in the field, as researchers increasingly recognize the unique advantages that content generation models offer in addressing long-standing challenges in digital pathology, including limited data availability, high annotation costs, and the demand for high-fidelity, consistent synthetic images.

Despite the rapid proliferation of content generation approaches in pathology, a focused review that systematically organizes methods and assesses their impact is still lacking. Prior surveys in computer vision often examine content generation models from broad or algorithmic perspectives [21], [22], leaving pathology-specific challenges underexplored. Notably, Chanda et al. (2024) [23] present a broad catalog of foundation and vision-language models,

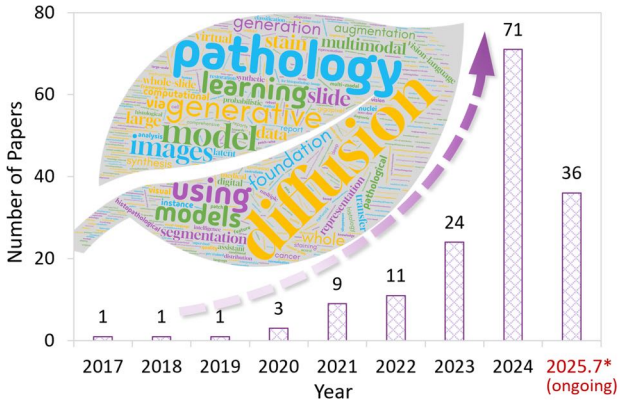

Fig. 2. Number of published papers on generative models in pathology from 2017 to July 2025. The red label "2025.7* (ongoing)" marks ongoing work up to July, while the purple curve shows the exponential growth trend. The accompanying word cloud highlights the most frequent terms, extracted from the titles of the collected articles.

where generative tasks are listed as one among many applications, whereas Bilal et al. (2025) [24] concentrate on evaluation frameworks for foundation models rather than their pathology-specific use. In contrast, our review is centered on content generation models as a distinct methodological paradigm. Given the unique demands of pathology, such as cell-level structural fidelity, high-resolution WSIs, and stain consistency, there is a pressing need to consolidate current developments under a pathology-centric lens. With the advent of foundation models and cross-modal generation strategies, the content generation paradigm in pathology is undergoing a significant shift, warranting timely reflection and synthesis.

In this survey, we present a comprehensive and systematic overview of content generation models in computational pathology. We analyze over 150 peer-reviewed papers encompassing diverse generation targets, with the data sources and search strategy detailed in Supplementary Materials Section II. To facilitate a structured understanding of this rapidly evolving field, we propose a multi-dimensional taxonomy that organizes prior work by generation targets, algorithmic approaches, application tasks, and datasets. This taxonomy delineates four major categories: image generation, text generation, molecular profile-morphology generation, and other specialized tasks. Within these categories, representative applications include synthetic image augmentation, mask- or text-guided synthesis, stain style transfer, image captioning and report generation, virtual molecular profiling, and spatial or semantic layout generation, among others. The scope of this review is limited to models with generative capacity, defined by the presence of a decoder that can synthesize new images or text. In contrast, encoder-only or contrastive foundation models, such as CLIP [25], ALIGN [26], or DINOv2 [27], are considered non-generative, as they learn aligned representations or similarity scores but do not create new content. We further review representative modeling paradigms, including GANs, VAEs, diffusion models, and large language and vision–language models. Their historical development is presented in chronological order, as shown in Fig. 4. In addition, we compile information on widely used datasets, providing a practical resource for researchers entering the field. Finally, the Discussion synthesizes demonstrated capabilities and impact, analyzes critical limitations and risks, and outlines future directions toward trustworthy clinical adoption.

Our contributions are as follows:

- We provide the first comprehensive survey of content generation models in computational pathology, covering more than 150 publications up to July 2025.

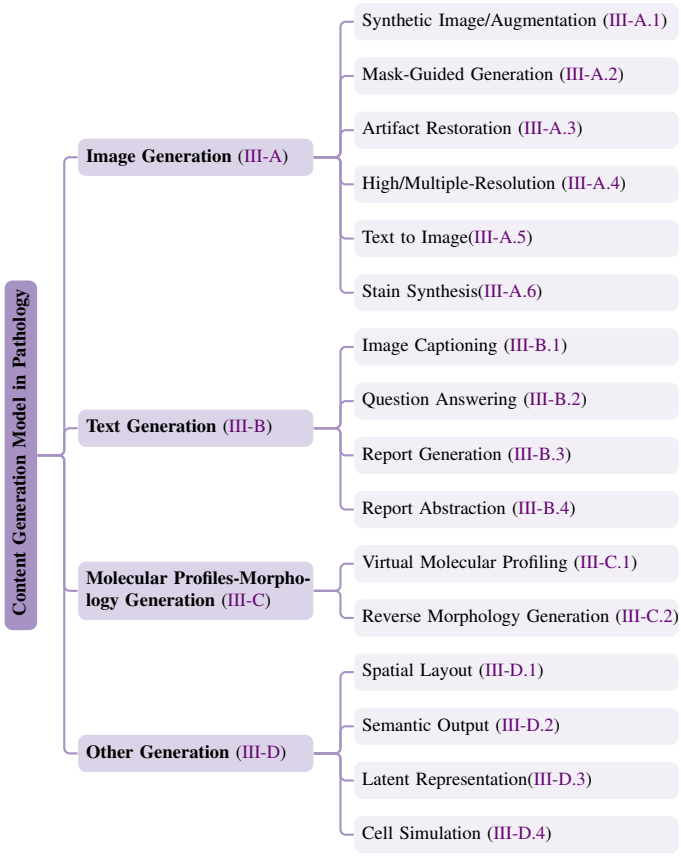

Fig. 3. Taxonomy of content generation models in computational pathology.

- We propose a unified taxonomy that organizes existing studies by generation targets and modeling approaches, offering a structured view of methodological advances in pathology.
- We curate commonly used datasets across major generation tasks and critically discuss open challenges and future directions.

The remainder of this survey is organized as follows. Section II reviews representative generative models in a historical context. Section III categorizes applications across image, text, molecular profile-morphology, and other emerging generation tasks. Section IV summarizes commonly used datasets. Section V discusses capabilities, limitations, and pathways toward clinical deployment.

II. Generation Model

Content generation models in computational pathology have undergone a rapid evolution over the past decade, marked by successive paradigm shifts driven by architectural breakthroughs. This trajectory is summarized in the timeline shown in Figure 4 and further illustrated by representative model families in Figure 5. The field's foundations were established around 2014 by **Variational Autoencoders** and **Generative Adversarial Networks**, which pioneered the synthesis of images from learned latent spaces. A subsequent shift occurred around 2020 with the advent of **Diffusion Models**, enabling a new level of high-fidelity image synthesis through an iterative denoising process. More recently, since 2022, the rise of **Large Language and Vision-Language Models** has reoriented the field from purely pixel-level tasks towards complex, multimodal reasoning and text generation. Beyond these three mainstream paradigms, a set of other distinct approaches, including novel hybrid architectures and advanced training strategies, is also being explored. In the

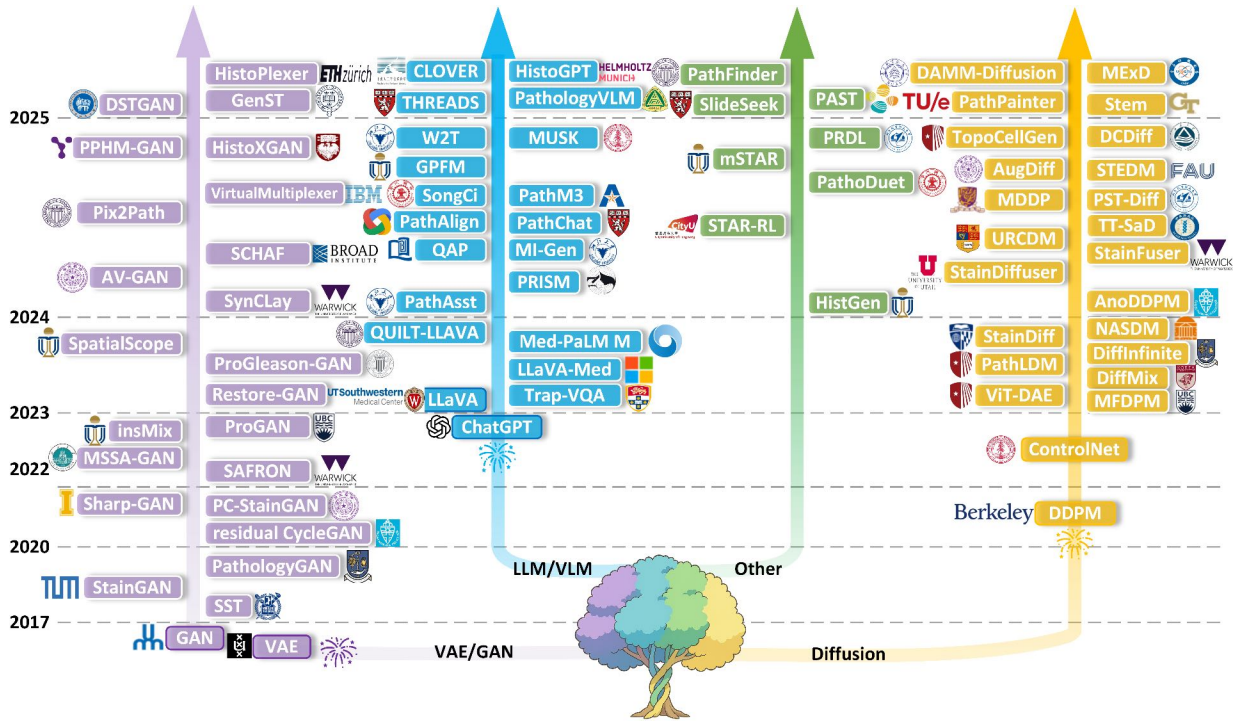

Fig. 4. Timeline of key developments in pathological generation (2017-2025). This timeline illustrates the major milestones and developmental trajectory of generative models across methods in pathology.

following sections, we describe these model families, focusing on their algorithmic innovations and key applications within pathology.

A. VAE and GAN-based Models

Variational Autoencoders [28] introduced the idea of encoding data into a latent distribution and reconstructing it through probabilistic decoding. While VAEs typically generate blurrier images compared to GANs or diffusion models, their latent structure is well-suited to representation learning [14] and conditional data synthesis [29]. In pathology, VAEs have been used for self-supervised pretraining and feature-space augmentation for classification tasks [30].

Generative Adversarial Networks [31], proposed in 2014, introduced adversarial training between a generator and a discriminator to produce realistic samples. Conditional variants such as cGAN [32], Pix2Pix [33], and CycleGAN [34] advanced GANs toward controlled, paired, and unpaired image-to-image translation. These architectures have been instrumental in stain normalization [35], [36], virtual staining [37], [38], and cross-domain adaptation [39] in histopathology. Later models such as StyleGAN [40] added style-based modulation to the latent space, enabling fine-grained control of image attributes. In computational pathology, StyleGAN and its variants have supported mitotic figure synthesis [41], rare-cell augmentation [42], and attribute-conditioned generation [43]. The discriminator components in GANs have also been repurposed for unsupervised learning of biologically meaningful features [44], making GANs a versatile tool for both generation and representation learning.

B. Diffusion-based Models

Diffusion models generate images by learning to reverse a stochastic noising process. This class of models, beginning with the Denoising Diffusion Probabilistic Model (DDPM) [45] in 2020, has demonstrated outstanding performance in producing visually realistic and diverse samples with strong semantic fidelity. Successors such as DDIM [46] and Score-based Generative Models [47] further accelerated sampling and extended generation to continuous domains.

Compared to GANs, diffusion models are more stable during training and less prone to mode collapse, making them increasingly favored in high-resolution pathology synthesis. In computational pathology, diffusion models have been applied to gigapixel-scale WSI generation [48], rare-class augmentation [49], and denoising or artifact removal [50], [51]. Recent models have further enabled synthesis conditioned on disease labels [52], segmentation masks [53], and genomic signals [54]. To enhance controllability, methods such as Classifier-Free Guidance (CFG, 2022) [55], Stable Diffusion [56], and ControlNet [57] were introduced. These allow models to accept user-defined constraints in the form of class labels, text prompts, edge maps, or spatial masks. In pathology, such control mechanisms have been used to synthesize mitosis stages [41], generate tumor tiles guided by transcriptomic profiles [54], and support counterfactual image generation for model interpretability [58].

C. LLM and VLM-based Models

Large language models and vision-language models represent a paradigm shift from pixel-level synthesis to knowledge-based generation. These models leverage large-scale autoregressive Transformers as their core engine, enabling complex multimodal reasoning and the synthesis of new, structured text sequences, such as diagnostic reports and answers to clinical queries. A generative vision–language model contains a decoding component capable of synthesizing novel outputs in at least one modality (vision or language) [59], [60]. For image-conditioned text generation, this involves a language decoder that generates sequences via autoregressive token prediction or masked-token infilling. Conversely, text-to-image generation relies on an image decoder, such as a diffusion model, to synthesize images from a textual prompt. It is crucial to distinguish these generative VLMs from a parallel lineage of non-generative, contrastive learning models [61]. While the latter have been pivotal in learning powerful

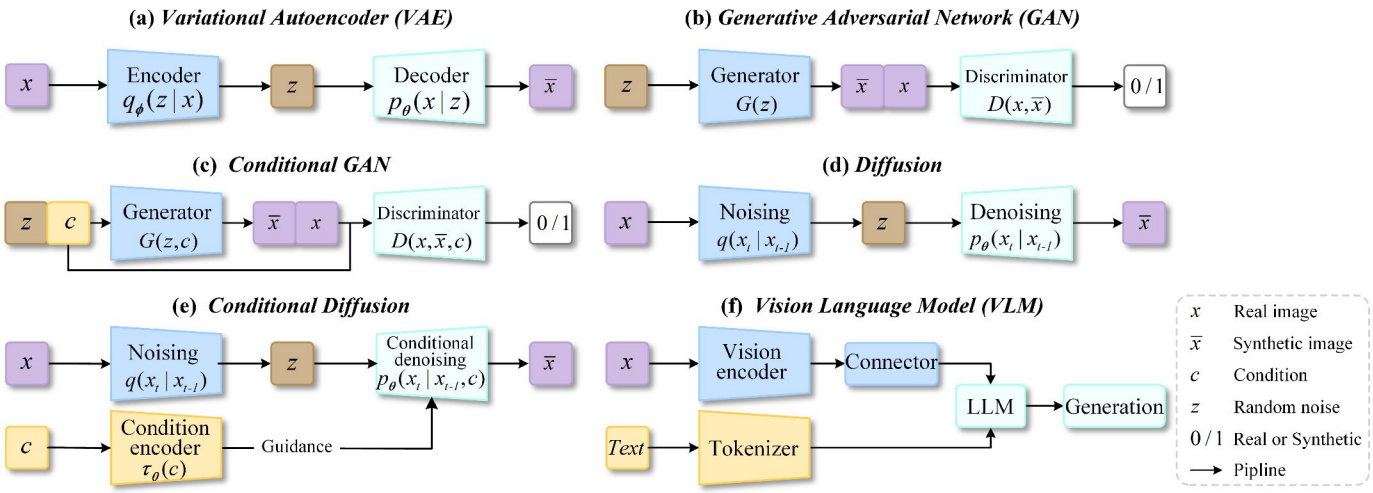

Fig. 5. Schematics of representative generative model architectures. (a) VAE learns a probabilistic latent space and reconstructs images through encoder-decoder optimization. (b) GAN consists of a generator and a discriminator trained in a minimax game to produce realistic outputs. (c) Conditional GAN introduces external conditioning to guide the generation process. (d) The diffusion model performs iterative denoising to generate data by learning the reverse of a noise injection process. (e) Conditional Diffusion extends the diffusion framework with conditioning inputs to steer the generative trajectory during denoising. (f) VLM fuses image and text inputs through an LLM for generation, supporting diverse multimodal tasks.

multimodal representations [62], they lack the decoder architecture necessary for synthesizing new content and thus fall outside the scope of this part.

The evolution of generative VLMs was driven by breakthroughs in general-domain models, beginning with the release of ChatGPT in 2022 and the open-source development of multimodal architectures such as LLaVA [63] in 2023. TraP-VQA [64] introduced transformer-based vision–language modeling to pathology, while Med-PaLM M [65] extended this paradigm across biomedical data. Domain-adapted models like LLaVA-Med [66] and Quilt-LLaVA advanced specialization for biomedical and pathology-specific VQA. These foundations enabled dedicated architectures such as HistoGPT [67], PathCHAT [18], and PathologyVLM [68], which integrated instruction-tuning and dialogue for report generation and interactive VQA. More recently, LLMs have been used as reasoning engines to guide generative processes, exemplified by TopoFM [69], which employs topological estimators to synthesize realistic cellular arrangements. This progression underscores the growing role of LLMs as central reasoning components in multi-stage generative pipelines.

D. Other Generative Approaches

Beyond the mainstream paradigms, several distinct methodological approaches are emerging. Novel hybrid architectures are being explored, exemplified by models like mSTAR [70] and PathoDuet [71], which employ advanced, multi-stage pre-training strategies to learn powerful, domain-aware representations before the final generative task. Reinforcement learning is also being employed not as a generative architecture itself, but as a powerful fine-tuning strategy to enhance factual accuracy and controllability by optimizing against domain-specific rewards [72]. Looking forward to 2025, the concept of AI Agents marks a leap, positioning optimized generative models as reasoning engines within autonomous systems such as SlideSeek [73] and PathFinder [74], capable of multi-step planning and tool use. These approaches collectively represent a systematic transition toward more modular, controllable, and intelligent generative systems.

III. Generation Task

A. Image Generation

Image generation plays a foundational role in computational pathology by enabling data augmentation, enhancing visual realism, and preserving structural fidelity, thereby supporting a wide range of downstream tasks (Fig. 6). Recent generative models have significantly advanced synthetic pathology imaging, addressing challenges such as limited annotations, class imbalance, and domain shifts. We categorize existing efforts into six sub-tasks based on their generation goals: (1) synthetic image generation for data augmentation (conditional and unconditional); (2) mask-guided generation for structural fidelity; (3) artifact restoration for removing artifacts and improving image quality; (4) high/multiple-resolution generation, including cross-scale images and gigapixel whole-slide generation; (5) text to image generation from diagnostic descriptions or prompts; and (6) stain style generation for normalization and transfer. Each category reflects distinct generative objectives and involves diverse conditioning mechanisms tailored to clinical needs, as illustrated in Tables I and II..

1) Synthetic Image and Augmentation:

a) Conditional Image Generation: Conditional generative models have become the mainstream approach in pathological image synthesis. Their core advantage lies in enabling controllable synthesis guided by semantic, spatial, or biological attributes, thereby providing targeted support for downstream tasks such as segmentation and classification. Early works were predominantly based on GANs, which enabled control over both high-level pathological features like Gleason grade [42], [75]–[77] and the fine-grained, realistic composition of individual nuclei for data augmentation [78]. Recently, however, the field has witnessed a paradigm shift toward diffusion models, which demonstrate significant advantages in generation quality, training stability, and flexibility of conditioning.

The conditioning mechanisms for diffusion models are evolving towards greater diversity and granularity. A significant body of work focuses on leveraging spatial and semantic priors. These methods aim to enhance the structural accuracy of generated images by introducing various forms of guidance, such as semantic features extracted by Transformers [79], cell nuclei structure maps from two-stage pipelines [80], custom instance segmentation maps [49], or masks provided by unsupervised models like SAM [81]. While these

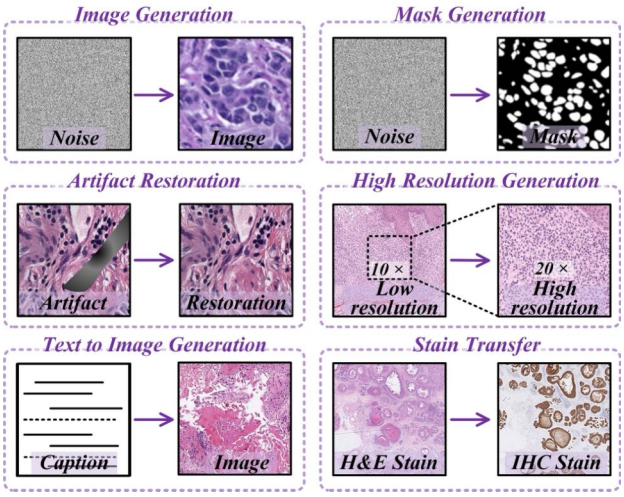

Fig. 6. Applications of image generation in computational pathology. Illustrative examples of key tasks, including mask-guided synthesis, resolution enhancement, text-to-image translation, and stain transfer.

approaches differ in the source and granularity of their conditional information, they share the common goal of enhancing structural plausibility and image-label alignment, particularly when addressing class imbalance or complex tissue architectures. To bypass costly fine-grained annotations, generation can be conditioned on self-supervised semantic representations, such as unsupervised histological prototypes, which serve as proxies for manual labels [82], [83].

More advanced research has begun to integrate biological knowledge and weak supervision signals, aiming to improve the clinical interpretability and utility of the models. Multiple studies are moving beyond mere visual realism to simulate authentic biological processes, for instance, by controlling diffusion depth to model the gradual progression of a disease [84], or using probabilistic mitotic scores as a condition to capture the morphological continuum of cell changes [41]. Furthermore, some research fine-tunes loss weights to emphasize subtle features associated with specific molecular states (e.g., HPV) [85], while others utilize frameworks like ControlNet with binary images and text prompts to achieve effective control without dense annotations [86], [87]. This trend signifies a shift in synthetic image generation from pursuing visually appealing results to creating functionally useful data that reflects underlying biological mechanisms or supports clinical decision-making. Moreover, some work sets the objective of conditional generation directly on functional applications, such as generating counterfactual visual explanations to investigate the impact of biomarkers [58] or performing anomaly detection via reconstruction error [88], further expanding the application boundaries of generative models.

b) Unconditional Image Generation: Unconditional generative models remain indispensable, particularly in unsupervised learning and data-scarce scenarios. Traditional GAN-based methods have played a significant role in learning interpretable latent spaces for phenotype manipulation [89], self-supervised field-of-view expansion [14], and image restoration [90]. Notably, some models cleverly bridge the gap between purely unconditional and strongly conditional generation by weakly embedding conditional signals (such as Gleason grades) directly into the generator [42]. More recently, unconditional diffusion models have been increasingly employed to address specific data challenges. For example, studies have used them to generate high-quality samples for augmenting training sets, enhancing realism through techniques like color normalization [91]; others have focused on generating data for rare modalities like deep ultraviolet (DUV) fluorescence [92]; and some have mitigated domain shift, for instance by improving segmentation robustness through training-time stain augmentation on synthetically diversified stain styles [34], [93]. In summary, pathological image synthesis is shifting beyond the pursuit of visual fidelity toward task-oriented and knowledge-driven generation, where models are designed not only to produce realistic images but to serve specific analytical or clinical goals. Recent work has demonstrated that synthetic images generated by diffusion models can successfully train downstream classifiers to a performance level comparable to those trained on real data [94]. The choice between GANs and diffusion models, and between conditional and unconditional strategies, is now shaped by the demands of downstream tasks, the level of annotation available, and the need for controllability and interpretability in real-world applications.

2) Mask-Guided Generation: Mask-guided generation represents a key strategy for enforcing structural awareness in synthetic pathology images, thereby supporting critical applications like data augmentation, segmentation model training, and artifact correction. The primary technical approaches can be categorized by their generation target: direct mask-to-image synthesis and the joint generation of image-mask pairs.

a) Mask-to-Image Generation: A major line of work focuses on synthesizing realistic pathology images directly conditioned on segmentation masks. Diffusion models are prevalent in this area due to their high fidelity. Initial methods employed conditional diffusion models to generate images from semantic masks, effectively addressing class imbalance by enriching underrepresented structures [97]. More sophisticated approaches have since refined this process by enhancing the conditioning mechanism. For instance, techniques include using self-distillation from separated conditions for more precise multi-grade synthesis [52], focusing on fine-grained nuclei-level semantics [53], or adopting multi-resolution hierarchical features to better preserve the integrity of lesion structures during coarse-to-fine generation [100]. In parallel, GAN-based methods offer an alternative pathway. Sharp-GAN, for example, replaces binary masks with normalized nucleus distance maps and introduces a sharpness loss, a design choice that specifically improves boundary resolution for challenging cases like overlapping nuclei [96].

b) Joint Image-Mask Generation: Another research direction aims to simultaneously generate images and their corresponding masks, producing fully paired data ideal for training segmentation models. A key strategy within this domain is the disentanglement of content (the structural mask) from appearance (the visual style). Diffusion models have achieved this by conditioning on both predefined semantic layouts and extracted visual styles [98], or by using cross-image attention to transfer appearance while preserving morphological structures from a reference image [101]. PathoPainter leverages a similar principle for mask-guided inpainting, ensuring strong alignment between the generated image content and the guiding mask [103]. To address scalability, hierarchical frameworks like DiffInfinite have been proposed, which first generate a scalable coarse segmentation mask and subsequently synthesize a high-fidelity image via patch-based diffusion conditioned on it [15]. This architecture enables artifact-free image assembly at arbitrary scales and supports privacy-preserving data augmentation. In contrast, GAN-based approaches provide computationally efficient alternatives with fine-grained control. Methods like Pathopix-GAN are designed for robust, large-scale data augmentation using semantically adaptive normalization [102], while AttributeGAN demonstrates exceptional control by conditioning generation on specific cellular attributes such as cell crowding and nuclear pleomorphism, enhancing biological interpretability [95]. More recent diffusion models further advance this controllability, enabling the co-synthesis of images and instance

TABLE I
OVERVIEW OF IMAGE GENERATION METHODS IN COMPUTATIONAL PATHOLOGY (PART 1).

Year	Method	Input	Arch.	Key Application
1. Synthetic Image and Augmentation				
2019	PathologyGAN [89]	Patch	GAN	Deep representations of cancer tissue for classification and generation.
2020	Xue et al. [75]	Patch	GAN	Dataset augmentation with selective synthetic images for classification.
2022	Porter et al. [85]	Patch, HPV Status	Diffusion	Generating images conditioned on HPV status for improved classification.
	Bouteldja et al. [34]	Patch	GAN	CycleGAN-based augmentation for stain-invariant segmentation.
	insMix [78]	Patch	GAN	Generating realistic data for augmenting nuclei instance segmentation tasks.
	MFDPM [91]	Patch	Diffusion	Synthesizing histopathology images preserving morphological details.
2023	ViT-DAE [79]	Patch	Diffusion	Transformer-driven diffusion autoencoder for augmentation and classification.
	Sun et al. [87]	Patch	Diffusion	Instance-aware diffusion model for gland segmentation.
	ProGleason-GAN [42]	Patch	GAN	Synthesizing prostate cancer images with controllable Gleason grades.
	DiffMix [49]	Patch	Diffusion	Synthesizing balanced image-label pairs to improve segmentation.
	Bahadir et al. [41]	Patch	Diffusion	Detecting and characterizing mitotic figures for improved classification.
2024	AnoDDPM [88]	Patch	Diffusion	Detecting out-of-distribution samples in digital pathology images.
	Ktena et al. [93]	Patch	Diffusion	Improving medical classifier fairness via generation for detection.
	Graikos et al. [82]	Patch	Diffusion	Representation-guided diffusion models for large-image generation.
	ADD [84]	Patch	Diffusion	Simulating progressive pathological transitions for classification tasks.
	MoPaDi [58]	Patch	Diffusion	Generating counterfactual images for mechanistic explanation of AI models.
	Li et al. [76]	Patch	GAN	Unified framework for histopathology image augmentation and classification.
	Yu et al. [80]	Patch	Diffusion	Diffusion-based data augmentation for nuclei image segmentation.
	USegMix [81]	Patch	Diffusion	Unsupervised segment mix for data augmentation in classification.
	Pozzi et al. [94]	Patch	Diffusion	Evaluating synthetic data quality for downstream classification tasks.
2025	PDSeg [86]	Patch, Mask	Diffusion	Patch-wise mask distillation for weakly-supervised tissue segmentation.
	dcGAN [77]	Patch	GAN	Data augmentation with synthetic Gleason patterns to improve Gleason grading.
2. Mask-Guided Generation				
2021	AttributeGAN [95]	Cellular Attribute	GAN	Synthesizing histopathology images with control over cellular attributes.
	Sharp-GAN [96]	Patch	GAN	Enhancing boundary resolution in mask-to-image synthesis for segmentation.
2023	Cechnicka et al. [97]	Patch, Mask	Diffusion	Realistic data enrichment from masks for robust segmentation.
	NASDM [53]	Patch, Mask	Diffusion	Generating nuclei-aware semantic tissue from segmentation masks.
2024	DISC [52]	Patch	Diffusion	Self-distillation for prostate cancer grading and classification.
	Min et al. [43]	Text, Point Maps	Diffusion	Co-synthesis of histopathology images and masks from text and point maps.
	DiffInfinite [15]	Patch	Diffusion	Large-scale mask-to-image synthesis using parallel random patch diffusion.
	STEDM [98]	Patch, Layout	Diffusion	Zero-shot, style-based synthesis of image-mask pairs from semantic layouts.
	SynCLay [99]	Layout	GAN	Interactive synthesis of images from cellular layouts with co-generated masks.
	HADiff [100]	Patch	Diffusion	Hierarchy-aggregated diffusion for mask-based segmentation.
2025	Winter et al. [101]	Patch, Mask	Diffusion	Zero-shot, in-silico histopathologic image generation guided by masks.
	Jehanzaib et al. [102]	Patch, Mask	GAN	Robust segmentation and data augmentation from masks.
	PathoPainter [103]	Patch	Diffusion	Tumor-aware image inpainting for segmentation data augmentation.

labels from intuitive inputs like point maps and text prompts [43], and even user-specified cellular layouts, as demonstrated by frameworks like SynCLay [99]. In summary, mask-guided synthesis enables structural control in pathology image generation. The choice between direct mask-to-image synthesis, joint image–mask generation, and model type (diffusion or GAN) depends on whether the goal is to augment annotations with realistic textures or to create diverse, fully annotated datasets.

3) Artifact Restoration: Histological artifacts are structural or stylistic distortions introduced during tissue preparation and digitization, such as folds, tears, or staining inconsistencies [138]. These artifacts can obscure critical features and impair both human and machine interpretation. Generative models address this issue through two main strategies. (1) Localized restoration focuses on correcting only affected regions while preserving surrounding tissue. GANs have been applied to inpaint occlusions like marker ink [104] and reconstruct broader artifact regions with stain and structure preservation [9], with refinements such as multi-scale self-attention enhancing the modeling of contextual dependencies [90]. Diffusion models adopt masked denoising [16], further enhanced by Transformer-based context modeling [51] and latent-space optimization [107], and progressing towards fully unsupervised pipelines like HARP [106] that automate artifact localization prior to restoration. (2) Holistic enhancement targets global degradations such as blur or inter-site stain variation. Models like Restore-GAN improve overall image quality [105], and federated GANs address distributed inconsistencies [39]. Recent advances integrate artifact correction and synthesis into unified frameworks that support both preprocessing and data augmentation, marking a shift toward end-to-end artifact management in computational pathology [108].

4) High/Multiple-Resolution Generation: The synthesis of diagnostically viable pathology images necessitates reconciling microscopic cellular fidelity with macroscopic anatomical coherence. This fundamental challenge is addressed through two complementary research thrusts: high-resolution generation, which focuses on creating large-scale images with fine spatial detail, and multi-resolution generation, which aims to ensure structural and semantic consistency across different scales of observation.

a) High-Resolution Generation: Research in high-resolution synthesis focuses on generating large-scale images with fine spatial detail at a single, high magnification. The predominant strategy to manage the immense pixel space is through progressive or cascaded generation. Early work applied progressive GANs to synthesize 1024×1024 patches [109] and VAEs to improve classification with

TABLE II
OVERVIEW OF IMAGE GENERATION METHODS IN COMPUTATIONAL PATHOLOGY (PART 2).

Year	Method	Input	Arch.	Key Application
3. Artifact Restoration				
2020	Venkatesh et al. [104]	Patch	GAN	Restoring marker-occluded regions in histopathology images.
2022	MSSA-GAN [90]	Patch	GAN	Pathology image restoration using multi-scale self-attention.
2023	ArtiFusion [16]	Patch	Diffusion	Restoring artifacts in histopathology images to improve classification.
	AR-CycleGAN [9]	Patch	GAN	Detecting and restoring various histological artifacts.
	Shen et al. [39]	Patch	GAN	Federated learning system for privacy-preserving stain normalization.
	Restore-GAN [105]	Patch	GAN	Enhancing pathology image quality (deblurring, super-res) for downstream tasks.
2024	Wang et al. [51]	Patch	Diffusion	A lightweight diffusion model for the selective inpainting of histological artifacts.
	HARP [106]	Patch	Diffusion	Artifact restoration through unsupervised detection and diffusion-based inpainting.
	LatentArtiFusion [107]	Patch	VAE/Diff.	Restoring histological artifacts efficiently in a low-dimensional latent space.
2025	ArtiDiffuser [108]	Patch	Diffusion	Unified framework for histological artifact restoration and synthesis.
4. High/Multi-Resolution Generation				
2021	ProGAN [109]	Patch, WSI	GAN	High-quality patch/WSI generation for data augmentation and classification.
	Lahiani et al. [110]	Patch	GAN	Synthesizing seamless WSI using perceptual consistency for segmentation.
	Boyd et al. [14]	Patch	VAE	Generating high-resolution images to improve classification performance.
2022	SAFRON [111]	Mask	GAN	Generating seamless, large-scale histology images via patch-stitching GAN.
2023	Harb et al. [48]	WSI	Diffusion	Generating gigapixel-scale whole-slide images from scratch.
2024	STAR-RL [72]	Patch	RL	Hierarchical reinforcement learning for interpretable super-resolution.
	PathUp [112]	Patch, Text	Diffusion	Synthesizing large, multi-class pathology images with high fidelity.
	URCDM [113]	Patch, WSI	Diffusion	Ultra-resolution WSI synthesis using a cascaded conditional diffusion model.
	Thakkar et al. [114]	Patch	Diffusion	Comparative analysis of diffusion models for synthetic data generation.
	Histo-Diffusion [50]	Patch	Diffusion	Super-resolution with quality assessment for classification.
2025	ToPoFM [69]	Patch	LLM/VLM	Generating images with cellular topology control for segmentation.
5. Text to Image				
2023	PathLDM [115]	Patch, Text	Diffusion	Generating histopathology images conditioned on descriptive text prompts.
2024	VIMs [116]	Patch, Text	Diffusion	Synthesizing virtual multiplex IHC stains from H&E images using text prompts.
6. Stain Synthesis				
2017	SST [117]	Patch	GAN	Stain normalization and style transfer to improve classification robustness.
2018	Zanjani et al. [35]	Patch	GAN	Stain normalization using generative adversarial networks.
2021	PC-StainGAN [38]	Patch	GAN	Unpaired stain transfer (e.g., H&E to IHC) with pathology-consistency constraints.
	Residual CycleGAN [118]	Patch	GAN	Improving domain transformation robustness for segmentation tasks.
	Runz et al. [119]	Patch	GAN	Normalization of H&E stained images for improved classification.
2022	CAGAN [120]	Patch	GAN	Semi-supervised stain normalization via dual-decoder consistency learning.
2023	StainDiff [17]	Patch	Diffusion	Transferring stain styles between histology images using diffusion models.
	Jeong et al. [121]	Patch	Diffusion	Stain normalization using a score-based diffusion model for classification.
2024	AV-GAN [122]	Patch	GAN	Virtual staining for unevenly stained tissue using an attention-based varifocal GAN.
	StainFuser [123]	Patch	Diffusion	Controlling diffusion for fast neural style transfer in multi-gigapixel images.
	StainDiffuser [124]	Patch	Diffusion	Multitask virtual staining and segmentation from the same model.
	PPHM-GAN [37]	Patch	GAN	High-resolution any-to-any stain translation for classification.
	MDDP [125]	Patch	Diffusion	Pre-training via H&E-to-IHC translation for improved WSI classification.
	VirtualMultiplexer [126]	Patch	Other	AI-based virtual multiplexing for tumor profiling, prognosis, and classification.
	TT-SaD [127]	Patch	Diffusion	Improving classification robustness through test-time stain adaptation.
	PST-Diff [128]	Patch	Diffusion	Achieving high-consistency stain transfer with pathological/structural constraints.
	F2FLD [129]	Patch	Diffusion	Unpaired frozen-FFPE translation for classification.
	ULSA [130]	Patch, WSI	GAN	Unsupervised latent stain adaptation for WSI segmentation.
	Diffusion-L [131]	Patch	Diffusion	Image-to-image translation with robust uncertainty quantification.
	DUST [132]	Patch	Diffusion	A unified diffusion framework for versatile (any-to-any) stain transfer.
	ODA-GAN [10]	Patch	GAN	IHC staining via orthogonal decoupling of morphological and staining features.
	Xiong et al. [133]	Patch	Diffusion	Unpaired multi-domain stain transfer using dual path prompting for classification.
2025	CC-WSI-Net [134]	Patch	GAN	Generating seamless virtual IHC whole-slide images with content consistency.
	VM-GAN [135]	Patch	GAN	A value mapping framework for large-scale histological image-to-image translation.
	DSTGAN [136]	Patch	GAN	Deeply supervised two-stage GAN for stain normalization and segmentation.
	RBDM [137]	Patch	Diffusion	Virtual staining from label-free polarimetric imaging for report generation.

high-resolution outputs [14], while parallel efforts focused on ensuring seamless WSI reconstruction by introducing novel losses to eliminate tiling artifacts [110]. This concept has been significantly advanced by diffusion models, which employ coarse-to-fine or cascaded frameworks to sequentially add detail, enabling the synthesis of whole-slide images (WSIs) at the gigapixel scale [48], [113]. A parallel effort integrates structural priors, as in ToPoFM, which embeds topology-informed priors into a latent diffusion model to generate high-resolution images with realistic cellular organization, representing a fusion of scalability and structural control [69].

b) Multi-Resolution Generation: Multi-resolution image generation in pathology addresses the challenge of ensuring semantic and structural coherence across diagnostic magnifications. Several technical strategies have been explored to achieve this goal. One approach involves explicit alignment and enhancement frameworks. Histo-Diffusion employs a dual-stage pipeline that first restores histological priors and then applies controllable super-resolution, while PathUp incorporates a dedicated "patho-align" module to embed expert priors and mitigate tiling artifacts [50], [112]. SAFRON [111] further integrates the stitching process into adversarial training, compelling the generator to produce patches that are seamless and globally consistent. A distinct line of work reformulates super-resolution as a reinforcement learning problem. STAR-RL employs a hierarchical agent that adaptively locates low-quality regions and determines termination criteria, enabling efficient and interpretable region-aware reconstruction [72]. Another approach emphasizes scale-aware representation learning. Class-conditioned diffusion models guided by patch size and label prompts explicitly capture relationships between spatial scale and visual features, thereby improving the robustness of downstream classifiers to resolution shifts [114]. Collectively, these strategies highlight complementary pathways toward generating diagnostically consistent images across magnifications. In summary, high-resolution and multi-resolution generation represent complementary dimensions of pathology image synthesis: the former constructs detail-rich gigapixel images, while the latter ensures coherence across diagnostic scales. Clinical applications will require integrated frameworks that preserve both fine detail and cross-scale consistency.

5) Text to Image: Text-to-image generation in pathology aims to bridge linguistic clinical descriptions and visual tissue synthesis, enabling controllable and semantically rich image generation. PathLDM [115] introduces a text-conditioned latent diffusion model tailored for histopathology, leveraging GPT-3.5 to summarize diagnostic reports and integrate contextual textual cues into the image generation process. In parallel, VIMs [116] develops a text-to-stain diffusion framework that synthesizes IHC markers from H&E-stained (Hematoxylin and Eosin) images using textual prompts, requiring only uniplex IHC training data. Consequently, language-driven generative models hold significant promise for interpretability, modality translation, and report-informed synthesis in digital pathology.

6) Stain Synthesis: Generative models help overcome the substantial visual variability in histological images, a key obstacle to developing generalizable AI [36]. Stain synthesis comprises two tasks: normalization, which standardizes appearances within a stain type, and transfer, which converts one stain modality into another.

a) Stain Normalization: The initial wave of data-driven stain normalization was led by Generative Adversarial Networks (GANs), with unsupervised frameworks like Stain-Style Transfer (SST) network [117] and CycleGAN [119] emerging as foundational tools for mapping H&E images between domains. However, a critical drawback of these early models was the introduction of structural artifacts. This limitation spurred a series of architectural and methodological refinements focused on structure preservation. For instance, Residual CycleGAN introduced a global skip connection to force the model to learn only the residual color transformation, thereby better preserving morphological content [118]; and frameworks like CAGAN [120] introduced semi-supervised training with dual-decoder consistency to leverage unlabeled source domain data for more robust normalization. Extending this, DSTGAN integrated a Swin Transformer backbone to better capture long-range contextual information [136], while other work used InfoGAN-inspired frameworks to explicitly disentangle structure from color attributes [35]. Beyond image fidelity, research has expanded to address challenges like data privacy, leading to federated cGANs for privacy-preserving normalization across institutions [39]. More recently, diffusion models have emerged as a powerful alternative, with score-based methods offering stable normalization [121], controlling diffusion for fast, gigapixel-scale style transfer [123], and innovations like TT-SaD enabling robust adaptation without model retraining [127], while semi-supervised frameworks like ULSA train models to be inherently stain-invariant by enforcing latent feature consistency on unlabeled data [130].

b) Stain Transfer: Representing a more ambitious goal, stain transfer aims to computationally predict one staining modality from another (e.g., H&E to IHC). Early GAN-based efforts were frequently impaired by tiling artifacts and structural inconsistencies. To combat this, researchers have introduced pathology-specific consistency constraints [38]. Advanced models such as ODA-GAN [10] enforce consistency at the feature level through orthogonal decoupling of staining and morphology, combined with multi-scale biological priors to ensure the clinical utility of virtual IHC panels [126]. Other approaches employ perceptual losses for improved semantic mapping [110], tiling strategies for whole-slide coherence [134], [135], and unified frameworks for multi-domain transformations [37]. Attention-based architectures, such as AV-GAN [122], further address structural inconsistencies in complex regions. A definitive paradigm shift is now underway, with diffusion models becoming the dominant technology due to their superior generation fidelity. This has enabled high-fidelity translation from challenging modalities like autofluorescence, label-free polarimetric or artifact-laden frozen sections into virtual stains or their gold-standard counterparts [129], [131], [137]. A key focus in this new paradigm is ensuring structural consistency, leading to training-free frameworks that leverage DDIM inversion [133] or enforce pathological constraints [128]. To tackle this from a training perspective, StainDiff [17] has adapted the cycle-consistency principle, enforcing structural preservation in unpaired stain transfer by applying constraints within the diffusion model's latent space. The flexibility of diffusion models has also facilitated unified, multi-task architectures capable of any-to-any stain conversion or simultaneous virtual staining and segmentation [37], [124], [132]. In a particularly innovative application, virtual staining is now being leveraged as a powerful pretext task for representation learning, where the synthesis objective itself helps a model learn richer semantic features [125]. In summary, generative stain synthesis has developed along two complementary paths: normalization, which preserves structural fidelity, and transfer, increasingly enabled by diffusion models with superior realism. Yet clinical deployment depends not only on technical performance; normalization may conceal artifacts and transfer can introduce spurious features, underscoring the necessity of rigorous expert validation.

B. Text Generation

Generative models are fundamentally transforming the interface between visual pathology and clinical language. The application of large language and vision-language models in this domain has evolved along a trajectory of increasing clinical utility. This progression begins with foundational tasks of translating visual features

TABLE III
OVERVIEW OF TEXT GENERATION METHODS IN COMPUTATIONAL PATHOLOGY.

Year	Method	Input	Arch.	Key Application
1. Image Captioning				
2021	Gamper et al. [139]	Patch	VLM	Generating descriptive captions for pathology patches using a MIL framework.
2023	SGMT [140]	Patch, WSI	VLM	Generating WSI reports guided by cancer information using a masked transformer.
2024	HistGen [141]	WSI	VLM	Employing a local-global hierarchical encoder for comprehensive WSI report generation.
	Ferber et al. [142]	Patch	VLM	In-context learning of large models for pathology image captioning and reporting.
2. Visual Question Answering (VQA)				
2023	TraP-VQA [64]	Patch, Text	VLM	Achieving interpretable VQA by a vision encoder with multimodal attention mechanisms.
	LLaVA-Med [66]	Patch, Text	VLM	Building a biomedical VQA assistant through cost-efficient instruction tuning.
2024	PathChat [18]	Patch, Text	VLM	Developing a general-purpose multimodal AI assistant for interactive VQA and reporting.
2025	PathologyVLM [68]	Patch, Text	VLM	A large vision-language model foundational for diverse pathology VQA and reports.
	PathCoT [143]	Patch, Text	VLM	Improving reliability of zero-shot CoT reasoning via expert-guided prompting.
	SlideChat [19]	WSI	VLM	Enabling interactive WSI-level VQA through a hierarchical processing approach.
	CLOVER [144]	Patch, Text	VLM	Cost-effective instruction learning for pathology VQA and object detection tasks.
	MUSK [145]	Patch, Text	VLM	Leveraging large-scale unpaired data via two-stage pretraining for robust VQA.
3. Report Generation				
2023	Sengupta et al. [146]	WSI	VLM	Achieving end-to-end pathology report generation directly from whole-slide images.
	Quilt-LLaVA [20]	Patch, WSI	VLM	Enabling spatially-grounded WSI reasoning for VQA and report generation.
2024	HistoGPT [67]	WSI	VLM	Generating human-level, clinical-grade pathology reports from multiple gigapixel WSIs.
	MI-Gen [147]	WSI	VLM	Utilizing a multiple instance generation model for creating detailed WSI reports.
	PathGen-LLaVA [148]	Patch	VLM	A visual question answering model trained on PathGen-1.6M dataset.
	PathAlign [62]	WSI	VLM	Aligning slide-level visual features with textual reports for improved report generation.
	PathInsight [149]	Patch, Text	VLM	Instruction-tuning a model on curated datasets for enhanced reporting and VQA.
	PRISM [150]	WSI, Text	VLM	A foundation model generating slide-level reports by aggregating tile embeddings.
2025	PolyPath [151]	WSI, Text	VLM	Adapting a large multimodal model to generate cohesive reports from multiple slides.
	Lucassen et al. [152]	WSI	VLM	Multimodal representation learning for report generation of melanocytic lesions.
	mSTAR [70]	WSI	VLM	Generating reports that directly link tissue morphology to its molecular context.
	KR-MLFS [153]	WSI	VLM	Integrating knowledge retrieval with multi-level feature selection for report generation.
	SlideSeek [73]	WSI, Text	VLM	A multi-agent system that navigates WSIs autonomously for diagnostics.
	TCP-LLaVA [154]	WSI, Text	VLM	Enabling efficient WSI-level VQA via a novel token compression module.
	Redekop et al. [83]	Prototype	Diffusion	Synthesizing images guided by unsupervised prototypes for data-efficient SSL.
	PathFinder [74]	WSI, Text	VLM	A multi-agent system that collaborates to generate comprehensive diagnostic reports.
4. Text-to-Text / Information Extraction				
2025	Saluja et al. [155]	Report	LLM	Automated assessment of cancer stage and prognosis from pathology reports.

into descriptive text, advances to enabling interactive, dialog-based inquiry of images, through the development of specialized assistants like LLaVA-Med [66], and culminates in synthesizing comprehensive diagnostic reports. A parallel thrust focuses on leveraging language models for the interpretation and reframing of existing clinical text. Representative examples of these applications are illustrated in Fig. 7 and Table III.

1) Image Captioning: The foundational task in bridging vision and language is generating accurate textual descriptions for pathological images. Methodological progress has advanced through distinct stages, addressing the unique challenges of gigapixel data and the need for clinical precision. Early work focused on patch-level captioning, adapting standard CNN-encoder and RNN/Transformer-decoder architectures to translate localized visual features into text [156], [157]. To address the broader context of Whole Slide Images (WSIs), the field shifted towards WSI-level interpretation via multi-instance aggregation. This paradigm treats a WSI as a bag of instances, developing sophisticated mechanisms to synthesize a holistic description. Innovations here include pioneering the use of Multiple Instance Learning (MIL) with dense supervision [139], and developing advanced Transformer-based architectures that use techniques like masked attention or query-based aggregation to efficiently link visual information to diagnostic captions [140], [141], [158]. Most recently, a paradigm shift has been introduced by general-purpose Vision-Language Models (VLMs) like GPT-4V, which leverage in-context learning to perform descriptive tasks with minimal prompting and no domain-specific fine-tuning [142]. In summary, image captioning in pathology has evolved from patch-level descriptions to sophisticated WSI-level interpretation. The field now faces a crossroads between specialized, fine-tuned models offering high precision but requiring extensive data, and flexible, generalist VLMs that raise concerns about reliability and consistency. The central challenge remains validating the clinical and factual accuracy of generated text, regardless of the underlying architecture.

2) Question Answering: Visual Question Answering (VQA) elevates text generation from description to interactive inquiry, enabling a dynamic dialogue with visual data. A primary challenge in this domain is the scarcity of large-scale, paired image-text data required for robust training. To mitigate this, MUSK [145] utilizes a two-stage pretraining strategy that first learns from vast unpaired corpora, while newer systems like PathGen-LLaVA [148] are trained on massive, synthetically generated datasets to overcome this bottleneck. Building upon such data-efficient foundations, another major challenge is scaling VQA from isolated patches to entire WSIs. This requires addressing both the effective processing of high-resolution patches without information loss and the aggregation of information across the slide. Accordingly, PathologyVLM [68] introduces a scale-invariant connector to better handle multi-magnification information

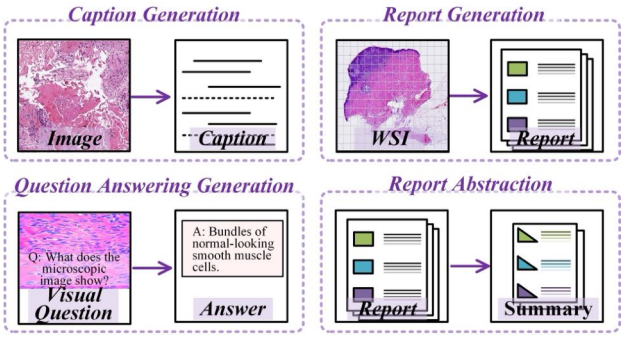

Fig. 7. Applications of text generation in computational pathology. Examples of primary text generation tasks include patch-level captioning, WSI-to-report generation, and text synthesis from image-text pairs.

inherent in pathology slides, while frameworks like WSI-VQA [159] and SlideChat [19] employ hierarchical processing to enable slide-level reasoning. PathChat [18] serves as a multimodal assistant that integrates a vision encoder with a large language model, enabling pathology-focused question answering and interaction. A critical component for clinical utility is spatial grounding, the model's ability to localize its answer to specific regions. Models like Quilt-LLaVA have been trained with spatially-aware instruction datasets to achieve this [20]. Interpretability is another key focus, with models like TraP-VQA integrating multimodal attention and visualization techniques to explain their reasoning [64], [160], prompting strategies like PathCoT [143] structure Chain-of-Thought to ensure its reliability. Furthermore, to address high computational costs, cost-effective instruction tuning methods leverage powerful LLMs to generate training data for smaller, more efficient models [144], while architectural innovations like the token compression in TCP-LLaVA [154] directly tackle the input sequence length bottleneck. In summary, pathology VQA is rapidly maturing from patch-level queries to interactive, slide-level dialogues with spatial awareness. While progress in model architecture and interpretability is significant, the field is constrained by data scarcity and the challenge of clinical validation. The ultimate success of VQA systems will depend on their ability to provide not just answers, but reliably accurate and interpretable answers that can be trusted in a diagnostic setting.

3) Report Generation: Automated report generation from visual evidence represents a pinnacle task, aiming to synthesize comprehensive, clinical-grade narratives. A key technical challenge is the effective fusion of global WSI context with salient local features. Foundational architectures address this aggregation challenge through methods like compressing tile embeddings with a perceiver network, as seen in PRISM [150], or by directly constructing multi-scale representations using hierarchical Vision Transformers (HIPT) [146]. HistoGPT [67] successfully scales report generation to the whole-slide level, processing multiple gigapixel WSIs to produce narratives of a quality comparable to human experts and exhibiting emergent zero-shot capabilities. In a conceptual leap beyond this, mSTAR [70] enriched its pretraining by incorporating gene expression data alongside images and reports, enabling the generation of narratives that directly link tissue morphology to its molecular context. Other approaches focus on enhancing spatial perception through hierarchical position-aware modules in the encoder [147] or augmenting the generation process with knowledge retrieval decoders [153]. A significant trend is the move towards emulating the pathologist's holistic workflow. This is exemplified by multi-agent AI frameworks like PathFinder, which uses collaborative agents to triage, navigate, and describe findings, thereby producing an interpretable diagnostic narrative [74]. Furthermore, models are being increasingly fine-tuned for specific clinical contexts, such as generating reports for melanocytic lesions or using parameter-efficient techniques like LoRA to adapt large VLMs for creating detailed, part-level reports from multiple slides [62], [149], [151], [152]. In summary, automated report generation has advanced from basic image-to-text mapping toward more complex diagnostic reasoning, integrating multi-scale feature aggregation and cross-slide contextual understanding. Despite notable gains in linguistic fluency and factual accuracy, clinical deployment remains hindered by the critical gap between generating plausible narratives and ensuring diagnostic validity. Achieving clinically error-free outputs demands rigorous, large-scale validation against expert standards.

4) Report Abstraction: Report abstraction has recently emerged as an important application of LLMs in computational pathology. Beyond generating reports from visual inputs, LLMs are now applied to transform unstructured pathology narratives into structured or semantically simplified representations. This enables two main functions: automated extraction of structured data for downstream analysis and decision support, and the generation of accessible summaries to improve communication with non-expert audiences.

For structured information extraction, foundation models such as GPT-4o and Llama 3 have demonstrated annotation-level accuracy in parsing free-text reports into machine-readable forms [161], [162]. Instruction-tuned variants, including Path-llama3.1-8B and Path-GPT-4o-mini-FT, further exhibit robust generalization in zero-shot settings for extracting clinically relevant parameters such as cancer type, AJCC stage, and prognosis [155]. Complementarily, LLMs have shown promise in patient-facing applications by rephrasing technical content into more interpretable language. Studies using GPT-4 indicate that such interpretive summaries can improve patient comprehension and engagement without compromising diagnostic fidelity [163], thus facilitating informed decision-making and health literacy. In summary, LLM-driven report abstraction enables both clinical efficiency and patient-centered communication. Broader adoption, however, requires rigorous validation across diverse diseases and institutions, with the critical challenge of preserving diagnostic nuance while simplifying language to balance accessibility and fidelity.

C. Molecular Profiles-Morphology Generation

The fundamental challenge of linking genotype to phenotype in biomedicine manifests acutely in oncology, where molecular alterations must be reconciled with observable morphological changes. Traditional approaches compartmentalize these domains: histopathological analysis extracts spatial architectural information from H&E-stained sections, while omics profiling quantifies molecular features at the expense of spatial context [169]. Spatial omics technologies, though promising, remain constrained by cost and technical complexity, limiting their clinical penetration [170]. Generative models learn the joint distribution of histology and molecular profiles, which enables conditional inference in two complementary directions in Table IV: inferring molecular profiles from histology, termed virtual molecular profiling, and synthesizing histology from molecular inputs, termed reverse morphology generation.

1) Virtual Molecular Profiling: Virtual molecular profiling aims to transform large repositories of low-cost H&E slides into computable molecular maps, offering transcriptomic or proteomic readouts without additional assays. A primary objective is the de novo synthesis of spatial transcriptomes. Large-scale foundation models like PAST achieve this at single-cell resolution [165], while alternative frameworks such as SCHAF employ graph-based architectures to incorporate structural priors for the same task [164]. Lightweight

TABLE IV
OVERVIEW OF GENERATIVE METHODS BRIDGING HISTOLOGY AND GENOMICS FOR CROSS-MODAL PREDICTION AND GENERATION. ABBREVIATIONS: ST, SPATIAL TRANSCRIPTOMICS; SCRNA-SEQ, SINGLE-CELL RNA SEQUENCING; TME, TUMOR MICROENVIRONMENT.

Year	Method	Input	Arch.	Key Application
1. Virtual Molecular Profiling				
2023	SpatialScope [13]	ST, scRNA-seq	VAE	Deconvoluting and imputing ST data to achieve single-cell, whole-transcriptome resolution.
2024	Diff-ST [11]	Low-Resolution ST, Patch	Diffusion	Super-resolving spatial transcriptomics using histology as guidance.
	SCHAF [164]	Patch, ST	VAE	Cross-modal prediction of single-cell gene expression from histology.
2025	PAST [165]	Patch, scRNA-seq	/	Single-cell gene expression and virtual IHC prediction from cell images.
	SPATIA [166]	Patch, scRNA-seq	/	Cross-modal generation between Patch and scRNA-seq (Bidirectional)
	GenST [12]	Patch, ST	VAE	Predicting spatial transcriptomics by aligning cross-modal latent spaces.
	Stem [167]	Patch	Diffusion	Spatial transcriptomics inference via conditional diffusion models.
	LD-CVAE [29]	WSI	VAE	Predicting surrogate genomic embeddings from WSI for survival prediction.
	HistoPlexer [168]	Patch, Protein Multiplex	GAN	Generating multiplex protein maps from H&E for TME characterization.
2. Reverse Morphology Generation				
2024	Pix2Path [7]	High-Resolution ST, Patch	GAN	Generating pathology images from ST for risk and perturbation analysis.
2025	RNA-CDM [54]	Bulk RNA-seq	Diffusion	Generating pathology images from bulk RNA-seq for augmentation.
	HistoXGAN [8]	Pathologic, genomic, and radiographic Feature	GAN	Reconstructing histology from latent multi-modal features for explainability and virtual biopsy.

models like GenST offer efficient prediction through aligned autoencoders [12]. To better capture biological heterogeneity, Stem [167] reframes deterministic regression as conditional diffusion-based generation, yielding distributions of plausible expression profiles from a single H&E patch. This generative capability extends to proteomics, where HistoPlexer synthesizes multiplexed protein maps while preserving biologically crucial co-localization patterns [168]. Beyond de novo synthesis, other models focus on enhancing existing molecular data. For instance, Diff-ST uses histology as a morphological prior to super-resolve low-resolution spatial transcriptomics [11], whereas SpatialScope computationally deconstructs mixed-cell data to achieve near single-cell resolution [13]. Generative imputation is critical for clinical robustness, as models like LD-CVAE [29] synthesize surrogate genomic embeddings from histology to stabilize multimodal predictions when molecular data are absent.

2) Reverse Morphology Generation: Reverse morphology generation serves as a powerful platform for basic biological discovery and model interpretability. By generating histology images conditioned on molecular profiles, these models provide visual hypotheses for how molecular states shape tissue architecture, creating an effective in silico experimental system. This task can be driven by inputs with varying levels of spatial information. Models like Pix2Path synthesize pathology images from high-resolution spatial transcriptomics, a capability that can be leveraged for virtual gene perturbation experiments to probe causal relationships between genes and phenotypes [7]. Addressing a more formidable challenge, RNA-CDM generates plausible tissue structures from non-spatial bulk RNA-seq data [54]. A distinct application of this reverse-generative principle is for model interpretability. HistoXGAN [8] reconstructs histology from latent vectors linked to molecular states, such as PIK3CA mutations, enabling visualization of subtle AI-learned morphological features.

Despite rapid progress, significant challenges persist, including the scarcity of high-quality, spatially aligned paired datasets [165], [166] and the difficulty in validating the biological fidelity of generated outputs. Future progress will depend on building efficient, general-purpose foundation models that support bidirectional inference, exemplified by frameworks like SPATIA [166]. The ultimate objective is to create a digital twin of tissues, a robust computational model capable of accurately simulating the morphological consequences of genetic interventions and drug responses.

D. Other Generation

Recent advances in generative modeling for pathology have expanded beyond traditional image and text synthesis to target more specialized data modalities and objectives. Current efforts can be grouped into four emerging directions, as summarized in Table V. The first focuses on spatial layout generation, simulating biologically realistic tissue and cellular organizations under structural constraints. The second emphasizes semantic output generation, producing interpretable artifacts such as prompts or embeddings to enhance model transparency and human–AI interaction. The third centers on latent representation generation, shifting synthesis from the pixel space to the latent space to improve efficiency, scalability, and downstream performance. Finally, cell simulation applies generative models to synthesize photorealistic microscopy data with programmatically defined ground truth, alleviating annotation bottlenecks for tasks such as tracking and segmentation.

1) Spatial Layout Generation: This category of generative models aims to simulate biologically realistic spatial configurations of cellular or tissue structures, enabling downstream analysis under controlled topological assumptions. The core challenge is to preserve complex inter- and intra-cellular spatial relationships. To achieve this, recent methods impose strong topological or spatial constraints on the generation process. For example, TopoCellGen integrates persistent homology into a diffusion framework to maintain spatial fidelity, while other diffusion-based approaches are guided by density-based cell layout maps [171], [172]. An alternative models the cellular environment as a graph, with DiGress [186] employing graph-based diffusion to generate cell graphs that preserve structural characteristics such as tertiary lymphoid structures. These techniques further generalize to multimodal settings, exemplified by DAMM-Diffusion [173], which predicts nanoparticle distributions by fusing structural information from the tumor microenvironment. In summary, generative spatial layout models have progressed from simple point patterns to structured tissue ecosystems, offering realistic synthetic microenvironments for biology and analysis, though validating their fidelity to complex biological rules remains a key challenge.

2) Semantic Output Generation: This area seeks to produce interpretable artifacts, including prompts and embeddings, that improve

TABLE V
OVERVIEW OF OTHER SPECIALIZED GENERATIVE METHODS IN COMPUTATIONAL PATHOLOGY.

Year	Method	Input	Arch.	Key Application
1. Spatial Layout Generation				
2024	Li et al. [171]	Patch, Layout	Diffusion	Guiding image generation and object detection using explicit spatial layout priors.
	TopoCellGen [172]	Cell Layout	Diffusion	Generating biologically plausible, topology-aware cell layouts for downstream tasks.
2025	DAMM-Diffusion [173]	Patch	Diffusion	Predicting nanoparticle distributions within tumor microenvironments for prognosis.
2. Semantic Output Generation				
2024	QAP [174]	Patch, WSI	VLM	Quantitative visual prompts from histological features for model adaptation.
	TQx [175]	Patch	VLM	Generating text-based image embeddings for explainable classification and clustering.
2025	MLLM4PUE [176]	Patch, Text	VLM	Generating universal embeddings for zero-shot classification and retrieval tasks.
3. Latent Representation Generation				
2019	Hu et al. [44]	Patch	GAN	Unsupervised visual representations from patches for segmentation and classification.
2023	PLIP [177]	Patch, Text	VLM	Learning joint image-text embeddings from web data for retrieval and classification.
2024	PRDL [178]	Patch, WSI	MIL	Prompt-guided sampling for diverse WSI representations in MIL pipelines.
	AugDiff [179]	Patch, WSI	Diffusion	Generating semantic feature augmentations to improve classification generalization.
	DCDiff [180]	Patch, WSI	Diffusion	Generating multi-resolution features for robust classification.
	Prov-GigaPath [181]	WSI	/	Generating slide-level features for downstream tasks.
	GPFM [182]	Patch	/	Generating universal feature backbone via multi-expert knowledge distillation.
2025	MExD [183]	Feature	Diffusion	Mixture-of-experts class distribution synthesis for generative WSI classification.
4. Cell Simulation				
2024	SynCellFactory [184]	Patch	Diffusion	Synthesizing realistic time-lapse cell videos with ground truth for tracking tasks.
	Bruch et al. [185]	Patch	GAN	Coherent 3D cellular structures and masks by incorporating biophysical constraints.

model transparency and facilitate human–AI interaction. One research direction translates histological features into structured visual prompts for downstream guidance, as exemplified by QAP [174]. A second line of work develops multimodal large language models that derive universal embeddings from summarization-style prompts, enabling robust zero-shot classification and retrieval [176]. A third direction aligns visual features with semantically rich spaces through training on large-scale, internet-curated image–text datasets, as in PLIP [177]. Frameworks such as TQx [175] extend this approach by generating text-based image embeddings constructed from domain-specific vocabularies, yielding representations that are inherently explainable. In summary, semantic output generation provides a bridge between computational models and clinical workflows by producing interpretable, knowledge-aligned outputs, though a key challenge remains to ensure that these semantic concepts are consistently grounded in true pathological features.

3) Latent Representation Generation: This paradigm reorients generation from pixel-level synthesis to the more abstract and efficient latent space, where intermediate feature representations are synthesized or augmented to enhance downstream robustness and performance. For instance, diffusion models can be applied directly to latent features to generate semantically coherent augmentations, a technique that has proven highly effective for improving WSI classification while avoiding costly image transformations [179]. Building on this idea, PRDL employs prompt-guided sampling in the latent space to create diverse WSI representations for multiple instance learning [178]. Recent advances have extended this paradigm to gigapixel WSIs, as demonstrated by Prov-GigaPath [181], which adapts the LongNet transformer to learn holistic latent representations from tiles, establishing a benchmark for slide-level feature generation. MExD [183] combines a mixture-of-experts aggregator with a diffusion-based generative classifier to synthesize class distributions for WSI classification, while DCDiff [180] employs a dual-granularity diffusion architecture to generate features at multiple spatial scales. Another line of work distills knowledge from multiple expert models to construct universal latent representations, exemplified by the Generalizable Pathology Foundation Model (GPFM) [182]. In summary, latent representation generation offers an efficient alternative to pixel-level synthesis, significantly enhancing downstream performance, although concerns remain regarding the biological interpretability of the resulting features.

4) Cell Simulation: This highly specialized application of generative models focuses on creating synthetic microscopy data tailored for specific downstream tasks, primarily to address the fundamental challenge of limited annotated data. The goal is to generate photorealistic data that comes with perfect, programmatically generated ground-truth annotations. For instance, to solve the cell tracking annotation bottleneck, SynCellFactory [184] uses ControlNets to decouple cellular appearance from motion dynamics, enabling the synthesis of time-lapse videos with corresponding ground-truth trajectories. Similarly, for 3D segmentation, physics-informed GANs can simultaneously generate coherent 3D cellular structures and their corresponding pixel-perfect segmentation masks, using biophysical constraints to ensure the results are biologically plausible [185]. In summary, generative cell simulation offers a targeted means of alleviating annotation bottlenecks in biomedical imaging by producing synthetic datasets with programmatically defined ground truth. Its utility, however, remains constrained by domain shift, since the reliability of downstream models depends on simulations that accurately reflect biological variability and imaging artifacts.

IV. DATASET

To support the diverse generative tasks in computational pathology, we summarize commonly used datasets spanning a wide range of applications, including classification, segmentation, stain transformation, caption generation, and question answering, as detailed in Table VI. Commonly used evaluation metrics are summarized in Supplementary Materials Table I. These datasets span a broad range of organs, modalities (e.g., whole-slide images, image patches, and textual reports), and staining types (e.g., H&E and IHC), as illustrated

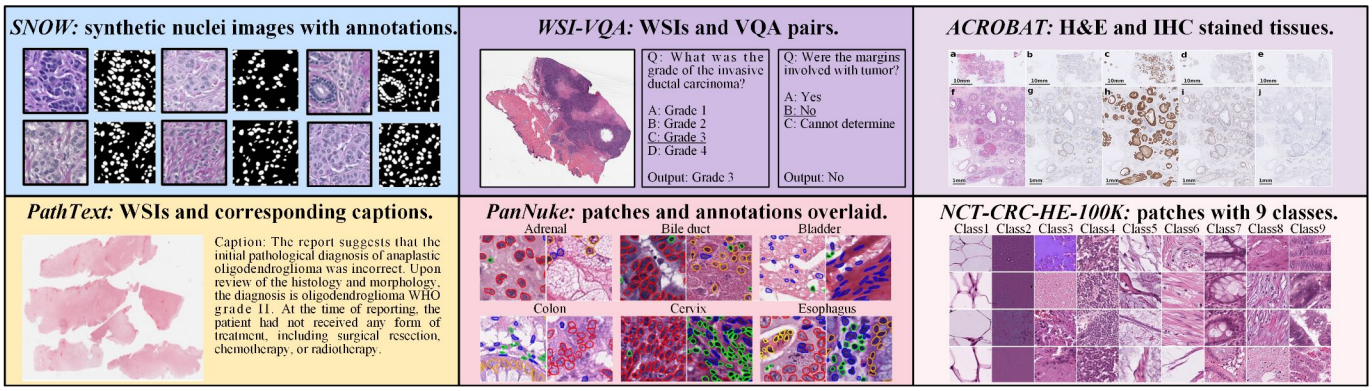

Fig. 8. Examples from representative pathology datasets showcasing variations in tissue type, staining, annotations, captions, and VQA pairs.

TABLE VI

DATASETS FOR CONTENT GENERATION MODELS IN COMPUTATIONAL PATHOLOGY, ORGANIZED BY TASK AND DATA TYPE. PAIRS INDICATE IMAGE–TEXT PAIRS.

Dataset	Organ	Modality	Stain Type	Size	Source
I. Real-World Benchmark Datasets					
A. Segmentation & Detection					
Camelyon16 [187]	Breast	WSI	H&E	400 WSIs	https://camelyon16.grand-challenge.org
CoNIC [188]	Colon	Patch	H&E	4,981 images	https://conic-challenge.grand-challenge.org
CoNSeP [189]	Colon	Patch	H&E	41 images	https://opendatalab.com/OpenDataLab/CoNSeP
DigestPath [190]	Multiple	Patch	H&E	1,559 images	https://digestpath2019.grand-challenge.org
GlaS [191]	Multiple	Patch	H&E	165 images	https://www.kaggle.com/datasets/sani84/glasmiccai2015-gland-segmentation
Lizard [192]	Colon	Patch	H&E	495,179 nuclei	https://www.kaggle.com/datasets/aadimator/lizard-dataset
MoNuSeg [193]	Multiple	Patch	H&E	29,000 nuclei	https://monuseg.grand-challenge.org/
PanNuke [194]	Multiple	Patch	H&E	7,901 images	https://huggingface.co/datasets/RationAI/PanNuke
TUPAC16 [195]	Breast	WSI	H&E	573 cases	https://tupac.grand-challenge.org
B. Classification					
BACH [196]	Breast	Patch, WSI	H&E	400+ images	https://zenodo.org/records/3632035
BreakHis [197]	Breast	Patch	H&E	9,709 images	https://web.inf.ufpr.br/vri/databases/breakhis
Camelyon17 [198]	Breast	WSI	H&E	1,399 WSIs	https://camelyon17.grand-challenge.org/
LC25000 [199]	Multiple	Patch	H&E	25,000 images	https://huggingface.co/datasets/1aurent/LC25000
NCT-CRC-HE-100K [200]	Colon	Patch	H&E	100,000 images	https://zenodo.org/records/1214456
PANDA [201]	Prostate	WSI	H&E	11,000 WSIs	https://panda.grand-challenge.org
PatchCamelyon [202]	Multiple	Patch	H&E	327,680 images	https://github.com/basveeling/pcam
TCGA-BRCA	Multiple	WSI	H&E	1,098 cases	https://portal.gdc.cancer.gov/projects/TCGA-BRCA
C. Stain Transfer					
ACROBAT [203]	Breast	WSI	H&E, IHC	4,212 WSIs	https://acrobat.grand-challenge.org
BCI [204]	Breast	WSI	H&E, IHC	9746 images	https://bci.grand-challenge.org
EMPaCT [205]	Prostate	Patch	H&E, IHC	420 images	https://zenodo.org/records/10066853
D. Multimodal & Other Tasks (Retrieval, Captioning, Biomarker Exploration)					
ARCH [206]	Multiple	Patch	/	15,164 images	https://warwick.ac.uk/fac/cross_fac/tia/data/arch
HEST-1k [207]	Multiple	WSI, omics	H&E	1,229 cases	https://huggingface.co/datasets/MahmoodLab/HEst
OpenPath [177]	Multiple	Patch, Text	/	208,414 images	https://huggingface.co/spaces/vinid/webplip
PathCap [208]	Multiple	Patch, Text	/	207K pairs	https://huggingface.co/datasets/jamessyx/PathCap
PatchGastricADC22 [209]	Stomach	Patch, Text	H&E	262K images	https://zenodo.org/records/6021442
QUILT-1M [210]	Multiple	Patch	/	1M pairs	https://zenodo.org/records/8239942
II. Generative & Synthetic Datasets					
A. Image-Synthetic					
SNOW [211]	Breast	Patch	H&E	20,000 images	https://zenodo.org/records/6633721
B. Text-Synthetic (for VQA & Report Generation)					
PathGen-1.6M [148]	Multiple	Patch, Text	Multiple	1.6M pairs	https://huggingface.co/datasets/jamessyx/PathGen
PathMMU [212]	Multiple	Patch, Text	/	24K images / 33K QA	https://huggingface.co/datasets/jamessyx/PathMMU
PathText [147]	Multiple	Patch, Text	/	9,009 WSI-text pairs	https://github.com/cpystan/Wsi-Caption
PathVQA [160]	Multiple	Patch, Text	/	5K images / 32K QA	https://huggingface.co/datasets/flaviagiammarino/path-vqa
PMC-VQA [213]	Multiple	Patch, Text	Multiple	227K VQA	https://huggingface.co/datasets/RadGenome/PMC-VQA
QUILT-VQA [20]	Multiple	Patch, Text	/	985 images / 1,283 QA	https://huggingface.co/datasets/wisdomik/Quilt_VQA
WSI-VQA [159]	Breast	WSI, Text	H&E	977 WSIs / 8,672 QA	https://github.com/cpystan/WSI-VQA

in Fig. 8, offering a comprehensive foundation for the training and evaluation of generative models in computational pathology. Notably, with the recent development of generative pathology, several synthetic benchmarks have emerged. These include both image-synthetic datasets, such as SNOW [211], which introduces a large-scale dataset of synthetic pathological images paired with semantic segmentation annotations for nuclei; and text-synthetic datasets, where generated textual reports or QA pairs are increasingly used to augment image-text training corpora. Examples include PathGen-1.6M [148], PathMMU [212], and WSI-VQA [159], which leverage

large-scale language models to generate descriptive captions and question-answer pairs.

V. Discussion

Generative artificial intelligence is increasingly reshaping computational pathology, impacting diagnostic imaging, knowledge extraction, and clinical decision-making. This discussion first examines current capabilities and demonstrated impact, then analyzes critical limitations and barriers, and finally outlines strategic pathways with future directions toward trustworthy clinical adoption.

A. Current Capabilities and Impact

1) Architectural Suitability and Task Performance: The optimal choice of a generative model in computational pathology is not absolute but is dictated by a fundamental trade-off among synthesis fidelity, computational efficiency, and architectural inductive bias. The absence of standardized, clinically meaningful benchmarks currently precludes direct numerical comparison across studies, making an understanding of these architectural trade-offs paramount. For image synthesis tasks that prioritize maximal fidelity and morphological diversity, such as virtual staining or the creation of novel cellular morphotypes, diffusion models currently set the state of the art [172]. Their iterative refinement principle enables precise modeling of complex, multi-scale image distributions, albeit at the cost of substantial computation and slow inference. Conversely, when efficiency is the primary constraint, such as in real-time data augmentation, GANs achieve lower latency through single-pass generation, although this advantage can be offset by training instability and the risk of mode collapse [214]. For text generation, autoregressive transformers excel in producing coherent and contextually consistent pathology narratives. This is primarily because their self-attention mechanism adeptly captures the long-range dependencies required to sustain logical continuity. However, these models remain critically vulnerable to factual hallucination unless rigorously constrained [215]. This landscape of architectural trade-offs underscores that claims of superior performance are contingent not just on the chosen model, but on the specific definitions, datasets, and evaluations employed.

2) Impact on Learning Paradigms: Generative models are fundamentally reshaping learning paradigms in computational pathology by shifting the objective from learning discriminative boundaries to modeling the underlying data distributions [216]. This paradigm shift translates into concrete benefits for diagnostic practice, including enhanced data augmentation, improved data efficiency, and richer multimodal representations. First, it redefines **data augmentation** from simple geometric transformations to semantic synthesis. This is critical in imbalanced tasks, such as mitotic detection or rare cancer subtype classification, where generative models synthesize realistic examples to enhance robustness [75]. Second, this focus on the data distribution inherently enables **data-efficient learning**. Generative objectives support self-supervised pre-training on unlabeled pathology data, yielding versatile backbones. These backbones can be fine-tuned for tasks such as grading, segmentation, and biomarker prediction with limited annotations. Finally, this paradigm enables deeper **multimodal representations**. By learning generative functions between modalities, models move beyond simple alignment to semantically grounded representations. This provides a more interpretable basis for tasks such as report generation, mutation prediction, and treatment response forecasting [18].

3) Methodological Advantages and Clinical Impact: Generative models derive substantial methodological value by modeling comprehensive data distributions. Their utility arises from three key methodological advantages. First, by capturing the underlying data distribution, generative models synthesize high-fidelity samples that augment rare classes and reduce annotation demands, thus improving efficiency under data scarcity and imbalance [217]. Second, reconstruction and cross-modal translation objectives yield structured and robust representations that transfer to downstream tasks such as grading, segmentation, and biomarker prediction. They also reveal morphology–molecular associations, providing a basis for hypothesis generation. Third, synthetic cohorts can be shared with reduced risk of exposing protected health information, alleviating privacy barriers and enabling multi-center collaboration.

Building on these methodological advantages, generative models deliver tangible benefits in clinical and societal impact. In the clinical environment, they enhance operational efficiency. Automated report generation can alleviate the clerical burden on pathologists [19], and techniques like virtual staining offer substantial savings in reagent costs and tissue consumption [11]. More fundamentally, generative approaches expand core diagnostic capabilities. Virtual molecular profiling enables computational inference of spatially resolved molecular maps from routine H&E images, unlocking archival resources for biomarker discovery. At the societal level, these models support standardization and broaden access to expertise. Computationally-driven stain normalization can reduce inter-laboratory variability, improving diagnostic consistency for multi-center trials and telepathology. The synthesis of diverse virtual slide libraries, including rare cases, provides a scalable and low-cost resource for global pathology education. By encapsulating specialist knowledge, these models can also function as critical assistive tools in resource-limited settings, thereby helping to mitigate global health disparities.

B. Critical Barriers to Clinical Translation

Despite the aforementioned advantages, generative models face enduring barriers to clinical translation. These challenges fall into three categories: intrinsic model limitations, methodological and technical bottlenecks, and systemic obstacles to real-world deployment.

1) Intrinsic Model Limitations: **Interpretability, Reliability, and Trust Deficit.** Generative models are trained to approximate data distributions rather than to establish verifiable input–output mappings, which inherently limits their transparency and hampers systematic auditing. Likelihood-driven objectives favor syntactic and visual plausibility over factual accuracy, increasing the risk of hallucinated outputs that compromise interpretation [215]. Domain shifts exacerbate these limitations, introducing instability that undermines robustness across diverse institutional and patient cohorts. In segmentation, models may produce spurious tumor boundaries with misleading precision, potentially misdirecting disease assessment. The high visual fidelity of such errors makes them difficult to detect, unlike obvious artifacts such as blurring, and thereby increases clinical risk. The convergence of limited traceability, poor controllability, and vulnerability to clinical variability jointly creates interpretability challenges that perpetuate institutional mistrust and impede clinical translation. Thus, this opacity raises particular concern in high-stakes diagnostics, where accountability is vital, as it perpetuates a trust deficit that slows clinical adoption.

2) Technical and Evaluation Bottlenecks: **WSI Generation Challenges.** The synthesis of realistic whole-slide images (WSIs) remains a significant challenge due to their gigapixel-scale resolution and the scarcity of comprehensively annotated training data. Current patch-based approaches effectively capture local histological patterns but still struggle to preserve global spatial coherence and tissue-wide architectural integrity. This limitation substantially impairs their utility for applications requiring long-range contextual information, such as tumor staging or morphological heterogeneity

assessment. While hierarchical generation frameworks and multiscale attention mechanisms offer promising directions by modeling cross-scale dependencies, they impose prohibitive computational demands that often render them largely impractical on standard hardware [48], [113]. Therefore, developing memory-efficient, resolution-aware architectures is a critical research priority. Such advances are essential for modeling complex spatial relationships in histopathological data and moving WSI synthesis toward clinical viability.

Evaluation Limitation. The absence of standardized, pathology-specific evaluation protocols remains a critical barrier to the clinical adoption of generative models in computational pathology. Conventional metrics such as FID, SSIM, and PSNR, originally developed for natural image assessment, capture only low-level similarity and fail to reflect diagnostic features such as cellular morphology, glandular architecture, or spatial organization [218]. This mismatch risks producing outputs that appear visually plausible but are clinically irrelevant or misleading. Although task-specific evaluation metrics tailored to pathology have been proposed, they remain fragmented and lack systematic validation across diverse datasets and pathological conditions. Furthermore, the field still lacks standardized benchmark datasets and reproducible evaluation pipelines, hindering fair comparison and robust progress [219]. Addressing these limitations will require the development of pathology-specific benchmarks and multidimensional evaluation frameworks that align with clinical needs.

3) Systemic Deployment Obstacles: **Barriers to Clinical Deployment.** The translation of generative models into clinical practice is constrained by systemic barriers that involve technical, regulatory, and socio-technical domains. Technically, state-of-the-art models require substantial computational resources for inference, often exceeding the infrastructure and budgets of typical clinical environments [220]. Performance degradation due to evolving clinical practices and patient populations necessitates continuous monitoring and periodic retraining, an operational burden that few institutions can sustain. Regulatory and legal uncertainties pose equally significant obstacles. Existing frameworks were built for discriminative algorithms with static performance, whereas generative models produce novel outputs, raising unresolved questions about validation, monitoring, and post-market surveillance [221]. Liability remains ambiguous, with responsibility for errors caused by generated artifacts not clearly assigned. Finally, socio-technical factors complicate deployment. Human–AI interaction is vulnerable to socio-technical risks, including deployment bias when systems are used beyond their validated scope and automation complacency when clinicians over-rely on algorithmic outputs [222]. These risks are compounded by persistent challenges of clinical validation. Appropriate endpoints for generative systems remain undefined, and large-scale prospective trials required to establish safety and utility are constrained by substantial logistical and financial demands. Confronting these barriers is essential to a credible path toward safe and effective deployment in pathology.

Computational Resource Constraints. Generative model development incurs substantial costs across three phases of the computational pipeline. (1) **Data infrastructure** is burdened by the acquisition, curation, and storage of gigapixel whole-slide images, which demand extensive storage systems and costly expert annotation workflows. (2) **Training costs** scale exponentially with model complexity [223]. Conventional CycleGAN (11.8M parameters) [224] remains feasible with the computational resources typically available in academic settings, whereas latent diffusion models (1.4B parameters) [56] and multimodal foundation systems demand multi-GPU training over several days. For instance, while Quilt-LLaVA [20] was trained on LLaVA-v1.5-7B within approximately 10 hours using 4 NVIDIA A100 GPUs, PathChat [18] required a more resource-intensive multistage process, including 32 hours on 32 A100 GPUs for pretraining and an additional 39 hours on 8 A100 GPUs for fine-tuning and joint training. These comparisons highlight the steep escalation in training costs as models progress from lightweight GANs to diffusion and foundation-scale architectures. (3) **Deployment costs** persist throughout clinical implementation. These are driven by the substantial computational resources required for high-latency inference, as well as the need for recurrent retraining and monitoring to address data drift and emergent biases.

These demands create systemic barriers that affect both scientific equity and clinical accessibility. Frontier model development remains largely concentrated within well-resourced institutions, limiting participation from smaller groups and constraining reproducibility [225], [226]. Clinically, high operational costs restrict deployment in resource-constrained healthcare systems, exacerbating diagnostic disparities and constraining the democratization of advanced pathology tools. Addressing these constraints will require advances in algorithmic efficiency, along with open dissemination of pretrained models and the establishment of standardized benchmarks. Ultimately, the clinical translation of generative pathology depends on reconciling technical innovation with economic sustainability and equitable accessibility.

Ethical, Legal, and Security Considerations. The integration of synthetic data in computational pathology raises interconnected challenges across ethical, legal, and security domains [227] that must be addressed prior to clinical implementation. Ethically, synthetic histopathology images may appear realistic yet contain biological inconsistencies, risking diagnostic errors by potentially misleading both algorithms and human experts. This concern is particularly acute in high-stakes clinical environments where interpretability and reliability are essential for trust. Legally, the governance of synthetic medical data remains underdeveloped, with unresolved issues spanning patient consent, intellectual property rights, and liability for diagnostic errors. From a security perspective, models trained on synthetic distributions may be less robust to adversarial manipulations, especially when the generated data fail to capture the full heterogeneity and rare edge cases present in clinical specimens.

C. Future Opportunities

To advance generative models in computational pathology, future efforts must prioritize the deeper integration of data modalities, model capabilities, and clinical needs. This roadmap encompasses unified architectures, multimodal generation, robust evaluation, and enhanced clinical interpretability.

Generative Foundation Model. The next stage of generative pathology depend on moving beyond task-specific architectures toward foundation-scale models that can serve as unified engines for diverse clinical and research needs. Current approaches remain fragmented, with models often optimized for narrow objectives or confined to representation learning [228]. Generative foundation models (GFMs) are not merely another technical direction but the logical culmination of recent progress, aiming to integrate whole-slide synthesis, mask-conditioned augmentation, semantic reporting, and molecular prediction within a shared and scalable framework. Through multimodal pretraining that aligns histology, genomic profiles, molecular markers, and clinical metadata, GFMs could support cross-modal tasks such as image-to-text, text-to-image, and gene-to-image synthesis. Such integration is crucial for capturing biologically grounded disease representations and for modeling temporal dynamics when sequential histology is combined with longitudinal clinical data. These models could address entrenched challenges of data scarcity, class imbalance, and fragmented pipelines, while enhancing few-shot generalization and cross-domain transferability with minimal supervision. Realizing clinically viable GFMs, however, requires

more than technical innovation. It demands strategies to manage the computational burden of foundation-scale training [219], ensure semantic consistency across modalities, and establish evaluation and governance frameworks appropriate for clinical contexts.

Agent-driven Generative Pathology The emerging field of Agent AI moves beyond static generation toward interactive systems that execute expert workflows through coordinated perception, reasoning, and action [229]. In pathology, this paradigm is already being applied to mirror the diagnostic process, with recent work revealing distinct strategies for action, evidence processing, and report generation [73], [74], [208]. They exemplify the integration of multi-role coordination, active exploration, and language-based reasoning. The next frontier lies in evolving from passive interpretation to active hypothesis generation and testing. Embedding generative capabilities into the agent loop could transform agents from observers into experimentalists, capable of conducting in silico experiments such as counterfactual simulations of tissue morphology or virtual synthesis of rare diseases. Realizing this vision and ensuring clinical deployability require advances in data, training, and evaluation to capture causal dynamics, unify perception and action, and ensure accurate and robust hypotheses. Maturing these components will enable agent-driven generative pathology to become a dependable partner, capable of delivering insights aligned with the reasoning logic of pathologists.

AI Virtual Cell. One long-term direction for generative pathology is the AI Virtual Cell, a multi-scale digital representation that simulates cellular states and perturbation responses in silico [230]. This paradigm shifts the field from static image synthesis to dynamic, hypothesis-driven modeling by leveraging advances in spatial omics and foundational cell atlases [231], [232]. Virtual cells could serve as priors for generating synthetic tissue at single-cell resolution, with controlled variation in cellular composition and molecular states. They could also enable perturbation-conditioned modeling to simulate disease progression and treatment response, supporting diagnostic training and virtual trials. Furthermore, aligning histology with molecular networks would facilitate mechanistic hypothesis testing, with synthetic microenvironments validated against multimodal ground truths. Realizing this vision requires standardized multi-omics datasets, unified generative–predictive frameworks, and evaluation protocols that assess morphological fidelity, molecular coherence, and causal validity. Together, these advances could transform generative pathology into a hypothesis-aware engine for decoding the cellular basis of disease.

Clinically Deployable Generative Systems. The translation of generative models from research prototypes to clinically deployable systems remains a defining challenge. This requires moving beyond isolated performance metrics toward a comprehensive framework that integrates validation and governance. First, a deployable system must demonstrate **technical robustness**, combining high-fidelity outputs with effective safeguards against factual hallucinations to ensure reliability in clinical use [233]. Such reliability must be sustained under real-world data shifts, a persistent obstacle in practice [2]. Second, future systems must prove their value through seamless **clinical integration**, functioning as controllable and interpretable tools for pathologists. Third, these capabilities must be reinforced by a **governance framework** encompassing regulatory approval pathways, ethical safeguards against bias, and explicit accountability mechanisms [234]. Therefore, establishing standards of accuracy, utility, and trustworthiness defines the path toward clinically deployable generative systems in pathology.

VI. Conclusion

Content generation modeling has rapidly gained prominence in computational pathology, becoming a central paradigm for advancing data efficiency, representation learning, and clinical translation. This survey systematically reviewed recent progress by organizing generative approaches into four major task domains: image generation, text generation, molecular profile–morphology generation, and other specialized generation tasks. The reviewed literature highlights methodological innovations across GANs, diffusion models, and language–vision architectures, showing a clear evolution from proof-of-concept synthesis toward clinically meaningful objectives such as data augmentation, report generation, and biomarker discovery. As the field progresses, content generation foundation models are anticipated to provide scalable, trustworthy, and clinically deployable systems. We hope this work offers both a reference framework and a foundation for future research in this rapidly advancing domain.

References

[1] B. van Breugel *et al.*, "Synthetic data in biomedicine via generative artificial intelligence," *Nat. Rev. Bioeng.*, vol. 2, no. 12, pp. 991–1004, Oct. 2024.

[2] A. Zhang *et al.*, "Shifting machine learning for healthcare from development to deployment and from models to data," *Nat. Biomed. Eng.*, vol. 6, no. 12, pp. 1330–1345, Dec. 2022.

[3] A. Kazerouni *et al.*, "Diffusion models in medical imaging: A comprehensive survey," *Med. Image Anal.*, vol. 88, Aug. 2023.

[4] S. P. Deshpande, "Generative ai for computational pathology," Ph.D. dissertation, University of Warwick, Coventry, U.K., Aug. 2023. [Online]. Available: https://wrap.warwick.ac.uk/185054/

[5] A. H. Song *et al.*, "Artificial intelligence for digital and computational pathology," *Nat. Rev. Bioeng.*, vol. 1, no. 12, pp. 930–949, Dec. 2023.

[6] F.-A. Croitoru *et al.*, "Diffusion models in vision: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 9, pp. 10 850–10 869, Sept. 2023.

[7] X. Fu and Y. Chen, "Pix2path: Integrating spatial transcriptomics and digital pathology with deep learning to score pathological risk and link gene expression to disease mechanisms," *bioRxiv*, 2024.

[8] F. M. Howard *et al.*, "Generative adversarial networks accurately reconstruct pan-cancer histology from pathologic, genomic, and radiographic latent features," *Sci. Adv.*, vol. 10, p. eadq0856, 2024, *Corresponding authors.

[9] J. Ke *et al.*, "Artifact detection and restoration in histology images with stain-style and structural preservation," *IEEE Trans. Med. Imag.*, vol. 42, no. 12, pp. 3487–3500, 2023.

[10] T. Wang *et al.*, "ODA-GAN: Orthogonal decoupling alignment GAN assisted by weakly-supervised learning for virtual immunohistochemistry staining," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, June 2025, pp. 25 920–25 929.

[11] X. Wang *et al.*, "Cross-modal diffusion modelling for super-resolved spatial transcriptomics," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.* Springer, 2024, pp. 98–108.

[12] R. Wood *et al.*, "GenST: A generative cross-modal model for predicting spatial transcriptomics from histology images," in *MICCAI Workshop on Computational Pathology with Multimodal Data (COMPAYL)*, 2025.

[13] X. Wan *et al.*, "Integrating spatial and single-cell transcriptomics data using deep generative models with spatialscope," *Nat. Commun.*, vol. 14, no. 1, p. 7848, November 2023.

[14] J. Boyd *et al.*, "Self-supervised representation learning using visual field expansion on digital pathology," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2021, pp. 639–647.

[15] M. Aversa *et al.*, "Diffinfinite: Large mask-image synthesis via parallel random patch diffusion in histopathology," in *Proc. Adv. Neural Inf. Process. Syst.*, 2023, pp. 78 126–78 141.

[16] Z. He *et al.*, "Artifact restoration in histology images with diffusion probabilistic models," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, 2023, pp. 518–527.

[17] Y. Shen and J. Ke, "StainDiff: Transfer stain styles of histology images with denoising diffusion probabilistic models and self-ensemble," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, 2023, pp. 549–559.

[18] M. Y. Lu *et al.*, "A multimodal generative ai copilot for human pathology," *Nature*, vol. 634, pp. 466–473, 2024.

[19] Y. Chen *et al.*, "SlideChat: A large vision-language assistant for whole-slide pathology image understanding," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, June 2025, pp. 5134–5143.

[20] M. S. Seyfioglu *et al.*, "Quilt-LLaVA: Visual instruction tuning by extracting localized narratives from open-source histopathology videos," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2024, pp. 13 183–13 192.

[21] C. He *et al.*, "Diffusion models in low-level vision: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2025, to be published.

[22] Z. Guo *et al.*, "Diffusion models in bioinformatics and computational biology," *Nat. Rev. Bioeng.*, vol. 2, no. 2, pp. 136–154, Feb. 2024.

[23] D. Chanda *et al.*, "A new era in computational pathology: A survey on foundation and vision-language models," *arXiv:2408.14496*, Aug. 2024.

[24] M. Bilal *et al.*, "Foundation models in computational pathology: A review of challenges, opportunities, and impact," *arXiv:2502.08333*, Feb. 2025.

[25] A. Radford *et al.*, "Learning transferable visual models from natural language supervision," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 8748–8763.

[26] C. Jia *et al.*, "Scaling up visual and vision-language representation learning with noisy text supervision," in *Int. Conf. Mach. Learn.* PMLR, 2021, pp. 4904–4916.

[27] M. Oquab *et al.*, "Dinov2: Learning robust visual features without supervision," *arXiv:2304.07193*, 2023.

[28] D. P. Kingma and M. Welling, "An introduction to variational autoencoders," *Found. Trends Mach. Learn.*, vol. 12, no. 4, pp. 307–392, Dec. 2019.

[29] J. Zhou *et al.*, "Robust multimodal survival prediction with conditional latent differentiation variational autoencoder," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, Vancouver, BC, Canada, 2025, pp. 10 384–10 393.

[30] K. Tang *et al.*, "Self-supervised representation distribution learning for reliable data augmentation in histopathology WSI classification," *IEEE Trans. Med. Imag.*, vol. 44, no. 2, pp. 462–474, Feb. 2025.

[31] I. J. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, Montreal, QC, Canada, 2014, pp. 2672–2680.

[32] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv:1411.1784*, Nov. 2014.

[33] P. Isola *et al.*, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Honolulu, HI, USA, Jul. 2017, pp. 1125–1134.

[34] N. Bouteldja *et al.*, "Tackling stain variability using CycleGAN-based stain augmentation," *J. Pathol. Inform.*, vol. 13, Dec. 2022.

[35] F. G. Zanjani *et al.*, "Stain normalization of histopathology images using generative adversarial networks," in *Proc. IEEE Int. Symp. Biomed. Imag.*, Washington, DC, USA, Apr. 2018, pp. 573–577.

[36] J. Breen *et al.*, "Generative adversarial networks for stain normalisation in histopathology," in *Applications of Generative AI*, A. Khanna *et al.*, Eds. Cham, Switzerland: Springer, 2024, ch. 10, pp. 227–247.

[37] M. Kawai *et al.*, "Virtual multi-staining in a single-section view for renal pathology using generative adversarial networks," *Comput. Biol. Med.*, vol. 182, Aug. 2024.

[38] S. Liu *et al.*, "Unpaired stain transfer using pathology-consistent constrained generative adversarial networks," *IEEE Trans. Med. Imag.*, vol. 40, no. 8, pp. 1977–1989, Aug. 2021.

[39] Y. Shen *et al.*, "A federated learning system for histopathology image analysis with an orchestral stain-normalization GAN," *IEEE Trans. Med. Imag.*, vol. 42, no. 7, pp. 1969–1981, Jul. 2023.

[40] T. Karras *et al.*, "A style-based generator architecture for generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, Long Beach, CA, USA, Jun. 2019, pp. 4401–4410.

[41] C. D. Bahadir *et al.*, "Characterizing the features of mitotic figures using a conditional diffusion probabilistic model," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, Vancouver, Canada, Oct. 2023, pp. 121–131.

[42] A. Golfe *et al.*, "ProGleason-GAN: Conditional progressive growing GAN for prostatic cancer Gleason grade patch synthesis," *Comput. Methods Programs Biomed.*, vol. 240, Aug. 2023.

[43] S. Min *et al.*, "Co-synthesis of histopathology nuclei image-label pairs using a context-conditioned joint diffusion model," in *Proc. Eur. Conf. Comput. Vis.*, Milan, Italy, Sept./Oct. 2024, pp. 146–162.

[44] B. Hu *et al.*, "Unsupervised learning for cell-level visual representation in histopathology images with generative adversarial networks," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 4, pp. 1316–1328, Jul. 2019.

[45] J. Ho *et al.*, "Denoising diffusion probabilistic models," in *Proc. Adv. Neural Inf. Process. Syst.*, Dec. 2020, pp. 6840–6851.

[46] J. Song *et al.*, "Denoising diffusion implicit models," in *Proc. Int. Conf. Learn. Represent.*, May 2021.

[47] Y. Song *et al.*, "Score-based generative modeling through stochastic differential equations," in *Proc. Int. Conf. Learn. Represent.*, May 2021.

[48] R. Harb *et al.*, "Diffusion-based generation of histopathological whole slide images at a gigapixel scale," 2023.

[49] H.-J. Oh and W.-K. Jeong, "Diffmix: Diffusion model-based data synthesis for nuclei segmentation and classification in imbalanced pathology image datasets," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, 2023, pp. 337–345.

[50] X. Xu *et al.*, "Histo-diffusion: A diffusion super-resolution method for digital pathology with comprehensive quality assessment," *arXiv:2408.15218*, 2024.

[51] C. Wang *et al.*, "Histology image artifact restoration with lightweight transformer based diffusion model," in *Artif. Intell. Med.*, 2024, pp. 81–89.

[52] M. M. Ho *et al.*, "Disc: Latent diffusion models with self-distillation from separated conditions for prostate cancer grading," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2024, pp. 1–5.

[53] A. Shrivastava and P. T. Fletcher, "NASDM: Nuclei-aware semantic histopathology image generation using diffusion models," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, Vancouver, Canada, Oct. 2023, pp. 585–594.

[54] F. Carrillo-Perez *et al.*, "Generation of synthetic whole-slide image tiles of tumours from rna-sequencing data via cascaded diffusion models," *Nat. Biomed. Eng.*, vol. 9, no. 3, pp. 320–332, March 2025.

[55] J. Ho and T. Salimans, "Classifier-free diffusion guidance," *arXiv:2207.12598*, 2022.

[56] R. Rombach *et al.*, "High-resolution image synthesis with latent diffusion models," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2022, pp. 10 684–10 695.

[57] L. Zhang *et al.*, "Adding conditional control to text-to-image diffusion models," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2023, pp. 3836–3847.

[58] L. Žigutytė *et al.*, "Counterfactual diffusion models for mechanistic explainability of artificial intelligence models in pathology," *bioRxiv*, 2025.

[59] F. Bordes *et al.*, "An introduction to vision-language modeling," *arXiv:2405.17247*, 2024.

[60] S. Wang *et al.*, "Learning visual grounding from generative vision and language model," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.* IEEE, 2025, pp. 8057–8067.

[61] S. Javed *et al.*, "Cplip: Zero-shot learning for histopathology with comprehensive vision-language alignment," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2024, pp. 11 450–11 459.

[62] F. Ahmed *et al.*, "Pathalign: A vision-language model for whole slide images in histopathology," in *MICCAI Workshop on Computational Pathology with Multimodal Data (COMPAYL)*, 2024.

[63] H. Liu *et al.*, "Visual instruction tuning," *Proc. Adv. Neural Inf. Process. Syst.*, vol. 36, pp. 34 892–34 916, 2023.

[64] U. Naseem *et al.*, "Vision-language transformer for interpretable pathology visual question answering," *IEEE J. Biomed. Health Inform.*, vol. 27, pp. 1681–1690, 2023.

[65] T. Tu *et al.*, "Towards generalist biomedical ai," *NEJM AI*, vol. 1, p. AIoa2300138, 2024.

[66] C. Li *et al.*, "LLaVA-med: Training a large language-and-vision assistant for biomedicine in one day," in *Proc. Adv. Neural Inf. Process. Syst.*, New Orleans, LA, USA, Dec. 2023, pp. 28 541–28 564.

[67] M. Tran *et al.*, "Generating dermatopathology reports from gigapixel whole slide images with HistoGPT," *Nat. Commun.*, vol. 16, no. 1, p. 4886, 2025.

[68] D. Dai *et al.*, "Pathologyvlm: a large vision-language model for pathology image understanding," *Artif. Intell. Rev.*, vol. 58, p. 186, 2025.

[69] J. Li *et al.*, "Topofm: Topology-guided pathology foundation model for high-resolution pathology image synthesis with cellular-level control," *IEEE Trans. Med. Imag.*, 2025.

[70] Y. Xu *et al.*, "A multimodal knowledge-enhanced whole-slide pathology foundation model," 2025.

[71] S. Hua *et al.*, "Pathoduet: Foundation models for pathological slide analysis of h&e and ihc stains," *Med. Image Anal.*, vol. 97, p. 103289, 2024.

[72] W. Chen *et al.*, "Star-rl: Spatial-temporal hierarchical reinforcement learning for interpretable pathology image super-resolution," *IEEE Trans. Med. Imag.*, 2024.

[73] C. Chen *et al.*, "Evidence-based diagnostic reasoning with multi-agent copilot for human pathology," *arXiv:2506.20964*, 2025.

[74] F. Ghezloo *et al.*, "Pathfinder: A multi-modal multi-agent system for medical diagnostic decision-making applied to histopathology," *arXiv:2502.08916*, 2025.

[75] Y. Xue *et al.*, "Selective synthetic augmentation with histogan for improved histopathology image classification," *Med. Image Anal.*, vol. 67, p. 101816, 2021.

[76] M. Li *et al.*, "Unified framework for histopathology image augmentation and classification via generative models," in *Proc. Int. Conf. Digit. Image Comput., Tech. Appl.*, 2024, pp. 462–469.

[77] D. J. Van Booven *et al.*, "Mitigating bias in prostate cancer diagnosis using synthetic data for improved ai driven gleason grading," *NPJ Precis. Oncol.*, vol. 9, no. 1, p. 151, 2025.

[78] Y. Lin *et al.*, "Insmix: Towards realistic generative data augmentation for nuclei instance segmentation," 2022.

[79] X. Xu *et al.*, "Vit-dae: Transformer-driven diffusion autoencoder for histopathology image analysis," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, 2023, pp. 66–76.

[80] X. Yu *et al.*, "Diffusion-based data augmentation for nuclei image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.* Springer, 2023, pp. 592–602.

[81] J. Wang and J. Kwak, "Usegmix: Unsupervised segment mix for efficient data augmentation in pathology images," in *MICCAI Workshop on Data Engineering in Medical Imaging*, 2024, pp. 54–63.

[82] A. Graikos *et al.*, "Learned representation-guided diffusion models for large-image generation," pp. 8532–8542, 2024.

[83] E. Redekop *et al.*, "Prototype-guided diffusion for digital pathology: Achieving foundation model performance with minimal clinical data," in *Proc. Conf. Comput. Vis. Pattern Recog. Workshops*, June 2025, pp. 5187–5195.

[84] Z. Liu *et al.*, "Generating progressive images from pathological transitions via diffusion model," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, 2024, pp. 308–318.

[85] V. Porter *et al.*, "Optimising diffusion models for histopathology image synthesis," in *Proc. Brit. Mach. Vis. Conf.*, 2024.

[86] W.-H. Li *et al.*, "Pdseg: Patch-wise distillation and controllable image generation for weakly-supervised histopathology tissue segmentation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2025, pp. 1–5.

[87] M. Sun *et al.*, "Enhancing gland segmentation in colon histology images using an instance-aware diffusion model," *Comput. Biol. Med.*, vol. 166, p. 107527, 2023.

[88] J. Linmans *et al.*, "Diffusion models for out-of-distribution detection in digital pathology," *Med. Image Anal.*, vol. 93, p. 103088, 2024.

[89] A. C. Quiros *et al.*, "Pathologygan: Learning deep representations of cancer tissue," *arXiv:1907.02644*, 2019.

[90] M. Liang *et al.*, "Multi-scale self-attention generative adversarial network for pathology image restoration," *Vis. Comput.*, vol. 39, pp. 4305–4321, 2023.

[91] P. A. Moghadam *et al.*, "A morphology focused diffusion probabilistic model for synthesis of histopathology images," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, 2023, pp. 2000–2009.

[92] S. S. Ghahfarokhi *et al.*, "Deep learning for automated detection of breast cancer in deep ultraviolet fluorescence images with diffusion probabilistic model," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2024, pp. 1–5.

[93] I. Ktena *et al.*, "Generative models improve fairness of medical classifiers under distribution shifts," *Nat. Med.*, vol. 30, pp. 1166–1173, 2024.

[94] M. Pozzi *et al.*, "Generating and evaluating synthetic data in digital pathology through diffusion models," *Sci. Rep.*, vol. 14, p. 28435, 2024.

[95] J. Ye *et al.*, "A multi-attribute controllable generative model for histopathology image synthesis," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, 2021, pp. 613–623.

[96] S. Butte *et al.*, "Sharp-gan: Sharpness loss regularized gan for histopathology image synthesis," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2022, pp. 1–5.

[97] S. Cechnicka *et al.*, "Realistic data enrichment for robust image segmentation in histopathology," in *MICCAI Workshop on Domain Adaptation and Representation Transfer*, 2023, pp. 63–72.

[98] M. Öttl *et al.*, "Style-extracting diffusion models for semi-supervised histopathology segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2024, pp. 236–252.

[99] S. Deshpande *et al.*, "Synclay: Interactive synthesis of histology images from bespoke cellular layouts," *Med. Image Anal.*, vol. 91, p. 102995, 2024.

[100] X. Zhang *et al.*, "Hadiff: hierarchy aggregated diffusion model for pathology image segmentation," *Vis. Comput.*, pp. 1–12, 2025.

[101] D. Winter *et al.*, "Mask-guided cross-image attention for zero-shot in-silico histopathologic image generation with a diffusion model," 2025.

[102] M. Jehanzaib *et al.*, "A robust image segmentation and synthesis pipeline for histopathology," *Med. Image Anal.*, vol. 99, p. 103344, 2025.

[103] H. Liu *et al.*, "Pathopainter: Augmenting histopathology segmentation via tumor-aware inpainting," *arXiv:2503.04634*, 2025.

[104] B. Venkatesh *et al.*, "Restoration of marker occluded hematoxylin and eosin stained whole slide histology images using generative adversarial networks," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2020, pp. 591–595.

[105] R. Rong *et al.*, "Enhanced pathology image quality with restore–generative adversarial network," *Am. J. Pathol.*, vol. 193, pp. 404–416, 2023.

[106] M. Fuchs *et al.*, "HARP: Unsupervised histopathology artifact restoration," in *Proc. Int. Conf. Med. Imag. Deep Learn.* PMLR, 2024, pp. 465–479.

[107] Z. He *et al.*, "Latentartifusion: An effective and efficient histological artifacts restoration framework," in *MICCAI Workshop on Deep Generative Models*, 2024, pp. 202–211.

[108] C. Wang *et al.*, "Artidiffuser: A unified framework for artifact restoration and synthesis for histology images via counterfactual diffusion model," *Med. Image Anal.*, vol. 102, p. 103567, 2025.

[109] A. B. Levine *et al.*, "Synthesis of diagnostic quality cancer pathology images by generative adversarial networks," *J. Pathol.*, vol. 252, pp. 178–188, 2020.

[110] A. Lahiani *et al.*, "Seamless virtual whole slide image synthesis and validation using perceptual embedding consistency," *IEEE J. Biomed. Health Inform.*, vol. 25, pp. 403–411, 2021.

[111] S. Deshpande *et al.*, "SAFRON: stitching across the frontier network for generating colorectal cancer histology images," *Med. Image Anal.*, vol. 77, p. 102337, 2022.

[112] J. Li *et al.*, "Pathup: Patch-wise timestep tracking for multi-class large pathology image synthesising diffusion model," in *Proc. ACM Int. Conf. Multimed.*, 2024, pp. 3984–3993.

[113] S. Cechnicka *et al.*, "Urcdm: Ultra-resolution image synthesis in histopathology," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, 2024, pp. 535–545.

[114] D. Thakkar *et al.*, "Comparative analysis of diffusion generative models in computational pathology," *arXiv:2411.15719*, 2024.

[115] S. Yellapragada *et al.*, "Pathldm: Text conditioned latent diffusion model for histopathology," 2023.

[116] S. Dubey *et al.*, "Vims: virtual immunohistochemistry multiplex staining via text-to-stain diffusion trained on uniplex stains," in *Proc. Int. Workshop Mach. Learn. Med. Imag.*, 2024, pp. 143–155.

[117] H. Cho *et al.*, "Neural stain-style transfer learning using gan for histopathological images," 2017.

[118] T. de Bel *et al.*, "Residual cyclegan for robust domain transformation of histopathological tissue slides," *Med. Image Anal.*, vol. 70, p. 102004, 2021.

[119] M. Runz *et al.*, "Normalization of he-stained histological images using cycle consistent generative adversarial networks," *Diagn. Pathol.*, vol. 16, p. 71, 2021.

[120] C. Cong *et al.*, "Colour adaptive generative networks for stain normalisation of histopathology images," *Med. Image Anal.*, vol. 82, p. 102580, 2022.

[121] J. Jeong *et al.*, "Stain normalization using score-based diffusion model through stain separation and overlapped moving window patch strategies," *Comput. Biol. Med.*, vol. 152, p. 106335, 2023.

[122] Z. Li *et al.*, "Av-gan: Attention-based varifocal generative adversarial network for uneven medical image translation," 2024.

[123] R. Jewsbury *et al.*, "Stainfuser: Controlling diffusion for faster neural style transfer in multi-gigapixel histology images," 2024.

[124] T. Kataria *et al.*, "Staindiffuser: Multitask dual diffusion model for virtual staining," 2024.

[125] W. Lou *et al.*, "Multi-modal denoising diffusion pre-training for whole-slide image classification," in *Proc. ACM Int. Conf. Multimed.*, 2024, p. 10804–10813.

[126] P. Pati *et al.*, "Accelerating histopathology workflows with generative ai-based virtually multiplexed tumour profiling," *Nat. Mach. Intell.*, vol. 6, pp. 1077–1093, 2024.

[127] C.-C. Tsai *et al.*, "Test-time stain adaptation with diffusion models for histopathology image classification," in *Proc. Eur. Conf. Comput. Vis.*, 2025, pp. 257–275.

[128] Y. He *et al.*, "Pst-diff: achieving high-consistency stain transfer by diffusion models with pathological and structural constraints," *IEEE Trans. Med. Imag.*, 2024.

[129] M. M. Ho *et al.*, "F2fldm: Latent diffusion models with histopathology pre-trained embeddings for unpaired frozen section to ffpe translation," 2024.

[130] D. Reisenbüchler *et al.*, "Unsupervised latent stain adaptation for computational pathology," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, 2024, pp. 755–765.

[131] N. Sridhar *et al.*, "Diffusion models for generative histopathology," in *Deep Generative Models*, 2024, pp. 154–163.

[132] X. Yan *et al.*, "Versatile stain transfer in histopathology using a unified diffusion framework," *bioRxiv*, 2024.

[133] B. Xiong *et al.*, "Unpaired multi-domain histopathology virtual staining using dual path prompted inversion," *Proc. AAAI Conf. Artif. Intell.*, vol. 39, pp. 8780–8787, 2025.

[134] S. Liu *et al.*, "Generating seamless virtual immunohistochemical whole slide images with content and color consistency," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2025, pp. 1–5.

[135] J. Wang *et al.*, "A value mapping virtual staining framework for large-scale histological imaging," *arXiv:2501.03592*, Jan. 2025.

[136] Z. Du *et al.*, "Deeply supervised two stage generative adversarial network for stain normalization," *Sci. Rep.*, vol. 15, p. 7068, 2025.

[137] X. Zheng *et al.*, "Diffusion-based virtual staining from polarimetric mueller matrix imaging," 2025.

[138] S. A. Taqi *et al.*, "A review of artifacts in histopathology," *J. Oral Maxillofac. Pathol.*, vol. 22, no. 2, 2018.

[139] J. Gamper and N. Rajpoot, "Multiple instance captioning: Learning representations from histopathology textbooks and articles," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2021, pp. 16 544–16 554.

[140] W. Qin *et al.*, "What a whole slide image can tell? subtype-guided masked transformer for pathological image captioning," 2023.

[141] Z. Guo *et al.*, "Histgen: Histopathology report generation via local-global feature encoding and cross-modal context interaction," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, 2024, pp. 189–199.

[142] D. Ferber *et al.*, "In-context learning enables multimodal large language models to classify cancer pathology images," *Nat. Commun.*, vol. 15, p. 10104, 2024.

[143] J. Zhou *et al.*, "PathCoT: Chain-of-thought prompting for zero-shot pathology visual reasoning," 2025.

[144] K. Chen *et al.*, "Cost-effective instruction learning for pathology vision and language analysis," *Nat. Comput. Sci.*, vol. 5, no. 7, pp. 524–533, 2025.

[145] J. Xiang *et al.*, "A vision–language foundation model for precision oncology," *Nature*, vol. 638, pp. 769–778, 2025.

[146] S. Sengupta and D. E. Brown, "Automatic report generation for histopathology images using pre-trained vision transformers and bert," 2024.

[147] P. Chen *et al.*, " WsiCaption: Multiple Instance Generation of Pathology Reports for Gigapixel Whole-Slide Images ," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, 2024.

[148] Y. Sun *et al.*, "Pathgen-1.6m: 1.6 million pathology image-text pairs generation through multi-agent collaboration," 2024.

[149] X. Wu *et al.*, "Pathinsight: Instruction tuning of multimodal datasets and models for intelligence assisted diagnosis in histopathology," 2024.

[150] G. Shaikovski *et al.*, "Prism: A multi-modal generative foundation model for slide-level histopathology," 2024.

[151] F. Ahmed *et al.*, "Polypath: Adapting a large multimodal model for multi-slide pathology report generation," *arXiv:2502.10536*, 2025.

[152] R. T. Lucassen *et al.*, "Pathology report generation and multimodal representation learning for cutaneous melanocytic lesions," 2025.

[153] D. Hu *et al.*, "Pathology report generation from whole slide images with knowledge retrieval and multi-level regional feature selection," *Comput. Methods Programs Biomed.*, vol. 263, p. 108677, 2025.

[154] W. Lyu *et al.*, "Efficient whole slide pathology vqa via token compression," 2025.

[155] R. Saluja *et al.*, "Cancer type, stage and prognosis assessment from pathology reports using llms," 2025.

[156] M. Tsuneki and F. Kanavati, "Inference of captions from histopathological patches," in *Proc. Int. Conf. Med. Imag. Deep Learn.*, 2022, pp. 1235–1250.

[157] S. Elbedwehy *et al.*, "Enhanced descriptive captioning model for histopathological patches," *Multimed. Tools Appl.*, vol. 83, pp. 36 645–36 664, 2024.

[158] Q. Zhou *et al.*, "Pathm3: A multimodal multi-task multiple instance learning framework for whole slide image classification and captioning," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, 2024, pp. 373–383.

[159] P. Chen *et al.*, "Wsi-vqa: Interpreting whole slide images by generative visual question answering," in *Proc. Eur. Conf. Comput. Vis.*, 2025, pp. 401–417.

[160] X. He *et al.*, "Pathvqa: 30000+ questions for medical visual question answering," 2020.

[161] F. Shahid *et al.*, "Using generative ai to extract structured information from free text pathology reports," *J. Med. Syst.*, vol. 49, p. 36, 2025.

[162] J. B. Balasubramanian *et al.*, "Leveraging large language models for structured information extraction from pathology reports," 2025.

[163] X. Yang *et al.*, "Enhancing doctor-patient communication using large language models for pathology report interpretation," *BMC Med. Inform. Decis. Mak.*, vol. 25, p. 36, 2025.

[164] C. Comiter, *Inference of single cell profiles from histology stains with the Single-Cell omics from Histology Analysis Framework (SCHAF)*. Massachusetts Institute of Technology, 2024.

[165] C. Yang *et al.*, "Past: A multimodal single-cell foundation model for histopathology and spatial transcriptomics in cancer," 2025.

[166] Z. Kong *et al.*, "Spatia: Multimodal model for prediction and generation of spatial cell phenotypes," 2025.

[167] S. Zhu *et al.*, "Diffusion generative modeling for spatially resolved gene expression inference from histology images," 2025.

[168] S. Andani *et al.*, "Histopathology-based protein multiplex generation using deep learning," *Nat. Mach. Intell.*, 2025.

[169] P. L. Ståhl *et al.*, "Visualization and analysis of gene expression in tissue sections by spatial transcriptomics," *Science*, vol. 353, no. 6294, pp. 78–82, 2016.

[170] A. Rao *et al.*, "Exploring tissue architecture using spatial transcriptomics," *Nature*, vol. 596, no. 7871, pp. 211–220, August 2021.

[171] C. Li *et al.*, "Spatial diffusion for cell layout generation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, 2024, pp. 481–491.

[172] M. Xu *et al.*, "Topocellgen: Generating histopathology cell topology with a diffusion model," 2025.

[173] J. Zhou *et al.*, "Damm-diffusion: Learning divergence-aware multi-modal diffusion model for nanoparticles distribution prediction," 2025.

[174] C. Yin *et al.*, "Prompting vision foundation models for pathology image analysis," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2024, pp. 11 292–11 301.

[175] A. T. Nguyen *et al.*, "Towards a text-based quantitative and explainable histopathology image analysis," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, 2024, pp. 514–526.

[176] Q. Zhou *et al.*, "Mllm4pue: Toward universal embeddings in digital pathology through multimodal llms," 2025.

[177] Z. Huang *et al.*, "A visual–language foundation model for pathology image analysis using medical twitter," *Nat. Med.*, vol. 29, pp. 2307–2316, 2023.

[178] K. Tang *et al.*, "Promptable representation distribution learning and data augmentation for gigapixel histopathology wsi analysis," *Proc. AAAI Conf. Artif. Intell.*, vol. 39, pp. 7247–7256, 2025.

[179] Z. Shao *et al.*, "Augdiff: Diffusion-based feature augmentation for multiple instance learning in whole slide image," *IEEE Trans. Artif. Intell.*, vol. 5, pp. 6617–6628, 2024.

[180] J. Fan *et al.*, "Dcdiff: Dual-granularity cooperative diffusion models for pathology image analysis," *IEEE Trans. Med. Imag.*, vol. 43, pp. 4393–4403, 2024.

[181] H. Xu *et al.*, "A whole-slide foundation model for digital pathology from real-world data," *Nature*, vol. 630, pp. 181–188, 2024.

[182] J. Ma *et al.*, "Towards a generalizable pathology foundation model via unified knowledge distillation," 2025.

[183] A.-T. Nguyen *et al.*, "Mgpath: Vision-language model with multi-granular prompt learning for few-shot wsi classification," 2025.

[184] M. Sturm *et al.*, "Syncellfactory: Generative data augmentation for cell tracking," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, 2024, pp. 304–313.

[185] R. Bruch *et al.*, "Improving 3d deep learning segmentation with biophysically motivated cell synthesis," 2024.

[186] M. Madeira *et al.*, "Tertiary lymphoid structures generation through graph-based diffusion," in *Graphs in BioMed. Image Anal., and Overlapped Cell on Tissue Dataset for Histopathology*, 2024, pp. 37–53.

[187] B. Ehteshami Bejnordi *et al.*, "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," *JAMA*, vol. 318, pp. 2199–2210, 2017.

[188] S. Graham *et al.*, "CoNIC Challenge: Pushing the frontiers of nuclear detection, segmentation, classification and counting," *Med. Image Anal.*, vol. 92, p. 103047, 2024.

[189] ——, "Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images," *Med. Image Anal.*, vol. 58, p. 101563, 2019.

[190] Q. Da *et al.*, "Digestpath: A benchmark dataset with challenge review for the pathological detection and segmentation of digestive-system," *Med. Image Anal.*, vol. 80, p. 102485, 2022.
[191] K. Sirinukunwattana *et al.*, "Gland segmentation in colon histology images: The glas challenge contest," 2016.
[192] S. Graham *et al.*, "Lizard: A large-scale dataset for colonic nuclear instance segmentation and classification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 684–693.
[193] N. Kumar *et al.*, "A multi-organ nucleus segmentation challenge," *IEEE Trans. Med. Imag.*, vol. 39, pp. 1380–1391, 2020.
[194] J. Gamper *et al.*, "Pannuke: an open pan-cancer histology dataset for nuclei instance segmentation and classification," in *Proc. Eur. Congr. Digit. Pathol.*, 2019, pp. 11–19.
[195] M. Veta *et al.*, "Predicting breast tumor proliferation from whole-slide images: The tupac16 challenge," *Med. Image Anal.*, vol. 54, pp. 111–121, 2019.
[196] A. Polónia *et al.*, "Bach dataset : Grand challenge on breast cancer histology images," May 2019. [Online]. Available: https://doi.org/10.5281/zenodo.3632035
[197] F. A. Spanhol *et al.*, "A dataset for breast cancer histopathological image classification," *IEEE Trans. Biomed. Eng.*, vol. 63, pp. 1455–1462, 2016.
[198] P. Bándi *et al.*, "From detection of individual metastases to classification of lymph node status at the patient level: The camelyon17 challenge," *IEEE Trans. Med. Imag.*, vol. 38, pp. 550–560, 2019.
[199] A. A. Borkowski *et al.*, "Lc25000 lung and colon histopathological image dataset."
[200] J. N. Kather *et al.*, "100,000 histological images of human colorectal cancer and healthy tissue," Apr. 2018. [Online]. Available: https://doi.org/10.5281/zenodo.1214456
[201] W. Bulten *et al.*, "Artificial intelligence for diagnosis and gleason grading of prostate cancer: the panda challenge," *Nat. Med.*, 2022.
[202] B. S. Veeling *et al.*, "Rotation equivariant cnns for digital pathology," Sep. 2018. [Online]. Available: https://doi.org/10.1007/978-3-030-00934-2_24
[203] P. Weitz *et al.*, "Acrobat – a multi-stain breast cancer histological whole-slide-image data set from routine diagnostics for computational pathology," 2022.
[204] S. Liu *et al.*, "Bci: Breast cancer immunohistochemical image generation through pyramid pix2pix," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog. Workshops*, 2022, pp. 1815–1824.
[205] S. Karkampouna and M. Kruithof-de Julio, "Dataset empact tma," 2023. [Online]. Available: https://doi.org/10.5281/zenodo.10066853
[206] J. Gamper and N. Rajpoot, "Multiple instance captioning: Learning representations from histopathology textbooks and articles," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2021.
[207] G. Jaume *et al.*, "Hest-1k: A dataset for spatial transcriptomics and histology image analysis," *Proc. Adv. Neural Inf. Process. Syst.*, vol. 37, pp. 53 798–53 833, 2024.
[208] Y. Sun *et al.*, "Pathasst: A generative foundation ai assistant towards artificial general intelligence of pathology," *Proc. AAAI Conf. Artif. Intell.*, vol. 38, pp. 5034–5042, 2024.
[209] M. Tsuneki and F. Kanavati, "Patchgastricadc22," Dec. 2021. [Online]. Available: https://doi.org/10.5281/zenodo.6021442
[210] W. O. Ikezogwo *et al.*, "Quilt-1m: One million image-text pairs for histopathology," 2025.
[211] K. Ding *et al.*, "A large-scale synthetic pathological dataset for deep learning-enabled segmentation of breast cancer," *Sci. Data*, vol. 10, no. 1, p. 231, 2023.
[212] Y. Sun *et al.*, "Pathmmu: A massive multimodal expert-level benchmark for understanding and reasoning in pathology," 2024.
[213] X. Zhang *et al.*, "Pmc-vqa: Visual instruction tuning for medical visual question answering," 2024.
[214] M. M. Saad *et al.*, "A survey on training challenges in generative adversarial networks for biomed. image anal." *Artif. Intell. Rev.*, vol. 57, no. 2, p. 19, 2024.
[215] Z. Ji *et al.*, "Survey of hallucination in natural language generation," *ACM Comput. Surv.*, vol. 55, no. 12, pp. 1–38, 2023.
[216] S. P. Deshpande, "Generative AI for computational pathology," Ph.D. dissertation, Univ. of Warwick, Coventry, U.K., Aug. 2023.
[217] Z. Cai *et al.*, "Generative adversarial networks: A survey toward private and secure applications," *ACM Comput. Surv.*, vol. 54, no. 6, pp. 1–38, 2021.
[218] Y. Deo *et al.*, "Metrics that matter: Evaluating image quality metrics for medical image generation," *arXiv:2505.07175*, 2025.
[219] H. Cao *et al.*, "A survey on generative diffusion models," *IEEE transactions on knowledge and data engineering*, vol. 36, no. 7, pp. 2814–2830, 2024.
[220] M. Matheny *et al.*, *Artificial intelligence in health care: The hope, the hype, the promise, the peril*. National Academies Press, 2022, vol. 2019.
[221] S. Benjamens *et al.*, "The state of artificial intelligence-based fda-approved medical devices and algorithms: an online database," *NPJ Digit. Med.*, vol. 3, no. 1, p. 118, 2020.
[222] K. Drukker *et al.*, "Toward fairness in artificial intelligence for med. image anal.: identification and mitigation of potential biases in the roadmap from data collection to model deployment," *J. Med. Imag.*, vol. 10, no. 6, pp. 061 104–061 104, 2023.
[223] J. Kaplan *et al.*, "Scaling laws for neural language models," *arXiv:2001.08361*, 2020.
[224] J.-Y. Zhu *et al.*, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2223–2232.
[225] T. Besiroglu *et al.*, "The compute divide in machine learning: A threat to academic contribution and scrutiny?" *arXiv:2401.02452*, 2024.
[226] O. E. Gundersen and S. Kjensmo, "State of the art: Reproducibility in artificial intelligence," in *Proc. AAAI Conf. Artif. Intell.*, vol. 32, no. 1, 2018.
[227] S. Hao *et al.*, "Synthetic data in ai: Challenges, applications, and ethical implications," *arXiv:2401.01629*, 2024.
[228] E. Zimmermann *et al.*, "Virchow2: Scaling self-supervised mixed magnification models in pathology," 2024.
[229] Z. Durante *et al.*, "Agent ai: Surveying the horizons of multimodal interaction," *arXiv:2401.03568*, 2024.
[230] C. Bunne *et al.*, "How to build the virtual cell with artificial intelligence: Priorities and opportunities," *Cell*, vol. 187, no. 25, pp. 7045–7063, 2024.
[231] J. E. Rood *et al.*, "The human cell atlas from a cell census to a unified foundation model," *Nature*, vol. 637, no. 8048, pp. 1065–1071, 2025.
[232] P. Zhang *et al.*, "Systematic inference of super-resolution cell spatial profiles from histology images," *Nat. Commun.*, vol. 16, no. 1, p. 1838, 2025.
[233] H. P. Foote *et al.*, "Embracing generative artificial intelligence in clinical research and beyond: Opportunities, challenges, and solutions," *JACC: Adv.*, vol. 4, no. 3, p. 101593, 2025.
[234] F. Liao *et al.*, "Governance of clinical ai applications to facilitate safe and equitable deployment in a large health system: key elements and early successes," *Front. Digit. Health*, vol. 4, p. 931439, 2022.