Neurocomputing 676 (2026) 133064

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Survey paper

A systematic review of machine learning for digital stain processing in pathology

Rabiah Al-Qudah [a], Abubakar Bala [b,*], Mrouj Almuhajri [c], Khiati Zakaria [a], Ching Y. Suen [d]

[a] Abu Dhabi University, College of Engineering, Department of Computer Science and Information Technology, P.O. Box 59911, Abu Dhabi, United Arab Emirates
[b] Monash University Malaysia, School of Business, Jalan Lagoon Selatan, 47500 Bandar Sunway, Selangor, Malaysia
[c] Saudi Electronic University, College of Computing and Informatics, Riyadh 13316, P.O. Box 11673, Kingdom of Saudi Arabia
[d] Concordia University, Gina Cody School of Engineering and Computer Science, 1455 Blvd. De Maisonneuve Ouest, Montreal, Quebec H3G 1M8, Canada

ARTICLE INFO

Communicated by P. Yan

Keywords:
Digital pathology
Generative adversarial networks
Machine learning
Survey
Virtual staining

ABSTRACT

Digital staining involves using methods such as Machine Learning (ML) to replace chemical staining in pathology. Staining adds contrast that makes cell details more visible under the microscope. However, chemical methods are slow, use toxic reagents, and require skilled personnel. In contrast, digital staining can generate images faster, reduce the need for reagents and specialized equipment, and minimize plastic and chemical waste, making the workflow more sustainable. This paper systematically reviews papers published on ML-based digital stain processing. We propose a new taxonomy that divides existing studies into five groups: stain normalization, stain augmentation, virtual staining, stain transformation, and hybrid approaches. In addition, we observed several trends from the reviewed papers. Finally, we outline open research directions.

1. Introduction

Digital staining enables the virtual application or conversion of stains on histological images without requiring physical reagents. This technique has gained traction recently due to advances in generative models. These models can learn complex mappings between unstained and stained tissue images.

In traditional workflows, staining involves a multistep manual process that includes fixation, embedding, sectioning, chemical staining, and microscopic review. This process is time-consuming, relies on skilled personnel, and involves toxic reagents that pose health and environmental risks. In addition, applying multiple stains requires additional tissue sections, increasing resource use and patient discomfort. Digital staining has the potential to reduce these limitations by enabling faster generation of clinically useful images, which can shorten diagnostic turnaround times and support more efficient laboratory operations. It also reduces the dependence on chemical reagents, offering a more eco-friendly workflow that minimizes waste and reduces exposure to hazardous materials. Additionally, because digital staining involves less repeated physical sectioning, it can reduce patient discomfort. Furthermore, by decreasing the number of consumables and the labour required for manual staining, digital staining can contribute to meaningful cost savings.

Digital staining offers a faster, non-destructive, and more sustainable alternative. Here, Machine Learning (ML)-based models are trained to produce stain-equivalent images in seconds and support downstream diagnostic tasks such as classification and segmentation. These benefits position digital staining as a tool for modern pathology, especially in places with limited resources. Most current virtual staining systems are based on Generative Adversarial Networks (GANs). A GAN pairs two neural networks: a generator that produces realistic images and a discriminator that tries to tell generated images from real ones. During training, they compete, and the generator gradually learns to create images that the discriminator can no longer distinguish from authentic stained slides. Section 2.2 reviews the GAN variants most widely used for digital staining.

1.1. Research questions, search strings, and inclusion criteria

In this section, we define how we discovered and filtered the relevant literature for the survey.

1.1.1. Research questions

First, we define our research topic: “Machine learning-based digital stain processing for digital pathology.” We then developed research

* Corresponding author.
Email address: abala.ele@buk.edu.ng; abubakar.bala@monash.edu (A. Bala).

https://doi.org/10.1016/j.neucom.2026.133064
Received 18 September 2025; Received in revised form 10 December 2025; Accepted 12 February 2026
Available online 16 February 2026

Table 1
Research questions according to PRISMA.

#	Research question
RQ1	How do machine learning-based digital stain processing systems for digital pathology compare to traditional chemical methods regarding diagnostic time and accuracy?
RQ2	What machine learning models have been used for digital stain processing in digital pathology and how do their performances compare?
RQ3	What are the challenges of applying machine learning algorithms to process stains in digital pathology?

questions from this topic by following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [1]. These research questions are in Table 1.

1.1.2. Search strings

We identified key concepts from the research topic and the questions in the previous section. These concepts are shown in Table 2. From the table, it can be observed that several alternative spellings and synonyms of keywords are also used in the search.

From Table 2, the following search was performed on Scopus: TITLE-ABS-KEY((virtual OR digital OR automated) AND (staining OR dyeing OR re-staining OR restaining) AND (gan* OR "Generative adversarial network*" OR "Machine learning" OR ml OR dl OR "deep learning" OR ai OR "artificial intelligence" OR "Generative AI" OR "Generative artificial intelligence") AND (pathology OR histopathology OR biopsy OR smear)).

The first search was conducted on 18 November 2024. We were able to find 677 documents. Fig. 1 shows the number of files per year in the initial Scopus search. The figure shows how interest in the topic has increased over the years, especially after 2020. However, some of these articles were not relevant to our study. Thus, we created a filtering process to select only suitable studies.

1.1.3. Paper inclusion criteria

For this systematic review, we defined explicit inclusion criteria for selecting papers. The flow chart in Fig. 2 summarizes the process. Before screening, we removed duplicate or retracted papers.

In the screening phase, we applied additional criteria, such as restricting publication years to 2020–2025 and including only articles that met these conditions. Following Fig. 2, this reduced the initial 677 papers to about 100.

Table 2
Survey main topics and keywords.

#	Concept	Keywords
1	Artificial Intelligence	Artificial intelligence, Computer vision, Deep learning, Machine learning, Generative adversarial network, AI, ML, DL, GAN
2	Digital Staining	Virtual staining, automated staining, virtual re-staining, digital re-staining, virtual dyeing, digital dyeing, automated dyeing
3	Pathology	Histopathology, smear, biopsy

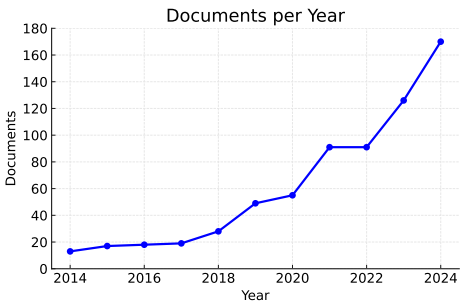


Fig. 1. The number of articles per year discovered in the initial Scopus search.

1.2. Related reviews

This section summarizes existing surveys on ML-based digital staining and explains how our review extends this literature. We repeated the search from Section 1.1.2 but limited it to review articles, obtaining 41 results. After screening, we identified seven relevant reviews and later added several recent ones during the paper revision.

One of such reviews is the work of Rivenson et al. [2]. The study highlighted that early methods of virtual staining used imaging modalities like multi-modal nonlinear microscopy to capture tissue contrast. However, they were limited to pixel-to-pixel mappings that often failed to capture the textures needed for accurate diagnostics. In contrast, CNN-based methods, such as GANs, solve this problem. They perform texture-based transformations by learning complex structural relationships between label-free and stained tissues. However, GAN-based staining has limitations, including the risk of misleading artifacts. The article suggests combining adversarial loss with structural or pixel-wise loss so that the network is constrained by both overall appearance and actual tissue features, reducing hallucinations.

In a related work, Jose et al. [3] reviewed how different GAN variants (standard GANs, cGANs, and CycleGANs) are used in digital pathology. They report that GANs improve image consistency by addressing staining variability, enhancing resolution, supporting virtual staining, removing artifacts such as ink marks, generating synthetic datasets to reduce labeled-data scarcity, and aiding nuclei detection, segmentation, and feature extraction.

Additionally, Morrison et al. [4] provided a survey of how generative deep learning transforms workflows in digital pathology. The survey states that a typical machine learning workflow in digital pathology includes several steps: data collection, preprocessing, model training, and interpretation. Each of these steps presents its own challenges. For example, before model training, images must be preprocessed to remove artifacts through stain normalization and data augmentation. Other issues include the large size of WSIs and the scarcity of annotated data. Generative models address some of these challenges through various applications. For instance, using GANs for color and intensity normalization reduces inconsistencies caused by staining protocols.

Furthermore, Pillar et al. [5] examine the use of machine learning for virtual tissue staining in pathology. They employ deep learning models to transform autofluorescence images of unstained tissue into virtually stained equivalents. The study describes how deep learning converts unstained images into bright-field images that mimic conventional stained slides.

In a related review, Waqas et al. [6] presented how generative AI and foundation models can revolutionize digital pathology. The article highlights the limitations of task-specific AI commonly used in computational pathology. Task-specific AI models, like CNNs, excel in specific tasks but struggle with generalization across different datasets. Furthermore, they require extensive annotated data and cannot use multimodal data (e.g., text).

In contrast to task-specific AI, foundation models, which are the backbone of generative AI, can solve some of these limitations. Foundation models are pre-trained on large, unannotated datasets and then adapted to tasks with minimal task-specific data. This allows them to handle multimodal data. These models can help standardize pathology workflows and help in diagnosis and report generation. In addition, they can also self-correct based on user feedback.

One challenge of this paper is that it covers the broad area of the use of generative AI in digital pathology. As a result, it struggles to provide specific challenges in applying these models to particular niches like digital staining. In addition, it does not present any detailed technical challenges that may be encountered during model training or deployment.

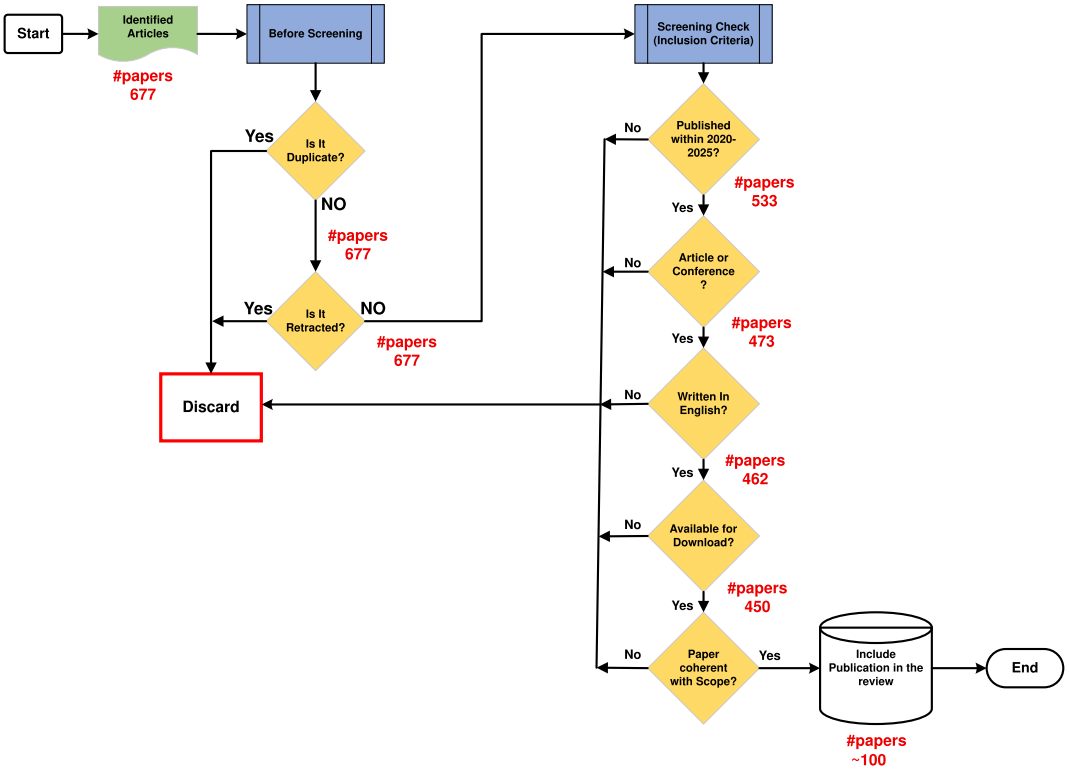


Fig. 2. A flowchart showing the PRISMA rules used in the relevant literature search.

Similarly, Bai et al. [7] reviewed works that apply deep learning for virtual staining. After highlighting the advantages of virtual staining over the traditional process, they categorize it into two: label-free and stain-to-stain transformation. They then go on to state that the ML models used can either be supervised or unsupervised methods. An interesting area the survey presented is virtual staining in vivo. This is where label-free imaging eliminates the need for invasive biopsies. Techniques like reflectance confocal microscopy allow virtual staining directly on live tissues, which enables quicker diagnoses without physical sample preparation.

Moreover, Kreiss et al. [8] provide a broad review of digital staining in optical microscopy, with a scope spanning many label-free imaging modalities, stain targets, and deep learning models. Their review treats digital staining as a single integrated field and explains both the underlying physical contrast mechanisms and the neural architectures. The paper also addresses practical issues such as data pairing, registration, metrics, and artifacts.

In contrast, our paper focuses on digital stain processing in pathology and presents a systematic review of recent work. This targeted scope allows us to define a clear pathology-specific taxonomy of tasks and to better compare existing methods.

Similarly, Latonen et al. [9] reviewed works that use deep learning for virtual staining. They highlighted that the methods mainly rely on Generative Adversarial Networks (GANs) to transform images. They classified the learning process into two categories: supervised and unsupervised methods. Supervised methods like Pix2Pix and cGANs rely on paired stained and unstained images. In contrast, unsupervised methods, like CycleGAN, enable style transfer without specific image matching.

One downside of the study is that it does not discuss the models used in detail nor does it present any technical challenges faced by model developers.

In a more recent review, Lin et al. [10] review virtual staining in pathology with a focus on the full workflow, typical models, available datasets, and evaluation methods. They analyze challenges that are specific to virtual staining, such as gigapixel whole slide images, feature homogeneity, and invisible diagnostic cues, and explain how these differ from generic image translation problems. They also link algorithmic design to clinical needs by writing from both engineering and pathology perspectives.

Table 3 compares our review with other surveys. From the table, it can be observed that our paper distinguishes itself from other surveys. Our main contributions can be summarized as follows:

- **Compared to other works, our paper focuses solely on stain processing, not the broader area of digital pathology. Thus, we are able to provide a more niche study.**
- **Additionally, unlike other review articles that restrict themselves to GANs or foundational models [3,6], this article includes other ML methods such as DenseNet, StainCUT, and RestainNet [11–13] together with GANs. This makes our work more encompassing.**
- **We also present a new taxonomy of the existing literature based on tasks being performed. To the best of our knowledge, we are the first to propose such a taxonomy.**
- **This review also presents trends in existing research. We also highlight open research areas that only a few researchers have explored.**
- **We also discuss some of the common benchmark datasets used to evaluate models.**
- **Finally, it is an up-to-date survey of machine learning-based digital stain processing that covers articles published between 2020 and 2025.**

The rest of the paper is organized as follows. Section 2 presents the analysis of studies that employ machine learning methods for digital stain processing. It also discusses the taxonomy of the field, commonly used models, and performance evaluation metrics. Section 3 outlines the widely used datasets and key trends. Section 4 presents open research issues. Finally, Section 5 concludes the paper.

Table 3
Summary of machine learning-based digital staining review papers.

Review	Focus area	Review type
Rivenson et al. 2020 [2]	Presented works that improve digital staining. Compared virtual staining process using multimodal imaging and that of deep learning. Also stated some of the challenges and open issues of using DL for digital staining.	Narrative
Jose et al. 2021 [3]	survey works that use GANs for DP. Presented areas where GANs are applied and their advantages. Presented limited challenge and research gaps. Focus is on the wide area of digital pathology	Systematic
Morrison et al. 2021 [4]	Surveyed works that aim to improve DP workflows. Presented challenges and future research areas. Focus is on many areas of DP.	Narrative
Pillar et al. 2022 [5]	Discussed the advantages of virtual staining over conventional methods. Presented challenges and future research areas	Narrative
Kreiss et al. 2023 [8]	Broad tutorial review of digital staining across label-free modalities, stains and deep learning models, including mechanisms and practical deployment. Did not develop a taxonomy.	Narrative
Bai et al. 2023 [7]	Discussed works that use DL for virtual staining. Categorized the methods. Also discussed in vivo staining. Presented challenges and research gaps.	Narrative
Waqas et al. 2023 [6]	Presented works that use foundational models for digital pathology. Highlights the shift from task-specific AI models to foundation models that use multimodal data. Presented challenges and future directions. Focus was on the wide topic of digital pathology	Narrative
Latonen et al. 2024 [9]	Reviewed works that use GANs for virtual staining. Presented a rough taxonomy and presented research gaps. Did not provide many technical details of the models.	Narrative
Lin et al. 2025 [10]	Covers workflows, typical models, datasets, evaluation methods, and challenges specific to whole slide images.	Narrative
Our study	**Focused on digital stain processing not downstream tasks. Covers entire ML-based methods, not only GANs. Provided a taxonomy of the area. Discussed famous benchmark datasets. Presented a comprehensive trend of the field. Outlined updated research gaps.**	**Systematic**

2. Machine learning methods for stain processing in pathology

This section reviews ML methods for stain processing in pathology, introducing a task-based taxonomy, the most commonly used models, their evaluation metrics, and an analysis of work in each category.

2.1. Taxonomy

In analyzing studies that use machine learning to address staining issues in pathology, we first categorize them based on the tasks they perform, as shown in Fig. 3. A summary table of the classes can also be seen in Table 4. The first group is stain normalization. In digital pathology, stain normalization is needed due to chemical stain variations across laboratories or devices. These inconsistencies often affect the analysis performed by deep learning models. Thus, papers in this class use DL to standardize these variations and help models focus on tissue structures rather than color differences. This improves accuracy and generalization across datasets.

The second category we identified is stain augmentation using ML. Although stain augmentation and stain normalization aim to address the problem of stain variability, they differ in their objectives and implementation. Normalization seeks to reduce variability by mapping all images toward a predefined stain appearance. Most normalization methods rely on a single reference image or a set of reference stain statistics, and all incoming images are standardized to match this reference. This process compresses the stain distribution into a narrower domain and ensures consistency across datasets.

In contrast, stain augmentation increases variability by generating additional synthetic images that capture a wide range of stain appearances. Instead of mapping images toward a fixed target, augmentation models create multiple variants of an input image to simulate natural diversity.

Another key distinction is where these methods are applied. Normalization is commonly used during both training and inference because its goal is to standardize all inputs before downstream analysis. In contrast, augmentation is typically applied only during training because its purpose is to expand the diversity of the training distribution and improve generalization.

The third class we identified is virtual staining. Here, label-free (unstained images) are stained using DL models. Most of the works in this group take a label-free image and convert it into a stained version one-to-one. However, some models can also simultaneously transform an unstained image into multiple stains.

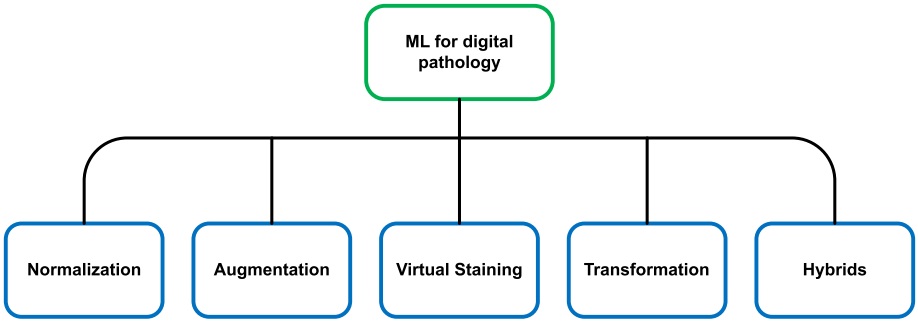


Fig. 3. Classification of works on ML-based stain processing in pathology.

Table 4
Summary of the proposed taxonomy for machine learning-based digital stain processing.

Category	Main goal	Typical input and output	Typical use case
Stain normalization	Reduce inter-laboratory and inter-scanner variation in stain appearance	Input: stained histopathology image with heterogeneous color characteristics. Output: stained image with standardized stain appearance while preserving tissue morphology.	Harmonizing multi-center datasets; improving robustness of downstream models.
Stain augmentation	Increase stain variability during model training	Input: stained image. Output: multiple synthetically re-stained variants of the same tissue.	Exposing models to diverse stain appearances to improve generalization.
Virtual staining	Generate stained images from label-free or unstained acquisitions	Input: label-free (e.g., autofluorescence or phase-contrast) image. Output: virtually stained image (e.g., H&E, special stain, or IHC).	Enabling non-destructive, faster, or cheaper staining workflows.
Stain transformation	Convert one stain type into another	Input: image with an existing stain (e.g., H&E). Output: image with a different target stain (e.g., PAS, IHC).	Reducing the need for additional physical staining of tissue sections.
Hybrid approaches	Combine two or more of the above categories in a single pipeline	Input and output depend on the combination (e.g., normalization plus augmentation, or virtual staining followed by transformation).	Addressing multiple stain-related challenges within a unified framework.

The fourth group consists of those that perform stain transformation. The works here transform a stained image to another type of stain, e.g., an H&E-stained image to PAS. Similar to virtual staining, some studies here also perform multi-stain transformation, where a stained image is converted to multiple stains.

In the fifth class, called hybrid, studies here perform two or more of the tasks simultaneously, such as normalization and augmentation, using DL.

Before the detailed discussion of studies within each of these categories, the next section provides a brief overview of some commonly used GAN models.

2.2. Commonly used models

This section explains three popular models commonly used for stain processing in digital pathology: Conditional Generative Adversarial Network (cGAN), Pix2Pix, and CycleGAN.

2.2.1. Conditional generative adversarial network (cGAN)

cGANs [14] improve basic GANs (see Fig. 4(a)) by adding control over the generated output. In a standard GAN, the generator creates data from random noise. This often leads to limited diversity, where the generated output becomes too similar or lacks variation in important features. As a result, the model is less adaptable for tasks that require specific or varied output. In contrast, cGANs shown in Fig. 4(b) condition both the generator and the discriminator on additional information, such as class labels. This allows for more control over the generation process and makes cGANs useful for tasks like targeted image synthesis.

To understand its formulation, consider labeled samples (x, c) and a noise vector z. The generator G maps (z, c) to $\hat{x} = G(z, c)$. The discriminator $D(x, c) \in (0, 1)$ scores image–label pairs.

Conditional adversarial objective:

$$\mathcal{L}_{\text{cGAN}}(G, D) = \mathbb{E}_{x,c}\left[\log D(x, c)\right] + \mathbb{E}_{z,c}\left[\log\left(1 - D\left(G(z, c),\, c\right)\right)\right]. \tag{1}$$

Thus, G tries to minimize this, while D tries to maximize it.
Overall objective:

$$\arg\min_{G}\ \max_{D}\ \mathcal{L}_{\text{cGAN}}(G, D). \tag{2}$$

Despite their advantages, using basic cGANs for image generation has its challenges. First, they tend to generate similar images for the same condition, limiting the output diversity. Secondly, they require significant computational resources, especially when generating high-resolution images. This makes them resource-intensive and challenging to scale compared to Pix2Pix.

2.2.2. Pix2Pix

Pix2Pix is a type of cGAN used for image-to-image translation [15]. As shown in Fig. 4(c), it learns the mapping between paired images through a generator and a discriminator. In basic cGAN, the "condition" can be text, labels, or images, while in Pix2Pix, it is strictly images. The generator, which is U-Net-based [16], transforms the input images into the output images. In contrast, the discriminator, which is PatchGAN [17] evaluates the authenticity of the generated images. Furthermore, instead of using noise as in cGAN, Pix2Pix uses dropouts in the generator to simulate noise.

Given an input image x and a target image y, the model learns a mapping with a generator $G : \{x, z\} \rightarrow \hat{y}$ while a discriminator D looks at pairs and judges real vs. fake. Both G and D see the input image x, so D compares real tuples (x, y) against fake tuples $(x, G(x, z))$.

The main objective is the cGAN loss:

$$\mathcal{L}_{\text{cGAN}}(G, D) = \mathbb{E}_{x,y}[\log D(x, y)] + \mathbb{E}_{x,z}[\log(1 - D(x, G(x, z)))] \tag{3}$$

Pix2Pix employs a pixel reconstruction loss to keep the output near the ground truth:

$$\mathcal{L}_1(G) = \mathbb{E}_{x,y,z}\left[\|y - G(x, z)\|_1\right] \tag{4}$$

The final Pix2Pix objective is:

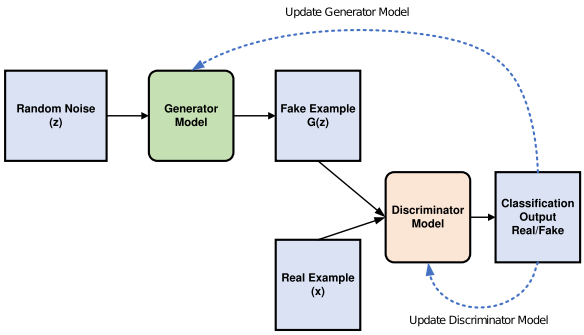
$$\arg\min_{G}\max_{D}\ \mathcal{L}_{\text{cGAN}}(G, D)\ +\ \lambda\,\mathcal{L}_1(G) \tag{5}$$

During training, Pix2Pix minimizes a loss function combining adversarial and pixel-wise losses to produce high-quality outputs. After training is finished, at the inference stage, the model takes an input image, applies the learned generator, and transforms it into the corresponding output image.

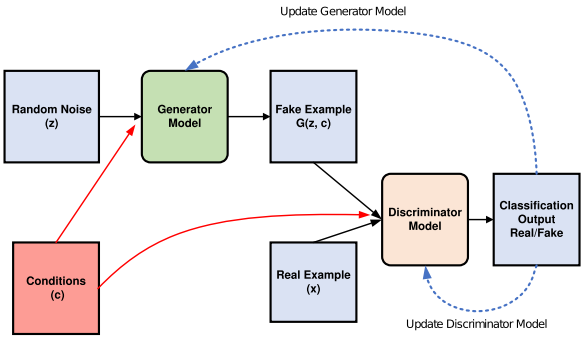
One of the main advantages of Pix2Pix is its ability to generate accurate and high-quality images when paired datasets are available, which is not often the case. This makes it particularly useful for tasks that require precise pixel-level translation, such as digital staining. Compared to basic cGAN, Pix2Pix is more focused on image-to-image translation with paired data, making it easier to implement for tasks where exact image pairs are available. However, Pix2Pix is less flexible than CycleGAN, as it relies heavily on paired data and struggles when there are variations in staining.

2.2.3. CycleGAN

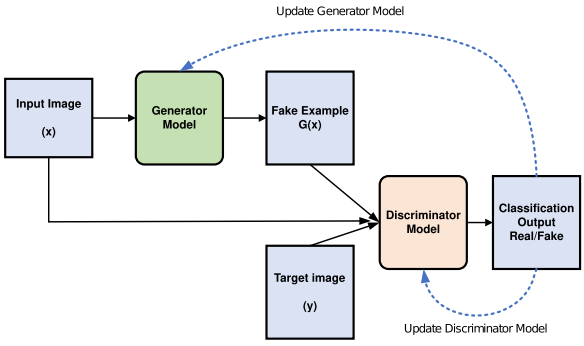
Although Pix2Pix offers good performance, it often requires paired images. However, these paired images are hard to obtain in most applications. In contrast, CycleGAN (see Fig. 4(d)) can perform image-to-image



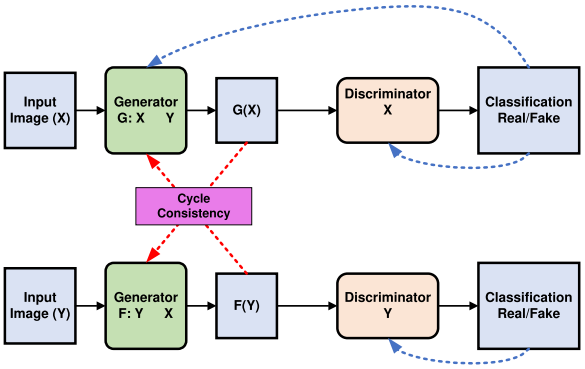
(a) A figure showing the basic GAN.



(b) An image of cGAN with the conditional block shown in pink. Also red arrows show how the condition block connects to both the generator and discriminator.



(c) The Pix2Pix GAN showing how the target image connects to the discriminator



(d) The CycleGAN. It operates without paired images. The figure shows the cycle consistency process between the two generators.

Fig. 4. The figures of the famous GANs used for processing digital stains. The images are summarized to aid understanding.

translations without the need for paired images [18]. It uses two generators and two discriminators to learn the mappings between two image domains. One generator converts images from domain X to domain Y, and the other does the reverse. The discriminators judge whether the images look real or fake. The key loss functions are adversarial loss and cycle consistency loss. The cycle consistency loss ensures that when an image is translated to the other domain and back again, it matches the original image.

Given two image domains X and Y, with unpaired samples $x \sim p_{\text{data}}(x)$ and $y \sim p_{\text{data}}(y)$. CycleGAN learns two mappings $G : X \to Y$ and $F : Y \to X$ together with adversarial discriminators D_Y and D_X. The first objective is an adversarial loss that makes the translated images indistinguishable from real images in the target domain. For the mapping $G : X \to Y$ and discriminator D_Y, the objective is

$$\mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) = \mathbb{E}_y\big[\log D_Y(y)\big] + \mathbb{E}_x\big[\log\big(1 - D_Y(G(x))\big)\big] \tag{6}$$

and symmetrically $\mathcal{L}_{\text{GAN}}(F, D_X, Y, X)$ for $F : Y \to X$ against D_X. This pushes $G(X)$ toward the distribution Y and $F(Y)$ toward X.

Adversarial loss aligns distributions at the set level, not per sample. A generator may map many inputs to one plausible target and still appear valid. Cycle consistency adds an inverse-like constraint: translate $X \to Y \to X$ and $Y \to X \to Y$ to recover the original sample. This anchors the mapping to each input, preserves content and structure, and still allows style changes. It also enables training without paired supervision because the reconstruction target is the input itself.

$$\mathcal{L}_{\text{cyc}}(G, F) = \mathbb{E}_x\big[\,\|F(G(x)) - x\|_1\,\big] + \mathbb{E}_y\big[\,\|G(F(y)) - y\|_1\,\big] \tag{7}$$

encouraging $F(G(x)) \approx x$ (forward cycle) and $G(F(y)) \approx y$ (backward cycle).

Combining these yields the full objective

$$\mathcal{L}(G, F, D_X, D_Y) = \mathcal{L}_{\text{GAN}}\big(G, D_Y, X, Y\big) + \mathcal{L}_{\text{GAN}}\big(F, D_X, Y, X\big) + \lambda\,\mathcal{L}_{\text{cyc}}(G, F) \tag{8}$$

where λ balances distribution matching and cycle consistency. The overall optimization is the usual adversarial min–max over both domains:

$$\arg\min_{G,F}\ \max_{D_X,D_Y}\ \mathcal{L}(G, F, D_X, D_Y) \tag{9}$$

Training alternates updates of G, F and D_Y, D_X on unpaired mini-batches from X and Y under the adversarial and cycle-consistency objectives. At inference, the discriminators are discarded, and translation is performed with the relevant generator ($G : X \to Y$ or $F : Y \to X$) to produce the transformed image for a given input.

2.3. Performance metrics

This section presents the performance measures commonly used to judge how well a model performs.

1. Structural Similarity Index Measure (SSIM): A metric used to assess the perceptual similarity between two images (x, y). In our case, it is often between the virtual stain generated and the actual stain. Its equation is given by:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{10}$$

where μ_x and μ_y are the mean intensities, σ_x^2 and σ_y^2 are the variances and σ_{xy} is the covariance of the images, with c_1 and c_2 as

stabilizing constants. A high SSIM score (close to 1) indicates that images are perceptually similar, while a low score (close to 0) suggests significant structural differences. SSIM is widely used due to its alignment with human visual perception. However, it can be sensitive to small variations and is computationally complex compared to simpler metrics such as PSNR.

2. Peak Signal-to-Noise Ratio (PSNR): Another common metric used to evaluate the quality of virtual staining methods. It compares the generated stained images with the ground-truth stained images. PSNR measures the relationship between the maximum possible pixel value of the image and the noise introduced during the image generation process.

$$PSNR = 10\log_{10}\left(\frac{MAX_I^2}{MSE}\right) \tag{11}$$

where MAX_I represents the maximum pixel value (typically 255 for 8-bit images) and MSE is the mean square error, which calculates the average squared differences between the generated and real images. A higher PSNR indicates better image quality, implying that the generated image is closer to the ground truth with minimal distortion. However, while PSNR is a widely used metric for its simplicity and computational efficiency, it has limitations in assessing perceptual quality. Thus, PSNR focuses on pixel-level differences and may not always align with how human observers perceive the quality of stained tissue images.

3. Fréchet Inception Distance (FID): used to assess the quality and diversity of generated stained images. It is obtained by comparing their feature distributions with those of real images.

$$FID = \left\|\mu_r - \mu_g\right\|^2 + \mathrm{Tr}\left(\Sigma_r + \Sigma_g - 2\left(\Sigma_r\Sigma_g\right)^{1/2}\right) \tag{12}$$

where μ_r and μ_g are the mean feature vectors for real and generated images, and Σ_r and Σ_g are their covariance matrices. In addition, Tr is the trace of a matrix, which is the sum of its diagonal elements. A lower FID score indicates better image quality, meaning the generated images are more similar to the real ones. Although effective in capturing overall image quality and diversity, FID may not detect local artifacts that could affect diagnostic accuracy.

4. Kappa Score (Cohen's Kappa): A statistical measure used to assess the agreement between two raters or models, adjusting for the chance agreement. In our context of digital staining, it evaluates how closely virtual staining results match real stained images. The Kappa score is calculated as:

$$\kappa = \frac{P_o - P_e}{1 - P_e} \tag{13}$$

where P_o is the observed agreement (the proportion of matching classifications) and P_e is the expected agreement by chance. A high Kappa score (close to 1) indicates strong agreement, while a low score (close to 0) suggests poor agreement. Negative values indicate worse than chance agreement. Kappa is useful for measuring the consistency of virtual staining methods with real results. However, it can be sensitive to imbalanced data and is limited to categorical data.

5. Learned Perceptual Image Patch Similarity (LPIPS): A metric used to evaluate perceptual differences between images based on human visual perception [19]. It measures the similarity between image patches by using deep learning to extract features from pretrained neural networks. In most of the works reviewed here, the network used for computing the LPIPS is VGG, AlexNet, or SqueezeNet, pretrained on the ImageNet dataset. In some cases, the model undergoes further linear calibration on the human perceptual Berkeley-Adobe Perceptual Patch Similarity (BAPPS) dataset [20].

The LPIPS is better at capturing perceptual quality compared to metrics based on pixel-wise differences. It is calculated by comparing the distance between the feature representations of image patches:

$$\mathrm{LPIPS}(x, y) = \frac{1}{N}\sum_{i=1}^{N}\left\|f_i(x) - f_i(y)\right\|_2^2 \tag{14}$$

where $f_i(x)$ and $f_i(y)$ are the feature representations of images x and y at the i-th layer of the neural network, and N is the total number of layers considered.

A lower LPIPS score indicates higher perceptual similarity between the images, meaning they appear more alike to human observers. Similar to SSIM, LPIPS aligns with human perception. On the other hand, it requires more computational resources than simple pixel-based metrics, and it may not always detect small, noticeable image distortions at a pixel level.

Although the common image similarity metrics mentioned here are useful for assessing how closely generated stains resemble reference images, they do not always correlate with diagnostic performance. For this reason, many studies complement image-based measures with downstream task evaluation, such as classification AUC for disease prediction or subtype identification, and segmentation Dice scores for detecting structures like nuclei, glands or lesions. This type of evaluation is especially important in stain normalization and augmentation, where matched target stains are often unavailable and image-level metrics alone provide limited insight into whether the modified stains improve model performance. Some virtual staining and stain-to-stain transformation works also report downstream results to confirm that visual similarity gains translate into meaningful diagnostic improvements. Several studies further employ expert or pathologist assessment to capture clinical relevance, sometimes quantified using measures such as SSIM for image similarity or Cohen's kappa for agreement.

2.4. Stain normalization using DL

As mentioned earlier, stain normalization helps the models to generalize to unseen data. This section presents works that employ DL models to perform stain normalization.

For example, Ziaei et al. [21] presented a study comparing different color normalization methods. They applied these techniques to a set of images scanned with two different scanners. The effectiveness of each method was then tested by comparing the normalized colors with the original scans using the Commission Internationale de l'Éclairage (CIE) color difference metric.

They tested the models on patch images of breast tissue scanned using two different scanners. The study found that while other color normalization methods showed promise, StyleGAN outperformed them.

Similarly, Salehi et al. [22] developed a Pix2Pix-based Stain-to-Stain Translation (STST) for stain normalization. The authors pointed out that most traditional normalization methods rely on a single reference image and may distort tissue structures. To address this, they proposed STST, a Pix2Pix-based model trained on paired grayscale and color-stained patches, which learns the mapping directly without a reference image.

The method was tested on the MITOS-ATYPIA-14 dataset. The STST approach outperformed traditional methods such as Reinhard and Macenko.

In a related work, Zheng et al. [23] introduced the Stain Standardization Capsule (SSC) for stain normalization. The SSC is trained jointly with a CNN to minimize a combined loss function. This includes a reconstruction loss that compares the reconstructed image to the ground truth.

The study used publicly available datasets such as CAMELYON16, ACDC-LungHP, and a gastric dataset. Experiments showed that SSC improved classification accuracy and maintained computational efficiency due to its lightweight design.

In a similar work, Patil et al. [11] proposed a method to address the challenges of color variations in H&E-stained histopathology images. The paper stated that normalization methods based on GANs are often computationally expensive. Their technique is based on a self-supervised, CNN-based model named ColorNormNet. The training process involved generating synthetic source images by adding offsets to the target images, which helps to avoid artifacts and improve speed. Additionally, the ColorNormNet relies on DenseNet-based blocks with few parameters. This makes it easier to integrate the model into deep learning pipelines.

They used two datasets, CAMELYON17 and MoNuSeg. The results showed that ColorNormNet improved classification and segmentation tasks, outperforming previous methods in accuracy and processing time.

Similarly, Runz et al. [24] investigated using CycleGANs to address variations in H&E-stained histological images. They used CycleGANs to map images between domains while ensuring that reconstructed images closely resemble the originals. The approach was tested on publicly available datasets like the MITOS-ATYPIA-14 challenge and an internal thyroid carcinoma dataset. Furthermore, a ResNet classifier trained on the CAMELYON16 dataset was used to measure the effect of normalization on classification tasks.

Experiments showed that normalization using CycleGAN improved image similarity indices to over 0.9 and reduced FID scores by up to 96%. Furthermore, when used in classification tasks, the normalized images increased kappa scores by over 50% in some instances.

Similarly, Mao et al. [25] introduced a Single Generative Network (SGNet) to improve stain normalization and the visual quality of histopathology images. The SGNet framework uses a single encoder-decoder architecture with three modules: pre-augmentation, max pooling skip-connections, and a Positional Normalization Dynamic Moment Shortcut (PONO-DMS). These components enhance color registration and preserve structural information. The dataset comprises patches of placental samples.

The results showed that SGNet outperformed existing methods such as CycleGAN and SegCN-Net. In addition, it preserved tissue structures, reduced artifacts, balanced stain colors, and improved segmentation performance.

StainCUT, a method for stain normalization in histopathological images, was introduced in Pérez et al. [12]. The method eliminates the need for paired datasets or reference images by using contrastive learning, setting it apart from template-based methods.

Additionally, the model uses a modified U-Net with fewer parameters for better memory efficiency and a two-step semi-supervised approach: training with unpaired images and applying a patch-wise contrastive loss to preserve content while learning domain-specific styles.

The dataset includes images from the MITOS-ATYPIA challenge [26]. The study also tested the model on a semantic segmentation task over the CAMELYON16 dataset [27].

The results demonstrated that StainCUT outperformed classical methods and StainGAN in terms of structural similarity and quality. Importantly, normalization during training produced better segmentation accuracy compared to normalization during inference.

In a related study, Zhao et al. [13] developed RestainNet for stain normalization in pathological images. The process started with destaining, where images were transformed into grayscale by extracting the luminance channel from the Lab color space. This step removed the color, but retained the structural details. Subsequently, re-staining was done by applying two digital dyes, H&E, to replicate the original stain.

The study tested RestainNet on the MITOS-ATYPIA-14 challenge, the Zhao dataset, and the MICCAI'16 GlaS challenge dataset. Results show that RestainNet preserved structural integrity and color accuracy compared to other approaches. Additionally, it improved performance in classification and segmentation tasks.

Similarly, Nazki et al. [28] introduced MultiPathGAN, a framework designed for stain normalization in histopathological images. This model handles multi-domain challenges using a single generator-discriminator network. Previous methods, such as StarGAN, are constrained to specific domains and fail to preserve delicate anatomical structures effectively. In contrast, MultiPathGAN employs an auxiliary feature extraction network that uses perceptual loss to ensure that salient structural details are retained during style translation. The auxiliary network provides feature maps from a pre-trained ResNet, which helps the generator maintain structural integrity. Training involves mapping input images to multiple domains using random target domain labels. Adversarial, reconstruction, and perceptual losses together ensure the model produces structurally accurate outputs. During inference, the generator translates input images into target domains.

The dataset comprises kidney cancer cases prepared into non-overlapping patches. The experiments showed that MultiPathGAN outperforms methods such as CycleGAN and StarGAN. Moreover, the model was able to normalize unseen domains without retraining.

In a similar research, a technique for stain normalization in histopathological images using a Score-based Diffusion model (SBD) was proposed in [29]. The model was enhanced by stain separation and overlapped moving window patch strategies. The method first applies Sparse Non-Negative Matrix Factorization (SNMF) to separate the hematoxylin and eosin components before normalization. This step reduces the risk of color mis-transfers, which can arise due to the generative diversity of the SBD model. An overlapped moving window patch method is employed to avoid grid artifacts common in patch-wise normalization approaches to further improve performance.

The dataset used in the study included 82 Whole Slide Images (WSIs) of the colon from Asan Medical Center collected between 2009 and 2019. External datasets like CAMELYON16 and PAIP2019 were used for additional validation. The results indicated that the SBD outperformed traditional methods like Macenko, Vahadane, and GAN-based approaches. The overlapping patch strategy also preserved high-quality normalization at less computational cost.

Elsewhere, Hetz et al. [30] proposed a MultiStain-CycleGAN, an approach based on CycleGAN for stain normalization across multiple domains. This method was more efficient than the traditional CycleGAN, which typically requires a separate model for each type of stain. The normal CycleGAN architecture was modified by converting images to grayscale to reduce input variation. This helped the model focus on color normalization instead of dealing with full RGB images. Additionally, the model used a color augmentation function to increase the diversity of input data, which helped it generalize better to unseen staining types.

The datasets used in the experiments included CAMELYON17 and SCC. The results showed that MultiStain-CycleGAN reduced domain shift with lower FID values of 44.964 for CAMELYON17 and 19.240 for SCC. It also achieved higher SSIM values of 0.947 for CAMELYON17 and 0.942 for SCC. Moreover, the downstream task of tumor classification accuracy was high at 0.897 for CAMELYON17 and 0.874 for SCC.

In a recent work, Du et al. [31] proposed DSTGAN for stain normalization. It builds on the Pix2Pix image-to-image framework and forms training pairs by converting well-stained target domain RGB patches to grayscale inputs while using the original RGB patches as targets. The generator is a U-Net-like network with deep supervision. It is trained with two discriminators that separately enforce realistic stain color and preservation of tissue texture. After this supervised phase, the authors introduce a two-stage semi-supervised scheme that also uses unpaired source domain images. At inference, only the generator is used, and a single pass converts any grayscale source patch into a normalized RGB patch.

DSTGAN achieves state-of-the-art results on four datasets, and it outperforms classical methods such as Macenko and Reinhard as well as GAN-based approaches like StainGAN and CAGAN.

2.5. Stain augmentation using DL

Stain augmentation uses DL models to generate synthetic images from existing data, improving robustness and cross-domain generalization.

Articles in this category include Wagner et al. [32], where HistAuGAN, a GAN-based approach to color augmentation in histopathological imaging, was presented.

HistAuGAN employs a multi-domain GAN architecture and improves upon CycleGAN-based approaches by separating tissue structure from stain color, matching color distributions while better preserving morphology.

During training, separate encoders extract content and stain-specific attributes, which are combined with domain information to generate augmented images.

The CAMELYON17 dataset, consisting of 50 WSIs from five medical centers, was used for evaluation. The results showed that HistAuGAN outperformed traditional augmentations and normalization techniques in tumor classification. In addition, it provided better generalization across out-of-domain datasets and reduced variability in performance metrics.

In a similar study, Vasiljevic et al. [33] proposed a technique to address the challenge of domain shifts in histopathological image analysis which is often caused by stain variability. They proposed an unsupervised domain augmentation method using GANs. The method uses a CycleGAN and StarGAN for stain-to-stain translation.

CycleGAN trains a dedicated model for each pair of stains, while StarGAN uses a single unified model with domain labels to handle multiple stains.

A U-Net segmentation model was trained on the augmented images to evaluate the effectiveness of the stain translations. The model successfully learned stain-invariant features, enabling it to generalize and perform segmentation on previously unseen stain types. The evaluation focused on glomeruli segmentation in kidney tissue samples.

Experimental results indicated that CycleGAN outperformed StarGAN and other methods. Its use of separate models allowed for better adaptation to individual stain characteristics. On the other hand, the broader generalization of StarGAN sometimes came at the cost of reduced precision.

2.6. Virtual staining using DL

Virtual staining converts label-free images into their stained counterparts.

Studies in this area include those by Bayat et al. [34].

This paper introduces an end-to-end DL framework for virtual H&E staining, tumor classification, and localization in non-stained prostate biopsy images, combining a cGAN, a ResNet-18 classifier, and a weakly supervised localization algorithm that avoids pixel-level annotations.

The dataset contained 95,605 image patches created from 46 WSIs collected from 38 patients.

The classification model achieved an accuracy of 86.3% and a precision of 85.05%. In contrast, tumor localization reached a Dice Index of 65.07, which is slightly lower than a fully supervised U-Net.

In another study, Li et al. [35] introduced a virtual histology framework to replace invasive skin biopsies. Their approach uses Reflectance Confocal Microscopy (RCM), a non-invasive imaging technique that captures detailed skin structures in grayscale. However, the grayscale nature of RCM makes interpretation difficult for pathologists. To address this, the authors developed a method that transforms RCM images into H&E-like visuals which are more familiar in clinical practice.

The method used Pix2Pix GAN and acetic acid staining to label nuclei in extracted skin tissues to serve as ground truth images for training. This method enabled an accurate mapping of unstained RCM images to their histological equivalents. The dataset included 1085 training images, 137 validation images, and 199 testing images collected from 36 patients.

The results showed that the model successfully generated virtual histology images resembling traditional H&E-stained slides. Furthermore, it achieved approximately 80% sensitivity and 70% precision in predicting nuclei positions and exhibited strong alignment with ground-truth nuclear morphology metrics such as size and compactness.

In a related research, Kaza et al. [36] developed a framework for the automated virtual staining, segmentation, and classification of White Blood Cells (WBCs) in deep ultraviolet (UV) microscopy images.

The pipeline comprises three interconnected tasks: virtual staining with a cGAN, segmentation with a U-Net-like CNN, and five-class WBC classification with a fine-tuned ResNet-18.

The dataset consisted of single-channel UV microscopy images with approximately 400 WBCs.

The authors reported an overall classification accuracy of 93%, based on five donors' test data. Additionally, experiments showed that the virtually stained images resembled the ground truth closely.

Using a similar architecture on a larger dataset of blood smear images from 23 individuals, Kaza et al. [37] reported virtually stained images with an average SSIM of 0.91 and strong downstream performance (Dice scores of 0.9899/0.9718 for cellular/nuclear masks and 94% classification accuracy).

Additionally, Zhang et al. [38] developed a technique for virtual staining of autofluorescence images of tissue samples. The technique employs a cascaded network architecture.

A virtual autofocusing network (Deep-R) first refocused defocused images, which were then fed into a GAN-based virtual staining network. Both models were trained jointly with a style loss so that Deep-R produced features well-suited for the staining model.

The dataset consisted of 5832 fields of view from human lung tissue, each with 512×512 pixels. A high-resolution objective lens was used to acquire both in-focus and defocused images for training and validation. Additionally, the dataset is composed of paired histochemically stained images for benchmarking.

Experiments revealed that the new framework produced results comparable to traditional methods. Moreover, it decreased image acquisition time by about 32% and autofocusing time by 89%.

Elsewhere, Soltani et al. [39] employed two methods for identifying prostate cancer. These methods are: optical staining through Principal Component Analysis (PCA) and virtual H&E staining using a CycleGAN-based deep learning model called UTOM.

Optical staining generates unique contrast maps that highlight critical tissue components such as nuclei, stroma, and glands. This provides nanoscale insights into the tissue alteration methods and aids in understanding tumor aggressiveness. On the other hand, virtual H&E staining transforms UV images into diagnostic-quality H&E-equivalent images without chemical reagents or lengthy staining. Together, these methods improve efficiency and tissue preservation compared to traditional histopathology.

The data included multispectral imaging data from 15 prostate cancer patients. The authors conducted several experiments. As validated by a panel of pathologists, the virtual H&E images closely matched traditional H&E-stained images in quality and diagnostic utility. Moreover, the virtual images accurately captured critical diagnostic features such as gland structure and achieved comparable or improved accuracy for Gleason grading.

Similarly, Boktor et al. [40] introduced a method for virtual histological staining using Total Absorption Photoacoustic Remote Sensing (TA-PARS) and Pix2Pix. To prepare the dataset, unstained TA-PARS images were co-registered with their chemically stained H&E counterparts using image registration tools in MATLAB. This alignment ensured accurate pixel-level mapping between input and target images for training. The approach allowed the model to generate virtual stains resembling standard histological images.

The dataset included 15,000 training patches and 6000 validation patches. They were derived from formalin-fixed paraffin-embedded human skin tissue sections. The findings showed agreement between

virtual and chemically stained images, with an average SSIM of 0.91 and an RMSE of 14.28. In addition, independent evaluations by four dermatopathologists confirmed the quality of virtual stains.

In another study, Martell et al. [41] presented a method for virtual histology using ultraviolet photoacoustic remote sensing (UV-PARS) microscopy combined with CycleGAN. The UV-PARS provided hematoxylin-like contrast for nuclear imaging. In contrast, ultraviolet scattering microscopy created eosin-like contrast for cytoplasm and extracellular matrix visualization. Together, these channels served as input to the CycleGAN. The GAN then uses a forward and reverse training process to transform unpaired UV-PARS data into virtual H&E-stained histology. The CycleGAN framework preserves tissue morphology using a cycle-consistency loss and improves color accuracy through an additional identity loss. The method enabled fast acquisition of virtual H&E images within seven minutes per square centimetre without requiring tissue staining or extensive preparation.

The dataset included 16,000 breast tissue images and 20,000 prostate tissue images, all cropped to 256 × 256 pixels for training. Experiments showed that the virtual H&E images closely matched the quality of brightfield-stained histology with a Median Multi-scale Structural Similarity Index (MS-SSIM) of 0.76 and Peak Signal-to-Noise Ratio (PSNR) of 21.6 dB. In addition, pathologists rated the virtual stains as comparable to or superior to frozen sections.

In another study, Tomczak et al. [42] designed a virtual staining technique for White Blood Cells (WBCs) using a disentangled GAN with confidence estimation. The model separated style and structural features into distinct latent representations through specialized modules for style, background, and WBC structure. Traditional GANs often require paired images, but this approach uses pseudo-segmentation masks and random latent codes to map unstained images to their stained counterparts. They estimated confidence by corrupting latent representations with noise and measuring mutual information across multiple generated samples, enabling low-quality virtual stains to be discarded.

The dataset comprises WBC images from 24 healthy donors. These were acquired using Differential Interference Contrast (DIC) microscopy for unstained images and brightfield microscopy for chemically stained ground truth. The proposed method performed similarly to paired-data models like Pix2PixHD [43] regarding structural preservation and offered greater flexibility in unpaired setups. It outperformed baseline models, including CycleGAN and StarGANv2 [44], in preserving clinically relevant structural details.

In a related work, Abraham et al. [45] proposed a method for virtual staining of quantitative Oblique Back-illumination Microscopy (qOBM) images to H&E-stained images. Traditional histology relies on time-consuming processes, whereas qOBM provides real-time, label-free imaging of thick tissues without complex preparation. Moreover, it combines simple LED illumination with high-resolution imaging to capture subcellular details, making it suitable for rapid intraoperative and diagnostic applications. The authors employed the CycleGAN model to bridge the gap between the grayscale qOBM phase-contrast images and the familiar appearance of H&E-stained histology.

The researchers used datasets including 2358 qOBM and 1737 H&E tiles from mouse liver, rat gliosarcoma, and human glioma specimens.

To evaluate the method, experiments compared the virtual H&E images to traditional H&E. Furthermore, the diagnostic relevance of the produced H&E images was tested with classifiers and neuropathologist reviews. The results confirmed the ability of the model to preserve histological detail and differentiate between healthy and tumor tissues.

MulHiST claims to be the first model designed to generate multiple histological stain images from thick tissue samples [46]. It eliminates the need for extensive preparation and avoids issues with traditional methods that require thin sections. Additionally, it uses light-sheet microscopy images as input to produce stained outputs for H&E, PAS, and MT. It is built on a GAN architecture with an AdaIN layer in the decoder to improve style transfer while preserving content. During training, the model processes shuffled multi-domain datasets to learn domain-invariant features and uses reconstruction loss to maintain the content across styles. For inference, it only requires light-sheet images to generate stained outputs in different styles. This approach eliminates the need for paired images and makes it more practical for thick tissue applications.

The dataset comprised six thick tissue slabs scanned into high-resolution images of 15,000 × 15,000 pixels. The experiments demonstrated that MulHiST produced clear and accurate histological features, such as PAS-positive and PAS-negative regions. The results were evaluated using FID and KID scores, and the MulHiST outperformed baseline models like CycleGAN and StarGAN in quality and accuracy.

Similarly, Min et al. [47] propose a study on wide-field Diffraction Phase Microscopy (DPM) for multi-contrast digital histopathology of mouse organs. The method combines Quantitative Phase Imaging (QPI) and deep learning for virtual staining. The researchers applied DPM to capture detailed phase delay maps of ten major mouse organs without using traditional staining methods. Subsequently, they used a GAN-based model to transform these maps into H&E-equivalent images. Training involved a mix of pre-training with mean absolute error loss and adversarial learning with a PatchGAN discriminator. This approach allowed for effective feature matching and image reconstruction. During the inference stage, QPI images are divided into overlapping patches and processed through the trained model. They are then stitched into virtually stained whole-slide images.

The dataset comprised tissue sections from ten organs, which were augmented to include 50,000 training and 10,000 validation patches per organ, with 256 × 256 pixel patch sizes. Results demonstrated that virtual staining achieved a high SSIM with H&E-stained images.

Furthermore, Khan et al. [48] evaluated the impact of different neural network architectures on virtual H&E staining using the Pix2Pix model. Three variants of the Pix2Pix model were tested. First is the baseline model, which uses a single convolutional layer in the encoder and decoder. The second is a double convolution variant, which adds an extra convolutional layer in each block. Finally, the third version employs a dense convolution inspired by DenseU-Net to enhance network capacity and improve the reconstruction of fine image details. Each variant was trained using paired datasets of unstained and chemically H&E-stained tissue sections.

The study utilized 81 WSIs of prostate tissue for training and evaluation. Results showed that the dense convolution variant outperformed the other two models, achieving the highest PSNR (22.865), SSIM (0.746), and Pearson correlation coefficient (0.916).

Furthermore, Martell et al. [49] proposed a method that employs Ultraviolet Photoacoustic Remote Sensing (UV-PARS) and Scattering Microscopy to generate a realistic virtual histology without chemical staining. UV-PARS detects DNA absorption, while UV scattering provides optical contrast that mimics H&E staining. These pseudo-colored images were then processed using CycleGAN to create more realistic, virtually stained H&E images. The approach was tested on human and murine tissues.

Experiments reveal that virtual histology achieved high diagnostic accuracy, with a sensitivity of 89%, specificity of 91%, and overall precision of 90% in a pathologist concordance study. Furthermore, pathologists preferred the CycleGAN-enhanced virtual stains over frozen section images for their detail and quality.

In a similar study, Wang et al. [50] introduced UBMSVStain, a method for virtual staining of unpaired bone marrow smears using a Content and Attention-Guided Staining (CAGS) module. The authors state that traditional CycleGAN frameworks can perform unpaired image translation, but face challenges in preserving structural details in different image regions. To address this, they designed the CAGS module to improve feature fusion after skip connections. The CAGS module combined the Content-Guided Staining (CGS) and Attention-Guided Staining (AGS) blocks. The CGS block refined low-level details, while the AGS block highlighted important areas of stain transfer. This combination ensured both fine-detailed structures and global style consistency. These

improvements made the framework highly effective for complex bone marrow textures.

The dataset consisted of 49 WSIs from Henan Provincial People's Hospital. It included 14 unstained and 33 Wright & Giemsa (W&G)-stained images. The results showed that adding the CAGS module to CycleGAN improved stain accuracy and structural preservation. In addition, blind evaluations by hematologists also confirmed that virtually stained images matched the quality of chemically stained ones.

A technique for virtual H&E staining of unstained tissue sections using both CycleGAN and Pix2Pix was proposed by Koivukoski et al. [51]. CycleGAN was used to map unstained tissue images to virtually stained H&E images without requiring paired data. However, to enhance the histological accuracy, the study also used the Pix2Pix model, which required paired images. These paired data were created through image registration by imaging the same tissue sections before and after chemical H&E staining. Moreover, the researchers tested tissue thicknesses (3–20 μm) and found that thinner sections (3–5 μm) produced better virtual staining accuracy compared to thicker ones.

The dataset consisted of histological images of tissues from wild-type FVB/N mice, including prostate, liver, kidney, and testis tissues. Findings showed that CycleGAN achieved good virtual staining but had limitations in finer histological details. On the other hand, Pix2Pix improved the representation of tissue morphology and the accuracy of cellular features.

In a related research, Salido et al. [52] evaluated three models: Pix2Pix, CycleGAN, and CUT [53] for creating digital H&E stained images from label-free breast tissue samples. The image pairs for Pix2Pix were obtained through computational alignment with affine transformations. The datasets comprised 1424 pairs of 1024x1024 images and 5294 pairs of 512x512 images.

CycleGAN performed best, achieving an SSIM score of 0.95 and a chromatic discrepancy of around 10%. Moreover, CycleGAN's ability to maintain content while adapting to the style made it the most effective model. Additionally, experts reviewed the results and found that the digitally stained images were suitable for diagnostic purposes.

Elsewhere, Fang et al. [54] introduced a GAN-based virtual peritoneal lavage cytology using Single-color Stimulated Raman Scattering (SRS) microscopy. Single-color SRS microscopy produces grayscale chemical images of tissues or materials by amplifying specific molecular vibrations with a single modulated laser, enabling label-free, high-contrast visualization. They developed a model called CellGAN based on CycleGAN. The model can virtually stain SRS images to mimic H&E cytology. SRS microscopy was chosen because it provides chemical contrast at the subcellular level without requiring dyes. However, its raw contrast does not match the familiar stained appearance in pathology, making virtual staining necessary for diagnostic accessibility. Instead of paired image datasets, which are difficult to obtain due to cell shifts and loss, CellGAN was trained on unpaired images. Additionally, to improve accuracy, a structural similarity loss function was introduced to ensure proper nucleocytoplasmic staining.

The dataset included 7319 exfoliated cells from 32 ascites specimens. The model achieved a 93.8% accuracy in virtual cytology diagnosis. It also demonstrated strong agreement with conventional staining methods, as shown by a Cohen's kappa of 0.782. Furthermore, the study was evaluated for single-cell analysis using deep learning models. The cell detection model achieved a mean average precision of 0.924, while the classification model reached an area under the receiver operating characteristic curve of 0.906.

Similarly, Koka et al. [55] compared deep learning-based virtual H&E staining (vHE) with traditional chemical H&E staining (cHE) for diagnosing lymphomas. The researchers explored a machine learning model that converts autofluorescence images of unstained slides into diagnostic-quality H&E-stained images. The dataset included 40 cases of nodal lymphoma, lymphoproliferative disorders, and reactive lymph nodes. Three board-certified pathologists from different institutions evaluated each case twice, once using cHE and once using vHE, with a two-week gap to reduce bias.

From the experiments they conducted, they found that vHE performed similarly to cHE in most diagnostic assessments. The stain quality pass rate was higher for vHE (92%) compared to cHE (79%). Furthermore, the binary diagnostic concordance between the two methods was nearly identical, with vHE achieving 90% and cHE reaching 92%. The study also found that vHE effectively assessed architectural features, necrosis, and sclerosis, showing a level of agreement comparable to traditional staining methods.

In a related work, Wang et al. [56] developed a virtual staining system to improve prostate cancer diagnosis. The paper focused on Gleason grading, a method that evaluates tumor severity by analyzing cellular patterns and structural organization. To achieve this, they used a custom-built hyperspectral fluorescence microscope to scan unstained prostate tissue and capture autofluorescence images. These images were then processed by a GAN model to generate virtual H&E and Prostatic Intraepithelial Neoplasia-4 (PIN-4) stains. The model was based on an enhanced Pix2Pix framework, which used conditional and unconditional GAN losses. This improvement helped produce more realistic virtual stains. In addition, the model implemented a shift-invariant loss function to correct minor misalignments between autofluorescence and bright-field images. To obtain image pairs for training, the same tissue section was scanned before and after staining. These images were then aligned through a co-registration process to ensure accurate learning.

The dataset included 799 slides from 796 cases, with 557 autofluorescence-H&E pairs and 602 autofluorescence-PIN-4 pairs. The results showed that the virtual staining system closely resembled real stains. In addition, an AI-based Gleason grading model, validated on real H&E images, achieved a high agreement between real and virtual stained images. The quadratic-weighted kappa score was 0.902, while accuracy reached 86.2%. For PIN-4, computational metrics confirmed that virtual images preserved key biological features. However, human pathologist assessments showed that intraductal carcinoma detections on virtual stains were slightly less consistent than on real stains.

Similarly, a technique for virtual H&E staining for glioma tissue from hyperspectral imaging was presented in [57]. The research used the Pix2Pix GAN. Additionally, within the Pix2Pix, a U-Net generator was employed because of its ability to retain both deep and shallow features, which improved image reconstruction. Moreover, the discriminator used was PatchGAN, which focuses on local regions instead of the entire image.

Since Pix2Pix relied on paired images, each hyperspectral image needed to be precisely aligned with its corresponding H&E-stained image. To achieve this, the study first identified a hyperspectral image with the highest contrast at 570 nm, making it features between the two images easier. Key points in both the hyperspectral and H&E images were then selected as reference markers. Using these reference points, a transformation process was applied to adjust for any shifts or rotations. This ensured that every detail of the hyperspectral image matched perfectly with the corresponding H&E image to create an accurate training pair for the model. The dataset included 10 glioma tissue sections, producing 627 registered image pairs.

The model achieved an SSIM of 0.7731 and a PSNR of 23.3. In addition, the U-Net-based generator outperformed other architectures. Moreover, the model reconstructed whole-slide images at a speed of 3.81 mm^2/s, making it suitable for real-time applications.

In a similar study, Yoon et al. [58] developed an automated virtual staining, segmentation, and classification in label-free Photoacoustic Histology (PAH) of human liver specimens. First, Explainable Contrastive Unpaired Translation (E-CUT) was developed as a deep learning model that transforms grayscale PAH images into virtually stained H&E images while maintaining morphological integrity. Compared to models like CycleGAN, E-CUT uses saliency loss and gradients to enhance structure preservation and explainability. The second step involves a U-Net-based segmentation model, which extracts

histological features such as cell area, cell count, and nuclear distance from the virtually stained images.

In the final step, a Stepwise Feature Fusion (StepFF) classification model combines deep feature vectors from PAH, virtual H&E (VHE), and segmentation outputs to improve classification accuracy.

The study relied on PAH images obtained from formalin-fixed paraffin-embedded human liver tissues. When evaluated, the framework achieved a classification accuracy of 98.0% using StepFF, outperforming conventional PAH classification, which reached 94.80%. Sensitivity was measured at 100%, showing its potential for real-world clinical applications. The E-CUT model also delivered the lowest FID and Kernel Inception Distance (KID) values, confirming good virtual staining quality.

In addition, Asaf et al. [59] proposed a GAN model for producing H&E-stained images. The DCLGAN framework builds upon CycleGAN and Contrastive Unpaired Translation (CUT) [53] by using dual learning, which enhances mutual information maximization between input and output images. Compared to CycleGAN, which depends on cycle consistency, DCLGAN adopts contrastive learning. This type of learning was initially introduced in CUT to learn domain-specific features effectively. The architecture consists of two generators and two discriminators for bi-directional image translation while maintaining domain-specific embeddings.

Training used adversarial, identity, and patch-wise contrastive (NCE) losses to preserve key image features.

Paired image data were necessary to train the model effectively, as it ensures the network learns from actual tissue correspondences rather than unstructured variations.

The dataset used for training included 56 pairs of unstained and H&E-stained skin tissue slides, further expanded into 92 tissue pairs and processed into 13,824 usable image patches.

The experiments showed that the FID between virtual and real H&E stains was 80.47. This value was much lower than the FID for unstained and virtual images (342.01) or unstained and H&E-stained images (320.4). A lower FID suggests greater similarity, indicating that virtual stains closely match actual stains.

Additionally, the Kernel Inception Distance (KID) results support this conclusion. The mean KID for virtual and H&E stains was 0.022. In contrast, the KID for unstained and H&E stains was 0.28, and for unstained and virtual stains it was 0.31.

In addition to quantitative metrics, dermatopathologists assessed the quality of the virtual stains. For side-by-side comparisons with real H&E-stained images, they reported an average agreement of 78.8% across key visual and diagnostic features. When evaluating virtual stains on their own, without seeing the real stained counterpart, they judged 90.2% of these images to be of sufficient quality for diagnostic purposes.

In another study, Fan et al. [60] developed a CycleGAN-based model for generating bright-field H&E stained images using unpaired Mueller Matrix (MM) polarization images. The model eliminates the need for pixel-level registration between the polarization and bright-field images, which is typically required in paired virtual staining techniques. To prepare the data, dimensionality reduction was applied to the MM polarization images through Principal Component Analysis (PCA), reducing the 16-channel data to three dimensions. This step was essential because bright-field images have only three channels, ensuring compatibility between the two image domains.

Additionally, stain normalization was performed on the bright-field images to address variations in staining intensity and hue. Finally, image registration was performed on a small subset of paired images to evaluate the model's performance, although the model was trained with unpaired data.

The dataset consisted of 9 human intestinal tissue slices, each yielding multiple regions for both MM and bright-field imaging. These images were augmented by cropping into 256×256 pixel patches, resulting in 13,500 MM polarization images and 13,500 bright-field images. The results showed that the model, especially when enhanced with skip-connection structures and SSIM loss functions, was able to produce high-quality images, achieving an SSIM of 0.71 and a PSNR of 31.24.

Elsewhere, Chen et al. [61] proposed an approach for diagnosing Lung Adenocarcinoma (LUAD) by using virtual H&E staining on label-free autofluorescence images. The proposed method employs a weakly supervised U-Frame learning framework for the virtual staining task and the CLAM-SB model for tumor classification. The U-Frame model is designed for complex image restoration and style transformation, which distinguishes it from models like Pix2Pix and CycleGAN. Its architecture consists of two generators and two discriminators built on CNNs. The two generators embed images into a shared latent space before transforming them into images of different domains. Through this process, autofluorescence images are converted into virtually stained H&E images. In the subsequent tumor diagnosis step, the CLAM-SB model classifies (binary) LUAD using virtually H&E-stained images.

The model was evaluated on a dataset that included 785 H&E-stained whole-slide images (WSIs) from The Cancer Genome Atlas (TCGA) LUAD dataset. The results show the CLAM-SB model achieving an average AUC of 0.973 on virtual H&E-stained WSIs. This closely matches the 0.977 AUC obtained from standard H&E-stained WSIs. The autofluorescence diagnosis model also achieved an average AUC of 0.930, showcasing the potential for direct diagnosis from autofluorescence images. Moreover, the attention heatmaps produced by the CLAM-SB model helped locate tumor regions in both virtual and ground-truth H&E-stained images, and they closely align with pathologist annotations.

In a related work, a Multi-channel CycleGAN (MC-GAN) for virtual histological staining of Photon Absorption Remote Sensing (PARS) images to H&E was developed by Boktor et al. [62]. This approach enhances the colorization of tissue structures by using input data from radiative signals, non-radiative signals, and extracted features. These signals are obtained through a modified K-means clustering method, which allows for intelligent extraction of multi-modal information from the time-domain signals. The extracted features enable the network to better distinguish between different tissue types, which is required for accurate staining. Compared to traditional CycleGANs that rely on fewer input channels, the MC-GAN handles multiple feature channels, thus increasing its ability to understand complex tissue structures. Furthermore, while image pairs are not strictly required for CycleGAN training, they used registered image pairs to improve performance and reduce artifacts such as hallucinations.

The dataset used in the study includes thin sections of Formalin-Fixed Paraffin-Embedded (FFPE) human skin and mouse brain tissues. The dataset for human skin includes approximately 500 overlapping patches, while the mouse brain dataset comprises around 2000 patches, all extracted from the PARS and H&E images. The best feature combination for the human skin dataset produced an SSIM of 0.89, PSNR of 22.92, RMSE of 18.22, and an LPIPS of 0.15. Similarly, the optimal results for the mouse brain dataset showed an SSIM of 0.85, a PSNR of 22.38, and an RMSE of 19.21.

Similarly, Li et al. [63] developed a model to generate virtual histological staining from OCT images of human coronary arteries called SCPAT-GAN. Compared to previous methods, SCPAT-GAN does not require pixel-wise paired training data. In addition, it employs a convolutional transformer architecture with structural constraining and pathological awareness modules to ensure accurate representation of both normal and pathological features. Moreover, the approach avoids the need for extensive manual annotation.

The model was trained on 194 OCT images from 23 patients, with 4297 patches used for training. SCPAT-GAN outperformed previous methods, achieving better quality virtual stained images. The model achieved an FID score of 175.7 and a PHV score of 57.35, outperforming Coronary-GAN and CycleGAN in normal and pathological datasets. Additionally, pathologists found SCPAT-GAN-generated images to be very similar to real ones.

In a related study, the impact of ground truth cell label derivation on deep learning models for immunophenotyping was investigated [64]. The focus was on virtual staining of CD3+ T-cells in lung cancer tissues from H&E images. The paper compared two Pix2Pix models: one trained using cell labels from the same tissue section as the H&E-stained section (the "same-section" model) and another using labels from an adjacent section (the "serial-section" model). The motivation was to address potential inaccuracies caused by section misalignment in conventional training approaches. The same-section method enabled precise cell mapping by reducing distortions introduced during sample preparation.

To evaluate these models, the researchers trained them on a 57 lung carcinoma tissue sample dataset. The results showed that the same-section model outperformed the serial-section model across multiple evaluation metrics, achieving a Pearson correlation of 0.967 versus 0.648 in predicting CD3+ cell counts ($p < 0.005$). Similarly, the accuracy was higher at 98.2% for the same-section model compared to 96.3% for the serial-section model.

Similarly, Yang et al. [65] developed a cGAN model that transforms autofluorescence images into virtual brightfield images and birefringence-stained images. Although a standard cGAN does not inherently require paired images, their method uses a modified cGAN in which a dedicated registration step is applied to create pixel-aligned autofluorescence and target image pairs. This transforms the problem into a supervised paired image-to-image task. In this setting, the cGAN becomes preferable to CycleGAN, since CycleGAN operates on unpaired data and relies only on cycle-consistency, which cannot ensure the level of structural correspondence needed for accurate stain synthesis.

The model consisted of three main components: a generator for image transformation, a discriminator for refinement, and an image registration module to correct spatial misalignments. The image registration process had two distinct phases. The first was a preprocessing step, in which brightfield and birefringence images were aligned with autofluorescence images using global registration techniques and elastic image transformation. The second was a training-integrated registration module, which learned residual misalignments during training to ensure better pixel-wise correspondence between the input autofluorescence images and their virtually stained outputs.

The dataset used for training and validation included 451 cardiac tissue samples from eight patients. The model achieved an MAE of 0.22 and an FID of 0.0066. Pathologist evaluations also showed that for birefringence image quality (M6), there was 91% agreement between the virtually stained and manually stained images.

In a recent work, Xu et al. [66] proposed a unified framework for digital H&E staining of dark-field images that works for both unpaired data and paired-but-misaligned data. It uses a teacher–student setup, where the student is a CycleGAN-style generator trained with a Least Squares GAN (LSGAN) objective. The teacher pipeline first enhances the image and then colorizes it with a U-Net. With this design, the student learns to mimic the teacher via knowledge distillation and is further guided by adversarial and cycle-consistency losses. For misaligned pairs, the framework adds a Learning-to-Align registration network that produces smooth deformations so that the images line up better.

As a result, the method outperforms several image-to-image and virtual-staining baselines in quality metrics and maintains downstream classification accuracy within about 1–2% of real stained images.

In a related study, Li et al. [67] performed virtual staining by transforming label-free autofluorescence images into bright-field H&E, Masson's Trichrome, and EVG stains for lung and heart transplant biopsies. They used a structurally conditioned GAN with a U-Net generator, trained on precisely registered autofluorescence–histochemical pairs, and uniquely created multiple perfectly aligned stains from a single tissue section. The dataset includes 5958 paired patches from 36 patients. Pathologists judged the virtual stains to closely match real histochemical stains, producing diagnostic concordance rate of 82.4% for lung and 91.7% for heart.

Furthermore, Zhang et al. [68] developed a method that converts low-resolution autofluorescence images of label-free lung and heart tissue into higher-resolution bright field H&E images in a single step. The method uses an image conditional Brownian Bridge diffusion model, where a shallow CNN upsamples the input and an attention U-Net denoiser learns the reverse diffusion process toward the target histology. The Brownian Bridge is conditioned on both the autofluorescence input and the histology endpoint, which tightens the mapping between domains and enables joint virtual staining and two to five times pixel super resolution. Trained on 1051 lung and 163 heart autofluorescence H&E pairs and tested on 180 lung and 155 heart patches, the Brownian Bridge diffusion model outperforms cGAN and Denoising Diffusion Probabilistic Models (DDPM) and better preserves small structures such as anthracotic pigment.

Similarly, Liu et al. [69] introduced Cytoland, a virtual staining framework that converts label-free three-dimensional quantitative phase images into synthetic fluorescence images of nuclei, cell membranes, and in some cases an infection reporter channel. Cytoland uses a UNeXt2-based architecture trained first with self-supervised masked autoencoding on large phase datasets and then with supervised learning on paired phase and fluorescence images. The training employs physics-inspired augmentations that replicate different microscopes and imaging conditions, which improve robustness. The authors trained variants of Cytoland on datasets that include human cell lines, induced pluripotent stem cell-derived neurons, virus-infected cells, and zebrafish neuromasts. Across tasks, they show high correlation with real fluorescence and segmentation performance that matches or exceeds models trained directly on fluorescence images. They also outperform simplified versions of the architecture. However, the method was not compared systematically to other virtual staining approaches.

Additionally, Park et al. [70] introduced a virtual staining pipeline that converts label-free 3D refractive index volumes from holotomography into 3D virtual H&E images of thick colon and gastric cancer tissues. Holotomography is first used to image unlabeled paraffin sections, reconstructing high-resolution refractive index maps of colon or gastric cancer slides from multiple intensity measurements. The resulting axial slices form the direct input to the deep learning model. The model used is a supervised cGAN whose generator is a 3D architecture search-based model tailored to medical volumes, so it can map refractive index stacks to brightfield H&E while preserving fine glandular and nuclear structures. Paired data were obtained by acquiring holotomography refractive index and single-channel brightfield images, registering them to whole slide H&E, and cropping them into thousands of patch pairs. In experiments, the trained model produced virtual H&E volumes that closely match chemical H&E in nucleus counts, areas, and anatomy. Compared with U-Net, Diffusion, and Pix2Pix baselines, the proposed cGAN attains similar or slightly lower performance, but it preserves glands, stroma, and nuclei better.

2.7. Stain transformation using DL

This section outlines studies that employ DL to transform one stain type to another. These articles include Mercan et al. [71]. In the article, a method to improve mitosis detection in breast cancer tissue images was developed. Mitosis detection is a strong prognostic indicator used in cancer classification. However, inter-observer variability and artifacts created by H&E stains often limit the process.

In contrast to H&E stains, Phospho Histone-H3 (PHH3) stained images are more reliable for mitosis detection. However, generating PHH3 stains from H&E requires a complex double staining process. To address this challenge, the authors employed GANs to generate synthetic histopathology images that transform H&E-stained tissue images into PHH3-stained images and vice versa.

The methodology presented involves using cGAN and CycleGAN to map images between the H&E and PHH3 stains. These synthetic images are then used to train CNNs for automatic mitosis detection. The dataset

used for the study consists of 18 WSIs from patients with triple-negative breast cancer.

The authors show that their approach of using synthetic H&E images performed similarly to or even better than models trained on real H&E images. Additionally, a second method using features extracted from the deep layers of the GANs produced even better performance than the baseline model that used real images.

In another work, Burlingame et al. [72] proposed a method called SHIFT. The technique simplifies and speeds up cancer diagnosis by converting H&E-stained histological images into virtual Immunofluorescence (IF) images. The SHIFT model uses cGANs to infer molecular marker distributions.

They developed a unique protocol to spatially register H&E and IF images from the same tissue section to train and test the SHIFT model. They worked with a small dataset comprising samples from four pancreatic cancer patients. To improve generalization across heterogeneous samples, a Variational Autoencoder (VAE) was used to extract compact feature representations from image tiles to guide the selection of representative training samples. This approach allowed them to optimize model performance despite the limited dataset.

The results show that SHIFT could generate virtual IF images comparable to real IF images. Furthermore, when trained on optimal sample combinations identified using the VAE, the model achieved higher SSIM scores than less representative samples. Additionally, the SHIFT model outperformed a competing supervised deep learning method called Label-Free Determination (LFD).

In a similar research, Wang et al. [73] presented an approach to enhance the application of Fourier Ptychographic Microscopy (FPM) for digital pathology. FPM is known for generating high-resolution and wide Field-Of-View (FOV) images without mechanical scanning. However, challenges like coherent artifacts and color accuracy prevent its broader application. The authors proposed an unsupervised deep learning framework to address these issues. The method takes an FPM monochromatic reconstruction image and virtually stains it. Not only does the process stain the images, but it also enhances image quality, reduces artifacts, and shortens acquisition time.

The study employed a CycleGAN with a multiscale structural similarity loss function to perform image-to-image translation. The technique bypasses the need for paired training data, which is often challenging to acquire. It uses two unpaired image sets: monochromatic FPM reconstructions and high-resolution images from standard microscopes. During training, the network reduces coherent artifacts and enhances image quality.

The dataset consists of 1500 high-resolution images for each staining type (immunohistochemistry brown, red, and H&E) and 1000 fluorescence images.

Experimental results showed that the network achieved higher SSIM compared to conventional FPM outputs. Furthermore, the approach reduces acquisition time by 67%, making it more practical for clinical settings.

Elsewhere, Levy et al. [74] explored the use of unsupervised virtual trichrome staining technologies for liver biopsy analysis in patients with Non-Alcoholic Steatohepatitis (NASH). They stated that traditional methods of physical staining, H&E to trichrome for evaluating disease severity, are resource-intensive. Thus, the research implemented a computational approach using a CycleGAN model to transform H&E-stained images into trichrome-stained representations. The model was chosen for its ability to handle unpaired image translations.

The dataset comprises 574 WSIs from 273 patients with biopsy-proven NASH. These images were collected over 12 years from Dartmouth Hitchcock Medical Center and supplemented with clinical and demographic data.

The results showed a high correlation between pathologists' evaluations of real and virtual stains. Additionally, the virtual staining was validated through Turing tests, and it produced comparable clinical biomarker correlations and disease progression predictions.

In another study, Picon et al. [75] proposed a method for virtually staining reconstructed autofluorescence images to resemble H&E microscopy images. The authors stated that the technique would aid optical biopsy in clinical applications and enable in vivo biopsy. They developed a framework using a Siamese network architecture. The method avoided using GANs due to their tendency to produce visually appealing but less accurate images.

Rather, they used a fully convolutional DenseNet architecture with additional regularization to ensure pixel and feature levels similarity. This approach addressed perceptual distortion issues often encountered with GANs.

The researchers used two datasets, i.e., an extensive H&E microscopy dataset from the Basque Biobank, which contained over 259,000 image tiles from different organs, and a simulated autofluorescence dataset derived from the H&E images.

The results showed that the method achieved a Balanced Accuracy (BAC) of 0.95 and a negative predictive value of 1.00 when detecting colorectal tumor tissues. These results matched the accuracy obtained using original H&E images.

In another study, Lo et al. [76] proposed a method that combines CycleGAN-based stain translation and Faster R-CNN to detect glomeruli in renal pathology images. The CycleGAN model translated the images between four different staining styles (H&E, PAS, Masson, and Silver). This approach addresses the challenge of limited stain-specific training data and enhances detection accuracy by converting less-common stains into the more familiar H&E or PAS styles. The Faster R-CNN was used to detect glomeruli across varied magnifications. During inference, the method first translates test images into the desired stain style using CycleGAN before detection by the Faster R-CNN.

The dataset comprised over 4300 images with different magnifications. Experiments showed that the CycleGAN-translated images were identical to real ones. Furthermore, the Faster R-CNN demonstrated better detection performance with F1 scores exceeding 0.9 in some experiments, outperforming previous methods.

Elsewhere, Kaza et al. [77] invented a framework to virtually stain single-channel UltraViolet (UV) microscopy images for hematological analysis. The method employs a cGAN trained on image pairs of grayscale UV images acquired at 260 nm and their corresponding pseudocolorized images. This design simplifies the traditional three-channel process by focusing on a single channel. Thus, it increases imaging speed threefold without losing accuracy. The dataset consists of approximately 35,000 image patches, each of 256x256 pixels, extracted from larger 1040x1392-pixel tiles.

Testing was done on 1800 patches. The results showed that the virtually stained images closely resembled Giemsa-stained images, with minor differences in nuclear contrast. Additionally, SSIM values were used to evaluate the results, and they showed good results for most color channels.

Similarly, a study to address the challenge of domain shift in medical image analysis caused by differences in image acquisition and processing protocols was presented in [78]. It proposed improving bias transfer using generative models to avoid frequent retraining. The authors evaluated three models: CycleGAN, U-Net CycleGAN, and Fixed-Point GAN (FPG). In addition, they enhanced the models with additional loss functions like MS-SSIM, structure loss, and identity loss to preserve image content and structure and thus improve segmentation and classification performance.

The models were trained using annotated datasets of kidney and prostate biopsy images. Finally, the models were tested for content preservation, domain adaptation, and impact on downstream tasks. After tests, the U-Net CycleGAN and FPG with structural loss performed best.

In a related study, de Haan et al. [79] developed a technique to transform H&E-stained tissue images into three special stains: PAS, Masson's Trichrome, and Jones' Silver.

Firstly, virtual staining networks were employed to generate perfectly paired datasets of H&E and special stain images from autofluorescent tissue sections. This ensured structural consistency and eliminated errors from unpaired data. Eight CycleGAN-based style transfer networks were applied to address variability in staining techniques and imaging equipment augmenting the training dataset. This step allowed the network to generalize across different histological conditions.

Next, a stain transformation network was trained using the augmented dataset. The network processed 256x256 pixel patches, randomly selected from over 7800 patches. During training, original and augmented H&E images were used as inputs, with paired special stain images serving as ground truth. Finally, the trained network was tested on unseen H&E images from 58 kidney biopsy samples. Pathologist evaluations showed that the network improved diagnostic accuracy and efficiency compared to traditional methods.

In a similar work, Hong et al. [80] proposed a deep learning-based method for measuring the tumor-Stroma Ratio (TSR) in gastric carcinomas. TSR measures the proportion of tumor cells to stromal tissue, a critical factor in cancer prognosis. They stated that traditional inspection by pathologists could lead to variability issues. Thus, they used Pix2Pix to generate cytokeratin (CK)-stained images from H&E slides. This conversion enabled the automated segmentation of tumor and stroma regions. Subsequently, the TSR can be calculated by determining the proportion of stromal tissue within the tumor microenvironment.

The study used a dataset of 373 gastric cancer patients, with a training set of 13 image pairs and a test set of 358 images. The results showed a strong agreement between the deep learning-based TSR (dTSR) and those obtained from pathologists called visual TSR (vTSR), with a kappa value of 0.623 and an AUC of 0.907 for classification. Additionally, tumors with a high TSR (more stroma) were linked to worse survival outcomes, which proved their prognostic value.

In a related study, Wieslander et al. [81], a method for creating fluorescence-like images from bright-field images of adipocyte cells was developed. This reduces the need for traditional staining, which is expensive. Three separate models were used for nuclei, cytoplasm, and lipid droplets. The nuclei model used Learning Using Privileged Information (LUPI) with segmentation masks to improve feature accuracy. The lipid model employed adversarial training and gradient losses for better edge detection and defect reconstruction. The cytoplasm model used a simpler U-Net architecture to avoid losses and prevent artifacts. These models were combined by generating their respective fluorescence channels, which were then joined into a single virtual stain.

They tested a dataset of 96 bright-field images. The models worked together to create realistic virtual fluorescence images that matched real stained images.

In another work, Lahiani et al. [82] proposed an approach for virtual WSI synthesis in digital pathology. The technique used is CycleGAN with Perceptual Embedding Consistency (PEC) loss. The additional PEC loss helped address tiling artifacts, which are visible discontinuities between adjacent image tiles caused by tile-wise processing in high-resolution images. Additionally, the PEC loss helped the network preserve features like cell size, morphology, connective tissue texture, and nuclear density. The dataset consisted of 50 WSIs from colorectal carcinoma cases.

The results showed that the PEC loss decreased tiling artifacts and generated seamless reconstructed images. In addition, the similarity index between virtual and real images improved. Furthermore, pathologists validated the method through clinical interpretation, and results revealed high agreement between virtual and real images, especially for CK and necrosis classifications. However, discrepancies were observed for Fibroblast Activation Protein (FAP) due to its complex patterns.

Similarly, Liu et al. [83] developed PC-StainGAN, a deep learning model for virtual stain transfer from H&E-stained images to Ki-67-stained images. The researchers recognized that traditional IHC staining is essential for accurate cancer diagnosis. However, the process is often expensive and tedious. In addition, existing methods like the CycleGAN cannot preserve structural details or maintain pathological consistency. To address these shortcomings, PC-StainGAN employed innovations such as skip connections for structural detail preservation and a pathology consistency constraint to ensure that the generated images aligned with actual pathological features. These advancements allowed the model to produce realistic Ki-67-stained images using minimal expert annotations.

The annotations ensured pathological consistency of the generated images. The model used a de-staining and re-staining approach, which virtually replicated clinical workflows while avoiding irreversible alterations to tissue samples.

The study used neuroendocrine and breast histopathology images, each containing approximately 54,000 training tiles. Additionally, unique loss functions, including pathological consistency loss and base-space aligned loss, were added to refine model performance. The PC-StainGAN outperformed existing methods.

In another study, Cetin et al. [84] invented a model called StainGAN for virtual re-staining of histopathological WSIs. The model combined self-attention GAN and contrastive learning to transform H&E-stained WSIs into CD8-stained images. The CD8-stained images are often used to analyze the presence of immune cells in tissues. They trained the StainGAN on a dataset of 36 WSI pairs of H&E and CD8-stained images. Subsequently, testing was performed on two additional WSI pairs to evaluate the generalizability of the model.

The results indicated that StainGAN achieved higher qualitative and quantitative performance compared to state-of-the-art models like CycleGAN and Pix2Pix. Furthermore, expert pathologists assessed the synthetic images and found them comparable to real CD8-stained images in terms of immune cell detection and overall structure.

Elsewhere, Lysik et al. [85] explored the use of GANs for the transformation of H&E-stained renal tissue images into Periodic Acid-Schiff (PAS). Four models were evaluated: Pix2Pix, modified Pix2Pix, CycleGAN, and modified CycleGAN. The Pix2Pix functioned as a supervised model requiring paired data. The second model, CycleGAN, was designed for unpaired data and employed dual generators and discriminators to map images between two domains. The third model, modified Pix2Pix, incorporated elements from CycleGAN by adding dual generators and discriminators to enable bidirectional transformations. Finally, the modified CycleGAN adapted the unsupervised approach of CycleGAN to a supervised context by using paired data and a simplified single generator and discriminator architecture.

The study used a unique dataset of 10 paired HE-PAS slides, consisting of 1014 annotated glomerular structures. The results showed that CycleGAN and modified Pix2Pix performed best, with CycleGAN achieving the lowest FID of 6.96 and a success rate of 96% in visual evaluations. On the other hand, Pix2Pix performed poorly, with an FID of 10.96 and only 52% accuracy in visual assessments.

In a related research, Zhang et al. [86] developed a technique for performing virtual staining on unlabeled human carotid atherosclerotic tissue. This approach was designed to replace traditional staining, which is time-consuming and requires multiple tissue sections. The researchers used the Pix2Pix GAN model to convert bright-field images into virtually stained ones. These stains included H&E, PSR, and EVG. The methodology used pseudo-labels for some stains to address the difficulty of obtaining multiple labels from a single tissue section. This ensured accurate multi-stain training without additional tissue preparation. The dataset used in the study contained 140 unstained WSIs paired with stained versions. It was divided into over 105,000 training pairs and 35,000 testing pairs.

The model outperformed other methods such as StarGAN [87] and StGAN [88]. In addition, it showed higher similarity to real stained images based on evaluation metrics. In addition, blind reviews by pathologists confirmed that the system could accurately identify features such as necrotic cores and collagen fibers.

Similarly, Zhang et al. [89] introduced MVFStain, a framework designed to generate virtual functional stain images from H&E-stained histopathology images. The framework includes several key

components: a content encoder, a style encoder, a decoder, and two discriminators (content and style). The content encoder extracts structural features from the input images, while the style encoder isolates the color-specific features of the stains. The decoder then combines these features to synthesize the virtual stains. Subsequently, the discriminators ensure that the generated images preserve the structural content and match the color distribution of the target stains. Additionally, they used KL loss to align style features with a Gaussian distribution, and histo loss was employed to refine domain-specific style features thus, enabling accurate and consistent stain generation.

Training was carried out using four datasets: lung lesions, lung lobes, breast tissue, and atherosclerotic lesions. These datasets consisted of 128x128 image patches, with non-overlapping patches generated for training and testing. The results demonstrated that MVFStain produced virtual stain images of comparable or better quality than existing methods. Moreover, it preserved structural details and achieved accurate predictions of positive signals.

In a related work, Zeng et al. [90] developed a semi-supervised H&E to Progesterone Receptor (PR) stain transformation for breast cancer diagnostics. This approach differs from traditional methods that rely on pixel-level paired images or expert annotations. In contrast, it uses unpaired H&E and PR slides. In the training phase, the model starts by aligning consecutive slides using a registration network. Then, positive/negative (Pos/Neg) labels are assigned to PR patches based on staining intensity and transferred to corresponding H&E patches to enable supervised training without explicit annotations. Subsequently, a semi-supervised GAN guided by tailored loss functions is used for the transformation. During inference, the trained model is able to translate H&E images into PR-like images.

The dataset comprised 30 pairs of pathology slides, divided into 22 for training and 8 for testing, with over 58,000 H&E and 62,000 PR patches processed. The model outperformed established methods such as CycleGAN and UGATIT [91] by achieving better scores on metrics like SSIM and PSNR.

Elsewhere, Zingman et al. [92] evaluated several Image-to-Image Translation (I2I) methods for stain transfer in histopathology. They focused on conversion between H&E and Masson's Trichrome (MT) stains. The study compared 12 methods consisting of three traditional algorithms and nine GAN-based approaches.

One fascinating part of the study is how the challenge of obtaining aligned image pairs, which are essential for training most cGANs, was addressed using a clever approach. First, they teach a cGAN model a simpler job first: "given a black-and-white picture of MT, learn how MT usually adds color." The model practices only on MT images by stripping each one to black-and-white and then learning to put the MT colors back.

Now, in order to convert H&E to MT, the H&E image is first transformed into black-and-white. This black-and-white image is given to the trained model to predict MT coloring. Finally, this predicted color is combined with the H&E black-and-white image to obtain something that looks like MT. Thus, the H&E is converted to its equivalent MT stain.

The dataset consisted of approximately 26,000 image tiles (256 × 256 pixels) for each stain type, extracted from the WSI of mouse liver tissues. Additional validation and test datasets containing about 1300 tiles were included.

Results showed that CycleGAN performed best, with minimal structural distortion and excellent color and texture matching. Other GAN-based methods, such as StainGAN and MUNIT, performed well but had some limitations. On the contrary, simpler traditional methods like Macenko and Vahadane struggled to replicate fine details and colorpatterns.

In a related research, a virtual staining method for generating Immunohistochemical (IHC) images from H&E images for breast cancer diagnosis was proposed in Duan et al. [93]. The methodology is based on a GAN with a structured generator and discriminator. The generator has a content encoder to extract tissue structure information from H&E images and a style encoder to capture staining styles from the target domain. These features are then combined through convolutional layers to produce IHC images. The discriminator is based on PatchGAN and is designed to handle two tasks. It evaluates the authenticity of generated images and classifies the expression of HER2 into five categories.

They employed 3896 paired H&E and IHC images for training and 977 paired images for testing. The dataset is part of the BCI dataset. Experiments show that the proposed method achieved the highest PSNR (21.522 dB) and SSIM (0.548) scores compared to other methods, such as CycleGAN and Pix2Pix variants. Moreover, it accurately identified the HER2 regions even in low-expression cases.

Similarly, Sloboda et al. [94] proposed xAI-CycleGAN, an explainable version of CycleGAN. The model was used to transform histopathology images from H&E stains to P63 stains for metaplastic breast cancer. It employs explainability features, such as interpretable latent variables and saliency-driven training, to improve transparency and usability. The model also uses context loss to ensure the structural integrity of tissue images during the transformation process. Furthermore, a modified SeFa algorithm [95] is used to edit the outputs. This allows users to fine-tune the generated images and align them more closely with ground truths.

The dataset comprises 110,564 image slices from 32 paired samples of H&E and P63-stained tissues. The results showed that xAI-CycleGAN produced high-quality images with better structural preservation and reduced artifacts compared to existing methods. Moreover, a survey involving histopathologists rated the generated images with an average realism score of 4.94 out of 6.

In a similar research, Biswas et al. [96,97] proposed a method for transforming hyperspectral H&E images to Verhoeff's van Gieson (EVG) images. The paper claims to be the first to transform hyperspectral H&E images to EVG at pixel level. They state that routine H&E cannot separate elastic and collagen fibers because both take on a similar pink color and pattern. In contrast, EVG stains elastic fibers deep blue and collagen an orchid color. However, EVG is complex and time-consuming. Thus, they developed a modified CycleGAN model to convert H&E to EVG.

First, they redesign CycleGAN with asymmetric, modality-aware generators and discriminators so it can translate from sixty-one band hyperspectral H&E to three-channel EVG and back while preserving stain detail. Next, they make identity loss workable across the channel mismatch by temporarily expanding EVG and by reducing H&E to three domain-guided basis channels, which stabilizes training. Finally, they add a supervised refinement after the unpaired unsupervised stage. This was achieved by training on a small set of carefully registered image pairs. With this, they were able to reduce noise, suppress false elastic fibers, and improve the realism of the synthesized EVG.

The dataset consisted of hyperspectral H&E and whole-slide EVG images from pancreatic tissues. Ground truth images were prepared through image registration to align hyperspectral H&E images with EVG images. Experiments revealed that the proposed method produced realistic and less noisy EVG images.

In a related study, Ling et al. [98] introduced a lightweight Self-supervised Disentanglement Network (SDN) for stain transfer. Traditional methods, such as GAN-based models, struggle to generalize to unseen staining domains. In comparison, SDN employs a unique approach of separating an image into stain features and content features. This approach allows the model to transfer staining styles between domains by swapping stain features while keeping the image's content intact. Thus, the SDN can perform style transfer even to domains unseen during training. During inference, the input image is split into stain and content features. Then, stain features from a reference image are combined with the content to generate a new image with the target staining style.

The evaluation was performed on two publicly available datasets: MITOS-ATYPIA-14 and CAMELYON17. SDN showed better performance in stain transfer compared to state-of-the-art methods. Additionally, the model enhanced downstream classification tasks by improving AUC scores on unseen datasets without requiring additional tuning.

Additionally, Yan et al. [99] introduced a Prior-Guided GAN (PG-GAN) to convert H&E-stained images into virtual Masson trichrome-stained images. Masson trichrome staining is preferred for assessing liver fibrosis because it highlights important collagen fibers for accurate evaluation of fibrosis. They used Contrastive Unpaired Translation (CUT) as the model backbone of the PG-GAN. CUT was chosen over CycleGAN due to its better stability, faster training, and ability to avoid over- or under-staining. The model was trained in two stages. The first stage involved training both the encoder and decoder to extract and reconstruct structural features. In the second stage, the decoder was fine-tuned for stain normalization. In addition, pathology feature loss was introduced to focus the network on histological structure to improve translation accuracy.

The data consist of 72 cases of liver biopsies with varying fibrosis stages, resulting in 255,297 H&E and 214,850 Masson trichrome image patches. Pathologist evaluations showed high accuracy and consistency between the virtual and real stains.

Elsewhere, Liu et al. [100] proposed a method for virtual fluorescence translation using a cGAN. The cGAN framework predicted fluorescence images across different channels, including nuclei, membrane, and cytoskeleton. The generator used a U-Net architecture with channel attention layers to enhance feature learning. In contrast, the PatchGAN discriminator focused on evaluating smaller image patches for detailed accuracy. The technique produced realistic virtual fluorescence imaging, which could be extended to multi-label imaging by combining predicted channels.

The data used consisted of images from mouse kidney tissue sections. The proposed method outperformed CycleGAN and BiCycleGAN in quantitative metrics such as SSIM, PSNR, and MAE.

In [101], a method for improving epithelial tissue segmentation in H&E images using immunofluorescence (IF) guidance was presented. The authors used IF staining to create labels and then mapped them onto post-IF H&E images (terminal H&E). However, the IF process alters tissue appearance, making applying these labels to conventional H&E images hard. To fix this, the authors used a CycleGAN to transform conventional H&E images to terminal H&E. As a result, a segmentation model trained on terminal H&E could perform well on conventional H&E without extra human annotations.

They used nine WSIs from lung and breast cancer tissues to test the model. It performed well on terminal H&E with a Jaccard index of 0.733, but struggled with conventional H&E due to staining differences.

Similarly, Zeng et al. [102] introduced an automatic diagnostic framework for Nasopharyngeal Carcinoma (NPC) that digitally stains H&E images into Epstein-Barr Encoding Region (EBER) images. EBER staining is a technique for detecting Epstein-Barr Virus (EBV) in NPC diagnosis. They employed a Pathology–Fidelity GAN (PF-GAN) model to achieve this transformation. Compared to conventional GANs, PF-GAN employs pathology–fidelity constraints to ensure that the transformed images retain their original pathological features. This model excels at distinguishing between tumor and non-tumor regions while reducing issues like staining overflow and distortion, which often occur in traditional staining processes.

The data used for training and testing were obtained from Peking University Shenzhen Hospital and comprised 232 WSIs collected between 2020 and 2023. The performance of PF-GAN was evaluated against CycleGAN and CUT. PF-GAN achieved the best performance, with an RMSE of 0.01174, an SSIM of 0.8981, and a PSNR of 38.63. Additionally, when added into the overall diagnostic framework, PF-GAN demonstrated high classification accuracy, achieving an F1-score of 0.9143 and outperforming both traditional GANs and fully supervised models.

In a related work, Wang et al. [103] introduced CytoGAN, a GAN model designed for virtual staining in cytopathology images. They were able to use CytoGAN to transform H&E images to Papanicolaou staining and vice versa. The authors highlight that many existing stain transfer techniques fail to preserve the fine details of cellular structures while producing realistic images. CytoGAN takes a different approach by using a GAN with two important features: a structure preservation module and a stain adaptive module. The first keeps the layout of cells intact with a U-Net-like design and an attention mechanism. The adaptive module separates stain features from structural information, making it easier to transfer staining styles. While many GAN-based models struggle with cell size variations and background noise, CytoGAN maintains fine cytological details and adapts well to different imaging conditions.

To train and test the model, they used a dataset of WSIs from 139 patients, gathered over seven years. CytoGAN was compared with 11 other stain transfer models, including CycleGAN, StyleGAN, and StainGAN, and consistently produced better results in both structure preservation and stain accuracy. The evaluation included both numerical measurements and expert reviews from pathologists. CytoGAN reached the highest MSSIM (0.993) and PSNR (34.5) scores, making it the top performer among tested models. In an endometrial cancer classification task, it also helped improve accuracy by 20%.

Similarly, Hou et al. [104] developed a multi-model pipeline to generate virtual panCK stains from H&E-stained images for cancer diagnosis. The framework consists of three main models: Mask R-CNN, Marker-DGAN, and CK-DGAN. The pipeline begins with Mask R-CNN, which segments key structures in H&E images, identifying cytokeratin-rich cell membranes and cytoplasm. This segmentation is important because cytokeratin distribution determines panCK staining patterns. The segmented regions are then used as inputs into Marker-DGAN, a modified dual-CycleGAN model designed to reconstruct the cytokeratin distribution while preserving structural details. Marker-DGAN improves upon standard CycleGAN by employing structure-preserving loss, identity loss, and a novel dynamic loss to ensure that only cytokeratin-positive regions are stained. In the final step, CK-DGAN refines the virtual stain by adding hematoxylin information to improve clarity and make the output closely resemble real panCK-stained images. They explain that the use of these three models is necessary because virtual stains generated from a single model often fail to maintain cytokeratin specificity, leading to inaccurate staining patterns.

The dataset used for training and evaluation consisted of 45 WSIs from the HPH cohort and 1000 WSIs from the publicly available CAMELYON17 dataset. A total of 40,000 image patches were extracted from these WSIs for training and testing. The proposed method achieved FID = 157.74, SSIM = 0.681, and recall = 0.578, outperforming CycleGAN and other existing virtual stain models. Moreover, a blind test conducted by pathologists showed that virtual panCK stains were 97.8% diagnostically accurate, closely matching chemical panCK. In tumor budding assessment, the framework demonstrated 95.1% accuracy, compared to 24.4% for CycleGAN.

In a related work, Li et al. [105] introduced a weakly supervised framework for virtual immunohistochemical (IHC) staining to overcome limitations in existing deep learning-based staining methods. Deep learning models have been proposed to generate virtual stains, but they often struggle to preserve essential pathological features from consecutive tissue slices.

To address this, the researchers developed a multi-layer Weakly Pathological Consistency Constraint (WPCC) that employs consecutive slices to maintain accurate pathological relationships. Since different regions of a tissue slice may have varying levels of alignment, the model applies adaptive weights to adjust the supervision strength at different layers of the network. This ensures that well-aligned regions contribute more to training, while less-aligned areas have a reduced influence. Additionally, a major challenge in GAN-based models is the instability of the discriminator, which weakens over time and provides poor feedback to the generator. To solve this, the study introduced a discriminator contrastive regularization loss that allows the discriminator to better differentiate between real and generated images across multiple layers. Furthermore, PatchNCE loss helped the generator preserve local pathological features by enforcing consistency between patches in input H&E and generated IHC images.

The dataset used for training and evaluation consisted of two publicly available sources: BCI (Breast Cancer Immunohistochemical) and MIST. In total, the study utilized 20,090 images. On the BCI dataset, the proposed model achieved PSNR of 19.132, SSIM of 0.499, FID of 50.1, and KID of 9.8, outperforming CycleGAN, CUT, and Pyramid Pix2Pix in most metrics. Similarly, on the MIST dataset, the model outperformed existing methods across multiple stain types.

A method for high-resolution medical image transformation of H&E-stained images into Immunohistochemistry (IHC) stained images was presented in [106]. The proposed Patch alignment-based Paired medical image-to-image Translation (PPT) model addresses several key challenges in medical imaging, including the misalignment of paired images. This misalignment often occurs in clinical settings where images are obtained from adjacent tissue sections. Thus, to create image pairs, additional image registration algorithms are required. To overcome these issues, the authors employ a combination of techniques, including the FocalNCE loss for bidirectional contrastive learning. This enhances the consistency between the input H&E and the output IHC images. They also introduce a patch alignment loss to ensure better synchronization between paired image patches. Additionally, they employ content and frequency loss to refine the fine details in the IHC-stained images and achieve more accurate and high-quality results.

The dataset used in this study includes the first high-resolution paired H&E to IHC dataset of canine lymphoma, called the HIT dataset. It comprises two sets of paired images: H&E-CD3 and H&E-PAX5.

The model was tested on various datasets, including public breast cancer datasets. The results of these experiments show that the PPT model outperforms existing state-of-the-art methods. For instance, the PPT model achieved the highest PSNR (17.79), FID (78.54), and LPIPS (0.2988) on the PAX5 dataset. Furthermore, expert evaluation from veterinary pathologists confirmed the model's potential for clinical application.

In another study, Fan et al. [107] developed a methodology for virtual staining using Mueller Matrix Guided GAN (MMG-GAN). The model transformed H&E-stained tissue images into Masson Trichrome (MT)-stained images. MT staining offers better contrast, particularly for fibrous tissue structures, which makes it more effective than H&E staining for highlighting connective tissues. The model employs Mueller matrix Microscopy (MMM) polarization data to capture structural features not visible in traditional bright-field images.

MMG-GAN improves upon CycleGAN, which struggles with non-bijective staining relationships. This enhancement allows better learning of the complex mapping between H&E and MT stains. The model was trained on 14 sets of intestinal epithelial tissue slices, with 1400 H&E and MT images in total. The evaluation results revealed that MMG-GAN outperformed other models, achieving pixel similarity scores between 0.8 and 0.9.

Similarly, Guan et al. [108] presented GramGAN, an unsupervised multi-domain stain transfer model designed to virtually transform histopathological images from one stain to another while maintaining tissue structure. It transforms an image from one staining type (for example, standard H&E) to another by passing it through a series of Style-Guided blocks, each equipped with a style encoding dictionary that stores the characteristics of all staining styles. The architecture gradually adjusts the image style toward the target stain, and its multi-domain design allows a single network to handle multiple stain types without the need to train a separate model for every pair of stains.

The model was trained and tested on kidney and lung tissue datasets. It was also tested on the segmentation of the glomeruli. In qualitative and quantitative evaluations, experimental results showed that GramGAN outperformed other stain transfer models like CycleGAN, StarGAN, and UGATIT. It also improved glomeruli detection and segmentation. When H&E-stained images were virtually transformed into PASM-stained images, glomeruli segmentation accuracy increased, with mAP@[0.50:0.95] scores rising from 0.623 to 0.679 for detection and from 0.545 to 0.612 for segmentation. Domain shift analysis also confirmed that the virtual GramGAN stains closely matched real stains.

In a similar research, Li et al. [109] proposed Confusion-GAN, a method designed to transform stains from H&E to immunohistochemistry (IHC) staining. Confusion-GAN uses unsupervised learning which eliminates the need for paired images, often required in other methods. Furthermore, it focuses on preserving pathological information and texture details through two key innovations: a Patch-level Pathology Information Extractor (PPIE) and a Multi-Branch Discriminator (MBD). The PPIE, combined with Multiple Instance Learning (MIL), extracts essential features at the patch level for accurate staining. Moreover, a dual spherical loss is used to refine the classification of pathology patches. In addition, the MBD includes a confusion discriminator that forces the generated images to match the real distribution and improve visual fidelity.

The dataset used for training and testing consisted of images from three sources: the BCI dataset for HER2 staining, the MIST dataset for ER staining, and a newly introduced Hepatocellular Carcinoma (HCI) dataset. Results showed that Confusion-GAN outperformed other unsupervised methods like CycleGAN, UGATIT, and DCD, achieving an FID of 77.3, LPIPS of 0.570, and SSIM of 0.2300 on the HCI dataset. In classification experiments, using generated GPC3 images increased AUC from 55.6% to 88.9%. In addition, pathologists could not distinguish between real and generated images, which confirmed the high quality of the stains.

In a recent research, Denholm et al. [110] introduced a virtual histology technique that converts H&E-stained renal biopsy WSI into PAS-like images using a CycleGAN-based model trained on unpaired data from many centers. Model development includes a human-in-the-loop selection phase, in which visually promising training checkpoints are shortlisted, and an expert renal pathologist selects the final network based on the accuracy of key PAS features, thereby ensuring histological authenticity. The realism of the virtual stains is further supported by an "imitation game," in which pathologists have difficulty consistently distinguishing real from synthetic PAS images. Segmentation on virtually stained H&E slides produced Dice scores that are slightly reduced but comparable to those on real PAS, and the method is evaluated primarily against this real-stain baseline rather than through head-to-head comparison with alternative virtual staining approaches.

2.8. Hybrids

This section presents articles that perform two or more tasks, either within a single model or with more than one model.

For example, Meng et al. [111] developed a technique that performs both stain transformation and virtual staining. The main target was virtual staining using an improved Pix2Pix. However, in order to create the required image pairs, stain transformation was performed using CycleGAN. This transformed H&E images into their corresponding unstained autofluorescence images. The process created image pairs that were then used to train the Pix2Pix model. This method provides a new way to create a paired dataset.

The CycleGAN model was improved by introducing domain consistency loss and input buffers to enhance its ability to produce realistic unstained images. The team further optimized the generator using depth-wise separable convolution layers to eliminate common artifacts in synthetic images.

In contrast, their Pix2Pix model replaced the vanilla U-Net generator with a new Parallel Feature Fusion Network (PFFN) to better preserve tissue textures in the virtual staining process.

A dataset of 400 H&E-stained images and 80 autofluorescence images was used. Experiments showed improved performance over previous techniques. In addition, the virtual staining achieved 97% accuracy based on doctors' assessments and produced a 95% accuracy when tested on a VGG16 classification model.

Similarly, Yang et al. [112] proposed a Cascaded Deep Neural Network (C-DNN) framework for virtual stain transfer. However, before performing the transformation, they conducted a virtual staining process to obtain image pairs.

Thus, the model was designed to transform H&E-stained images into special stains such as PAS. However, to obtain image pairs required for training, they used a pretrained network to stain autofluorescence images into H&E. After the image pairs are obtained, a second network is then trained to transform H&E stains into PAS stains.

The dataset contained approximately 25,000 image patches, taken from autofluorescence and histochemically stained images of kidney tissue sections. The experiments show that C-DNN produced highly accurate PAS images with better structural similarity and color accuracy than standard methods.

Similarly, Vasiljević et al. [113] developed HistoStarGAN, a model for multiple stain transfer in histopathology. The model combines stain transfer, normalization, and stain-invariant segmentation in a single framework. It extends the StarGANv2 [44] model by adding a segmentation module to preserve biologically important structures across various stainings. The architecture comprises a generator, discriminator, mapping network, style encoder, and segmentation module, all trained together in an end-to-end manner. During training, the mapping network generates a stain-specific style by transforming a random latent code. This style is then injected into the generator to produce diverse stain translations. In addition, the segmentation module is trained in a supervised manner so that features such as glomeruli remain intact throughout transformations.

The training was based on a synthetic stain variation dataset created using CycleGAN. The dataset contains 512x512 pixel histopathological images from renal pathology, covering five staining types: PAS, Jones H&E, Sirius Red, CD34, and CD68.

The results showed that HistoStarGAN outperformed earlier models such as UDA-GAN in stain-invariant segmentation. Moreover, the model could generalize to unseen stainings without additional training. However, minor issues like artifacts and normalizations challenges were noted, particularly in synthetic data generation.

In related work, Franchet et al. [114] developed a technique for reducing stain-induced bias in AI-driven histological image classification. They developed AugmentHE, a stain augmentation technique, and HEnorm, a deep learning-based normalization method. AugmentHE separates hematoxylin and eosin channels, modifies their color and concentration, and then recombines them to generate augmented images. This approach expands the range of color appearances in training data, improving the robustness of models.

In contrast, HEnorm ensures color consistency across images before training. It employs a U-Net model trained on hematoxylin-extracted images (extracted from H&E) to reconstruct the original stain appearance while maintaining tissue structure. This normalization process aligns images with a reference stain, preventing the model from relying on color variations. The method is speedy, processing images up to 78 times faster than conventional techniques, making it suitable for large-scale digital pathology applications.

The dataset used for the evaluation was the histological classification of breast cancer. Results showed that models trained with AugmentHE and HEnorm achieved better classification accuracy than conventional approaches. AugmentHE increased color dispersion by 81%, reducing the visibility of staining differences between datasets. The best performing models (using the HEnorm combined with AugmentHE and Vahadane's normalization combined with AugmentHE) reached an AUC of up to 0.85, outperforming standard augmentation and normalization methods.

3. Datasets and trends of research

In this section, we present details of some famous benchmark datasets, outline key research trends, and pinpoint specific gaps in the literature that remain unaddressed.

3.1. Famous datasets

This section summarizes several widely used datasets for benchmarking digital stain processing models. Table 5 provides an overview of some publicly available datasets commonly used in stain processing research.

1. CAMELYON16/17: These datasets are collections of H&E WSIs of lymph node sections [115]. These datasets were designed to detect breast cancer metastases in lymph nodes. Moreover, the dataset is used for metastasis segmentation and the development of computer-aided diagnosis tools, such as stain normalization and digital staining.

   Over time, the dataset has evolved to become more comprehensive. CAMELYON16 was released in 2016 and contained 399 WSIs, with pixel-level annotations for metastases. It was the first large publicly available dataset with expert annotations. In contrast, the CAMELYON17, released in 2017, expanded on this by providing 1000 WSIs from five medical centers.

   Moreover, the CAMELYON16 provides pixel-level annotations of metastases, which support tumor segmentation tasks. In

Table 5
Comparison of datasets for stain processing.

Dataset	Tissue type	Staining type	Data size	Resolution	Image format
CAMELYON16/17 [115]	Breast (lymph node metastases)	H&E	1400 WSIs total (CAMELYON16: 400 WSIs from 2 centers; CAMELYON17: 1000 WSIs from 5 centers)	40x; pixel size 0.243 μm/pixel	Whole Slide Images (WSIs), format: .tif
MITOS-ATYPIA-14 [26]	Breast tissue	H&E	1420 image patches	20× and 40×; each patch is 2048×2048 px; 0.23–0.25 μm/pixel	Microscopy image patches extracted from WSIs format: .tif)
MoNuSeg [116]	Breast, liver, kidney, prostate, bladder, colon, stomach, lung, brain	H&E	44 image patches (30 training, 14 testing), totaling 28,846 annotated nuclei	40x; each patch is 1000×1000 px	Image patches from WSIs; format: .tif
MIST [117]	Breast carcinoma	H&E and IHC (HER2, ER, PR, Ki67)	For each biomarker, >4000 train + 1000 test aligned H&E–IHC image pairs	20x; each patch is 1024×1024 px; 0.4661 μm/pixel	Aligned H&E–IHC image patches from WSIs; format: .png
NCT-CRC-HE-100K [118]	Colorectal cancer and normal colorectal tissue	H&E	100,000 non-overlapping patches in 9 tissue classes	20x; each patch is 224×224 px; 0.5 μm/pixel	Image patches from WSIs; format: .tif
BCI [119]	Breast carcinoma tissue (HER2)	Paired H&E and HER2 IHC	4870 paired H&E–HER2 IHC patches	20×; each patch is 1024×1024 px; 0.46 μm/pixel	Image patches from WSIs; format: .png

contrast, CAMELYON17 focuses on slide-level labels to evaluate model performance across various clinical settings.

One advantage of CAMELYON16/17 is its size and quality. The pixel-level annotations performed by expert pathologists that it contains are rare in histopathology datasets. This makes it a valuable resource for training and validating deep learning models.

However, it has some challenges. The WSIs are very large and may pose computational challenges in data processing and model training. In addition, class imbalance, where metastatic regions are much smaller than normal tissues, can complicate model learning. Also, despite being multi-center, its staining protocols and scanner types are still limited compared to the full range of clinical variability, which can affect the robustness of the model when applied to unseen data.

2. MITOS-ATYPIA-14 [26]: This dataset contains labeled images of breast tissue stained with H&E. It includes examples of mitotic figures and nuclear atypia, which are useful for training and testing computer models that detect cell division and abnormal cellular features in cancer diagnosis.

   According to the official dataset description, MITOS-ATYPIA-14 comprises 284 frames at 20× magnification and 1136 frames at 40× magnification. The images were acquired using two different whole-slide scanners: the Aperio ScanScope XT and the Hamamatsu NanoZoomer. Using different scanners at multiple magnifications offers a natural variation in image quality and appearance that reflects what is commonly seen in clinical practice.

   However, the dataset has its limitations. Although it offers many images, the total count may still be insufficient for training highly complex deep learning models without the risk of overfitting. Additionally, the annotations, although detailed, are subject to inter-observer variability, which can introduce noise into the training process.

3. MoNuSeg: This dataset was introduced to enhance the development of robust and generalizable algorithms for nuclei instance segmentation in histopathology images [116]. It consists of H&E-stained image patches from The Cancer Genome Atlas (TCGA) and covers various organs and disease types. MoNuSeg is mainly used for nuclei instance segmentation, which identifies and delineates individual nuclei in complex tissue images. It is also suitable for related applications such as digital stain normalization and testing model generalization across tissues and centers.

   The dataset was first released alongside the MoNuSeg 2018 challenge, comprising 30 training images with 21,623 annotated nuclei from 7 organs, and 14 test images including two unseen organs, summing to 44 images with over 28,000 labeled nuclei. Since then, no major expansion of the dataset has been reported, although it has inspired further studies.

   What sets MoNuSeg apart from many other datasets is its diversity in organ types and scanning centers, which makes it more applicable for training general-purpose segmentation models.

   However, on the other hand, it is relatively small, especially regarding image count and size, compared to modern deep learning standards. Although it spans multiple organs, each organ is represented by only a few samples, which may limit model resilience in cross-organ generalization.

3.2. Trends

After reviewing several papers on the topic, we were able to find certain recurring patterns, some of which are summarized in Table A.6. Fig. 5 illustrates the distribution of articles based on the taxonomy presented in Section 2.1. As shown, most papers focus on stain transformation tasks (40%), closely followed by virtual staining (38%).

In addition, in Fig. 6, we present the most commonly used model for digital staining. As explained in Section 2.2, the most used is

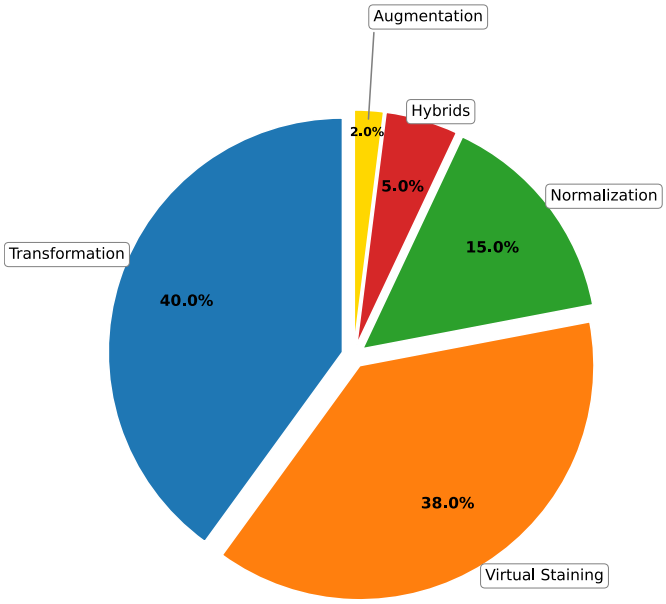


Fig. 5. The distribution of papers across five classes.

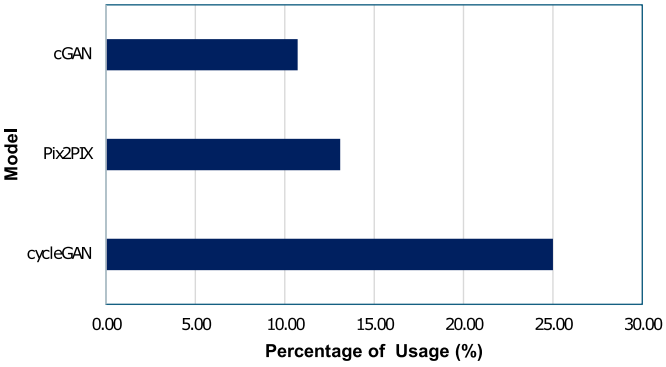


Fig. 6. The famous models used and their percentage of usage.

CycleGAN (25%), followed by Pix2Pix (13%), and cGAN (10.7%). Other identifiable trends are listed as follows.

1. Combining Digital Stain Processing with Downstream Tasks: The ultimate test for deep learning-based digital stain processing models is their performance on downstream tasks such as cancer detection, prediction, and segmentation. Thus, most reviewed studies included downstream task evaluation in their analysis. For models focusing on stain normalization or augmentation, downstream performance is often the only practical evaluation method. In contrast, virtual staining and stain transformation models are typically evaluated using both downstream performance and additional image similarity metrics, as outlined in Section 2.3.
2. Expert-Assessed Evaluation: We also observed that many papers involve experienced pathologists in evaluating model performance. This approach resembles a human-in-the-loop system. For example, a pathologist may be presented with both AI-generated and chemically stained slides. A well-performing deep learning model should ideally make it difficult for the pathologist to distinguish between them. Metrics such as Cohen's Kappa score and SSIM are commonly used to measure the agreement and perceptual similarity between real and virtual stains.

3. Use of new loss functions: As mentioned in Section 2.2, the basic loss functions used in GANs are adversarial loss, cycle consistency loss, and pixel reconstruction loss. However, using only these three often results in hallucinations and artifacts. As a result, researchers have explored other variants of loss functions to add to the mix. These loss functions include the perceptual loss used in [28], which helps preserve important structural features by comparing high-level representations instead of pixel-wise values. Another loss function is the identity loss used in [41] to ensure color fidelity and maintain input–output consistency. Other loss functions are the Structural Similarity (SSIM) loss, style loss, and patch-wise contrastive loss [12,25,49].
4. Common Generators and discriminators: From the reviewed articles, we observed that the most commonly used generator architecture in GAN-based models is the U-Net. This model uses an encoder–decoder structure with skip connections directly linking corresponding layers between the downsampling and upsampling paths. These skip connections are useful for preserving spatial resolution and fine tissue structures, which are important in histopathology. However, they can also carry over background noise or imaging artifacts, especially if the early features are not properly filtered.

   PatchGAN is often chosen for discriminator design. Instead of judging the image as a whole, PatchGAN evaluates small overlapping patches. This allows the model to emphasize texture and fine-grained details in local regions. Although this approach helps improve visual quality at the patch level, it may miss broader structural coherence across large tissue areas, affecting consistency in medical interpretation.
5. Use of proprietary data: Most articles surveyed employed local datasets, often collected from hospitals, alongside public benchmark data to evaluate model performance. From the surveyed papers, 80% of them use such private data either as a standalone source or in combination with publicly available datasets. Additionally, 70% of the surveyed papers strictly use private information that is not available online.

   This dual approach (using public and private data) provides readers with the benefit of understanding how models perform on standardized datasets and in real-world, small-scale clinical settings. However, these local datasets are frequently not publicly accessible, posing a reproducibility challenge. As a result, researchers with limited resources may be confined to using only public datasets, which can limit their ability to directly compare their models with others in the literature.

4. Open research challenges

This section presents some of the less explored areas in digital stain processing.

1. Explainability: This is important in digital stain processing because it directly impacts the trust and adoption of AI-powered tools in clinical practice. Pathologists need to understand how and why a model makes certain predictions. Despite its importance, only a few works have explored explainability in digital staining [58,61,94]. In future studies, researchers should focus on improving the transparency and interpretability of models. One approach could involve developing post-hoc explainability techniques like saliency maps or activation heatmaps to visualize which parts of the input image contributed most to the output. Moreover, interactive models that allow pathologists to query the system and receive understandable explanations for predictions could improve trust.
2. Employment of Diffusion Models: These models are a class of generative models that iteratively transform noise into structured data through gradual denoising. They have recently gained attention for their ability to generate high-quality images in various domains, including digital pathology [120]. Compared to GANs, diffusion models offer certain advantages regarding image quality and training stability [121,122]. Despite their potential, diffusion models face challenges. They require large and diverse datasets to achieve optimal performance. This can be a limitation in the medical domain, where annotated data is often scarce. Additionally, they need huge computational resources for training and inference. Future research direction should focus on improving efficiency and reducing the computational burden. One approach is using domain-specific knowledge, which involves incorporating expert insights about tissue characteristics and staining patterns into the model. This can guide the model to learn more efficiently with smaller datasets. Additionally, transfer learning using pre-trained models from similar domains can also be employed to reduce the need for extensive data and computational power.
3. Absence of Real-Time or In-surgery Deployment Studies: This remains a research gap in digital stain processing. In clinical settings such as cancer surgery or biopsy assessment, timely decisions depend on immediate tissue evaluation. Thus, fast and reliable virtual staining is highly valuable. However, only a few studies have attempted to address this need, [109,123]. Most models are still designed for offline retrospective analysis. Future studies should focus on developing lightweight and low-latency architectures suitable for deployment on clinical imaging devices and validating their performance in real-time workflows under surgical or diagnostic constraints.
4. Inadequate Quality Detection: Issues such as hallucinations and artifacts remain a significant challenge in digital stain processing. Virtually stained images are often used in clinical decision-making. Therefore, undetected distortions or fabricated structures can lead to misinterpretation and potential diagnostic errors [124]. Although the risk is well-recognized, only a few studies have addressed it directly. For example, Huang et al. [125] proposed AQuA as a framework to quantify anatomical and visual accuracy.

   In contrast, many other studies continue to rely on subjective evaluation by experts. Future work should focus on developing standardized automated methods for hallucination and artifact detection. Additionally, evaluation metrics that account for both visual quality and clinical reliability should be developed.
5. Limited Multi-modal Integration: The employment of multi-modal data, such as proteomics and cellular-level features, is essential for advancing machine learning-based stain processing. Recent studies have demonstrated that combining histopathological images with proteogenomic or text data can enhance predictive accuracy in identifying histological patterns associated with disease progression [126–128]. However, works in this area are scarce. Researchers should focus on developing robust algorithms that seamlessly employ diverse data types, such as genomic or textual pathology reports, with imaging data. This will provide more precise and holistic pathological assessments.
6. Federated Learning (FL): Although current approaches to digital stain processing predominantly rely on centralized machine learning models, there is a notable gap in exploring the potential of federated learning in this domain. FL supports privacy-preserving, decentralized learning, making it valuable in digital pathology, where data security and patient confidentiality are essential. In contrast to traditional methods, FL enables the collaboration of multiple institutions or healthcare providers without sharing sensitive medical data, thus facilitating the development of more robust and generalized virtual staining models. Recent studies in FL have shown promising applications in areas such as medical imaging [129–131], but its use in the virtual staining pipeline remains underexplored.

5. Conclusions

This paper reviews works that employ deep learning, mostly generative adversarial networks, for processing stains in pathology. We define stain processing as the tasks performed before downstream routines such as classification and segmentation. The articles analyzed show that using GANs for tissue staining offers several advantages over traditional chemical methods.

Our paper also presented a taxonomy of the field by classifying the works into normalization, augmentation, virtual staining, stain transformation, and hybrids. Moreover, we outlined key trends in digital stain research, which include: combining stain processing with downstream tasks, expert-assessed evaluation by pathologists, use of novel loss functions (such as perceptual and identity losses) to reduce artifacts, the dominance of U-Net generators paired with PatchGAN discriminators, and reliance on both benchmark and proprietary datasets.

Finally, we identified future research directions in digital stain processing. These include improving model explainability through saliency maps and interactive tools, employing diffusion models, developing lightweight architectures for real-time or intraoperative deployment, creating automated quality detection methods to flag hallucinations and artifacts, and employing multi-modal data such as spatial proteomics and genomic profiles.

CRediT authorship contribution statement

Rabiah Al-Qudah: Writing – review & editing, Formal analysis. **Abubakar Bala:** Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Mrouj Almuhajri:** Writing – review & editing, Visualization, Formal analysis. **Khiati Zakaria:** Writing – review & editing, Investigation. **Ching Y. Suen:** Writing – review & editing, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Appendix section

Table A.6
This table summarizes the studies that employ ML for stain processing. In the "Data Avail." column, we indicate a letter "H" if only part of the data is available online (or if the authors require emails and forms to be sent before sharing their data). A letter "F" if the full data used is available online, and a letter "N" if the data is not available online. The dataset weblinks were accessed on March 1, 2025.

Method	Model	Task	Data avail.	Data link	Dataset size
Ziaei et al. 2020 [21]	stainGAN	Normalization	F	https://mitos-atypia-14.grand-challenge.Org/Donwload/	310 H&E patches
Salehi et al. 2020 [22]	Pix2Pix	Normalization	F	https://mitos-atypia-14.grand-challenge.Org/Dataset/	12,720 H&E patches
Zheng et al. 2020 [23]	CapsNet	Normalization	H	https://camelyon17.grand-challenge.Org/Data/	1223 slides
Patil et al. 2021 [11]	DenseNet	Normalization	H	https://camelyon17.grand-challenge.Org/ & https://monuseg.grand-challenge.Org/	530 WSIs
Runz et al. 2021 [24]	CycleGAN	Normalization	F	https://heidata.uni-heidelberg.de/dataset.xhtml?persistentId=doi:10.11588/data/8LKEZF	14,114 tiles
Mao et al. 2021 [25]	SGNet	Normalization	N		49,000 patches
Perez et al. 2022 [12]	StainCUT	Normalization	F	https://mitos-atypia-14.grand-challenge.Org/Donwload/	424 slides
Zhao et al. 2022 [13]	RestainNet	Normalization	N		424 images
Kweon et al. 2023 [132]	MS-SST	Normalization	N		
Nazki et al. 2023 [28]	MultiPathGAN	Normalization	N		4350 patches
Jeong et al. 2023 [29]	SBD	Normalization	H	https://paip2019.grand-challenge.Org/Dataset/ & https://camelyon17.grand-challenge.Org/Data/	
Hetz et al. 2024 [30]	CycleGAN	Normalization	F	https://zenodo.org/records/7418555 & https://camelyon17.grand-challenge.Org/Data/	2M slides
Du et al. 2025 [31]	DSTGAN	Normalization	F	Four dataset links are provided in the original paper	≈ 100K patches
Wagner et al. 2021 [32]	HistAuGAN	Augmentation	F	https://camelyon17.grand-challenge.Org/Data/	50 WSIs
Vasiljevic et al. 2021 [33]	CycleGAN, starGAN	Augmentation	N		5296 patches
Bayat et al. 2021 [34]	cGAN	Virtual staining	N		46 WSIs
Li et al. 2021 [35]	Pix2Pix	Virtual staining	N		1521 images
Kaza et al. 2021 [77]	cGAN	Virtual Staining	N		35,000 patches
Kaza et al. 2022 [36]	cGAN	Virtual Staining	N		400 WBCs
Zhang et al. 2022 [86]	Pix2Pix	Virtual Staining	N		140 WSIs
Zhang et al. 2022 [38]	GAN	Virtual Staining	N		5832 tissues
Kaza et al. 2022 [37]	cGAN	Virtual Staining	N		77,600 patches
Soltani et al. 2022 [39]	UTOM	Virtual Staining	N		146,003 patches
Boktor et al. 2023 [40]	Pix2Pix	Virtual Staining	N		21,000 patches
Salido et al. 2023 [52]	CycleGAN	Virtual Staining	N		13,436 pathes
Martell et al. 2023 [41]	CycleGAN	Virtual Staining	F	https://zenodo.org/records/7981075	36,000 tissues
Tomczak et al. 2023 [42]	GAN	Virtual Staining	N		14,036 images

(continued on next page)

Table A.6 *(continued)*

Method	Model	Task	Data avail.	Data link	Dataset size
Abraham et al. 2023 [45]	CycleGAN	Virtual Staining	F	https://github.com/tmabraham/qOBMtoHE	5436 tiles
Shi et al. 2023 [46]	MulHiST	Virtual Staining	N		11,643 patches
Min et al. 2023 [47]	GAN	Virtual Staining	N		420,000 patches
Khan et al. 2023 [48]	Pix2Pix	Virtual Staining	F	https://etsin.fairdata.fi/dataset/4be10dd9-b0f8-401a-a74b-7dfe154d0b38/data	81 slides
Martell et al. 2023 [49]	CycleGAN	Virtual Staining	N		220 patches
Wang et al. 2023 [50]	UBMSVStain	Virtual Staining	N		12,294 patches
Koivukoski et al. 2023 [51]	CycleGAN, Pix2Pix	Virtual Staining	F	https://etsin.fairdata.fi/dataset/46d0ef41-a101-49c7-91fd-b5579be0958d/data	66,000 tiles
Fang et al. 2024 [54]	CellGAN	Virtual Staining	N		7319 cells
Koka et al. 2024 [55]	cGAN	Virtual Staining	N		40 images
Wong et al. 2024 [56]	Pix2Pix	Virtual Staining	N		1159 image pairs
Zhu et al. 2024 [57]	Pix2Pix	Virtual Staining	N		627 image pairs
Yoon et al. 2024 [58]	E-CUT	Virtual Staining	N		100 tiles
Asaf et al. 2024 [59]	DCLGAN	Virtual Staining	H	https://biomisa.org/index.php/downloads/	3824 patches.
Fan et al. 2024 [60]	CycleGAN	Virtual Staining	N		27,000 patches
Chen et al. 2024 [61]	U-Frame	Virtual Staining	N		117 image pairs
Boktor et al. 2024 [62]	CycleGAN	Virtual Staining	N		2500 patches
Li et al. 2024 [63]	SCPAT-GAN	Virtual Staining	N		8594 patches
Azam et al. 2024 [64]	Pix2Pix	Virtual Staining	N		12,058 patches
Yang et al. 2024 [65]	cGAN	Virtual Staining	N		451 images
Xu et al. 2025 [66]	CycleGAN	Virtual Staining	F	https://github.com/wwxb2012/digital_staining_knowledge_distillation	2020 patches
Li et al. 2025 [67]	cGAN	Virtual Staining	N		≈59,758 images
Zhang et al. 2025 [68]	Diffusion model	Virtual Staining	N		1549 images
Lui et al. 2025 [69]	Cytoland	Virtual Staining	N		
Park et al. 2025 [70]	cGAN	Virtual Staining	F	https://doi.org/10.57760/sciencedb.24217	≈ 230K patches
Mercan et al. 2020 [71]	cGAN & CycleGAN	Transformation	N		18 WSIs
Burlingame et al. 2020 [72]	cGAN	Transformation	F	https://gitlab.com/eburling/SHIFT	12,258 H&E and IF tiles
Wang et al. 2020 [73]	CycleGAN	Transformation	N		5500 images
Levy et al. 2021 [74]	CycleGAN	Transformation	N		574 WSIs
Picon et al. 2021 [75]	DenseNet	Transformation	N		1755 image tiles
Lo et al. 2021 [76]	CycleGAN	Transformation	N		4318 images
Thebille et al. 2021 [78]	CycleGAN, Fixed-Point GAN	Transformation	F	https://github.com/imsb-uke/bias-transfer-microscopy	4396 images
de Haan et al. 2021 [79]	CycleGAN with others	Transformation	F	https://github.com/kevindehaan/stain-transformation	7800 patches
Hong et al. 2021 [80]	Pix2Pix	Transformation	N		371 images
Wieslander et al. 2021 [81]	LUPI	Transformation	H	https://www.ai.se/en/project/adipocyte-cell-imaging-challenge	192 images
Lahiani et al. 2021 [82]	CycleGAN	Transformation	N		50 WSIs
Liu et al. 2021 [83]	PC-StainGAN	Transformation	N		716 images
Cetin et al. 2022 [84]	StainGAN	Transformation	N		38 WSIs
Lysik et al. 2022 [85]	Pix2Pix, CycleGAN	Transformation	N		10 HE-PAS slides
Zhang et al. 2022 [89]	MVFStain	Transformation	N		28,568 patches
Zeng et al. 2022 [90]	Semi-Superv.	Transformation	N		30 slides
Zingman et al. 2023 [92]	CycleGAN	Transformation	F	https://osf.io/byf27/	54,600 tiles
Duan et al. 2023 [93]	GAN	Transformation	N		3896 images
Sloboda et al. 2023 [94]	xAI-CycleGAN	Transformation	N		110,564 slices
Biswas et al. 2023 [96]	CycleGAN	Transformation	N		23 images
Ling et al. 2023 [98]	SDN	Transformation	F	https://mitos-atypia-14.grand-challenge.Org/Dataset/ & https://camelyon17.grand-challenge.Org/Data/	1016 slides
Yan et al. 2023 [99]	PG-GAN	Transformation	N		470,147 patches
Liu et al. 2023 [100]	cGAN	Transformation	N		5160 images
Wiedenmann et al. 2024 [101]	CycleGAN	Transformation	N		9 WSIs
Zeng et al. 2024 [102]	PF-GAN	Transformation	N		232,303 patches.
Wang et al. 2024 [103]	CytoGAN	Transformation	N		139 WSIs
Hou et al. 2024 [104]	Mask R-CNN, GANs	Transformation	H	https://camelyon17.grand-challenge.Org/Data/	52,000 patches
Li et al. 2024 [105]	GAN	Transformation	H	https://drive.google.com/drive/folders/146V99Zv1LzoHFYlXvSDhKmfIIL-joo6p	20,090 images
Zhang et al. 2024 [106]	PPT	Transformation	F	https://github.com/coffeeNtv/PPT	3584 image pairs

(*continued on next page*)

Table A.6 *(continued)*

Method	Model	Task	Data avail.	Data link	Dataset size
Fan et al. 2024 [107]	MMG-GAN	Transformation	N		3080 images
Guan et al. 2024 [108]	GramGAN	Transformation	H	https://dx.doi.org/10.21227/p7pw-y957	83,027 patches
Li et al. 2024 [109]	Confusion-GAN	Transformation	N		100 WSIs
Denholm et al. 2025 [110]	CycleGAN	Transformation	N		≈2140 images
Meng et al. 2021 [111]	CycleGAN, cGAN	Virtual Staining, Transformation	N		480 images
Yang et al. 2022 [112]	C-DNN	Virtual Staining, Transformation	N		25,040 patches
Vasiljević et al. 2023 [113]	HistoStarGAN	Transformation, Virtual Staining	N		30,249 patches
Franchet et al. 2024 [114]	U-Net	Augmentation, Normalization	N		64,000 images

Data availability

No data was used for the research described in the article.

References

[1] S. Agrawal, P. Oza, R. Kakkar, S. Tanwar, V. Jetani, J. Undhad, et al., Analysis and recommendation system-based on prisma checklist to write systematic review, Assess. Writ. 61 (2024) 100866.

[2] Y. Rivenson, K. de Haan, W.D. Wallace, A. Ozcan, Emerging advances to transform histopathology using virtual staining, BME Frontiers 2020 (2020) 9647163.

[3] L. Jose, S. Liu, C. Russo, A. Nadort, A. Di Ieva, Generative adversarial networks in digital pathology and histopathological image processing: a review, J. Pathol. Inform. 12 (1) (2021) 43.

[4] D. Morrison, D. Harris-Birtill, P.D. Caie, Generative deep learning in digital pathology workflows, Am. J. Pathol. 191 (10) (2021) 1717–1723.

[5] N. Pillar, A. Ozcan, Virtual tissue staining in pathology using machine learning, Expert Rev. Mol. Diagn. 22 (11) (2022) 987–989.

[6] A. Waqas, M.M. Bui, E.F. Glassy, I. El Naqa, P. Borkowski, A.A. Borkowski, et al., Revolutionizing digital pathology with the power of generative artificial intelligence and foundation models, Lab. Investig. (2023) 100255.

[7] B. Bai, X. Yang, Y. Li, Y. Zhang, N. Pillar, A. Ozcan, Deep learning-enabled virtual histological staining of biological samples, Light: Sci. Appl. 12 (1) (2023) 57.

[8] L. Kreiss, S. Jiang, X. Li, S. Xu, K.C. Zhou, K.C. Lee, et al., Digital staining in optical microscopy using deep learning-a review, PhotoniX 4 (1) (2023) 34.

[9] L. Latonen, S. Koivukoski, U. Khan, P. Ruusuvuori, Virtual staining for histology by deep learning, Trends Biotechnol. 42 (9) (2024) 1177–1191.

[10] W. Lin, Y. Hu, R. Zhu, B. Wang, L. Wang, Virtual staining for pathology: challenges, limitations and perspectives, Intell. Oncol. 1 (2) (2025) 105–119.

[11] A. Patil, M. Talha, A. Bhatia, N.C. Kurian, S. Mangale, S. Patel, et al., Fast, self supervised, fully convolutional color normalization of H&E stained images, in: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), IEEE, 2021, pp. 1563–1567.

[12] J.C.G. Pérez, D.O. Baguer, P. Maass, StainCUT: stain normalization with contrastive learning, J. Imaging. 8 (7) (2022).

[13] B. Zhao, C. Han, X. Pan, J. Lin, Z. Yi, C. Liang, et al., RestainNet: a self-supervised digital re-stainer for stain normalization, Comput. Electr. Eng. 103 (2022) 108304.

[14] M. Mirza, S. Osindero, Conditional generative adversarial nets, arXiv preprint arXiv:1411.1784, 2014.

[15] P. Isola, J.-Y. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1125–1134.

[16] P. Fischer, T. Brox, et al., U-Net: convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, vol. 9351, Springer, 2015, pp. 234–241.

[17] C. Li, M. Wand, Precomputed real-time texture synthesis with markovian generative adversarial networks, in: European Conference on Computer Vision, Springer, 2016, pp. 702–716.

[18] J.-Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2223–2232.

[19] R. Zhang, P. Isola, A.A. Efros, E. Shechtman, O. Wang, The unreasonable effectiveness of deep features as a perceptual metric, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 586–595.

[20] Y. Zhang, L. Huang, N. Pillar, Y. Li, Y. Li, L.G. Migas, et al., Virtual staining of label-free tissue in imaging mass spectrometry, Sci. Adv. 11 (31) (2025) eadv0741.

[21] D. Ziaei, W. Li, S. Lam, W.-C. Cheng, W. Chen, Characterization of color normalization methods in digital pathology whole slide images, in: Medical Imaging 2020: Digital Pathology, vol. 11320, SPIE, 2020, pp. 290–297.

[22] P. Salehi, A. Chalechale, Pix2Pix-based stain-to-stain translation: a solution for robust stain normalization in histopathology images analysis, in: 2020 International Conference on Machine Vision and Image Processing (MVIP), IEEE, 2020, pp. 1–7.

[23] Y. Zheng, Z. Jiang, H. Zhang, F. Xie, D. Hu, S. Sun, et al., Stain standardization capsule for application-driven histopathological image normalization, IEEE J. Biomed. Health Inform. 25 (2) (2020) 337–347.

[24] M. Runz, D. Rusche, S. Schmidt, M.R. Weihrauch, J. Hesser, C.-A. Weis, Normalization of HE-stained histological images using cycle consistent generative adversarial networks, Diagn. Pathol. 16 (2021) 1–10.

[25] X. Mao, J. Wang, X. Tao, Y. Wang, Q. Li, X. Zhou, et al., Single generative networks for stain normalization and quality enhancement of histological images in digital pathology, in: 2021 14th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), IEEE, 2021, pp. 1–5.

[26] G. Challenge, Mitos-Atypia 2014 challenge dataset, 2014, https://mitos-atypia-14.grand-challenge.org/Donwload/, (Accessed: 21 May 2025).

[27] B. Ehteshami Bejnordi, M. Veta, P.J. van Diest, B. van Ginneken, N. Karssemeijer, G. Litjens, et al., CAMELYON16 challenge, 2016, https://camelyon16.grand-challenge.org/, (Accessed: 24 March 2025).

[28] H. Nazki, O. Arandjelovic, I.H. Um, D. Harrison, MultiPathGAN: structure preserving stain normalization using unsupervised multi-domain adversarial network with perception loss, in: Proceedings of the 38th ACM/SIGAPP Symposium on Applied Computing, 2023, pp. 1197–1204.

[29] J. Jeong, K.D. Kim, Y. Nam, C.E. Cho, H. Go, N. Kim, Stain normalization using score-based diffusion model through stain separation and overlapped moving window patch strategies, Comput. Biol. Med. 152 (2023) 106335.

[30] M.J. Hetz, T.-C. Bucher, T.J. Brinker, Multi-domain stain normalization for digital pathology: a cycle-consistent adversarial network for whole slide images, Med. Image Anal. 94 (2024) 103149.

[31] Z. Du, P. Zhang, X. Huang, Z. Hu, G. Yang, M. Xi, et al., Deeply supervised two stage generative adversarial network for stain normalization, Sci. Rep. 15 (1) (2025) 7068.

[32] S.J. Wagner, N. Khalili, R. Sharma, M. Boxberg, C. Marr, W. De Back, et al., Structure-preserving multi-domain stain color augmentation using style-transfer with disentangled representations, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VIII 24, Springer, 2021, pp. 257–266.

[33] J. Vasiljević, F. Feuerhake, C. Wemmert, T. Lampert, Towards histopathological stain invariance by unsupervised domain augmentation using generative adversarial networks, Neurocomputing 460 (2021) 277–291.

[34] A. Bayat, C. Anderson, P. Shah, Automated end-to-end deep learning framework for classification and tumor localization from native non-stained pathology images, in: Medical Imaging 2021: Image Processing, vol. 11596, SPIE, 2021, pp. 43–54.

[35] J. Li, J. Garfinkel, X. Zhang, D. Wu, Y. Zhang, K. De Haan, et al., Biopsy-free in vivo virtual histology of skin using deep learning, Light: Sci. Appl. 10 (1) (2021) 233.

[36] N. Kaza, A. Ojaghi, F.E. Robles, Automated virtual staining, segmentation and classification of deep ultraviolet (UV) microscopy images for hematological analysis, in: Microscopy Histopathology and Analytics, Optica Publishing Group, 2022, pp. MW4A–5.

[37] N. Kaza, A. Ojaghi, F.E. Robles, Virtual staining, segmentation, and classification of blood smears for label-free hematology analysis, BME Front. 2022 (2022) 9853606.

[38] Y. Zhang, L. Huang, T. Liu, K. Cheng, K. de Haan, Y. Li, et al., Virtual staining of defocused autofluorescence images of unlabeled tissue using deep neural networks, Intell. Comput. (2022).

[39] S. Soltani, A. Ojaghi, H. Qiao, N. Kaza, X. Li, Q. Dai, et al., Prostate cancer histopathology using label-free multispectral deep-UV microscopy quantifies phenotypes of tumor aggressiveness and enables multiple diagnostic virtual stains, Sci. Rep. 12 (1) (2022) 9329.

[40] M. Boktor, B.R. Ecclestone, V. Pekar, D. Dinakaran, J.R. Mackey, P. Fieguth, et al., Virtual histological staining of label-free total absorption photoacoustic remote sensing (TA-PARS), Sci. Rep. 12 (1) (2022) 10296.

[41] M.T. Martell, N.J.M. Haven, B.D. Cikaluk, B.S. Restall, E.A. McAlister, R. Mittal, et al., Deep learning-enabled realistic virtual histology with ultraviolet photoacoustic remote sensing microscopy, Nat. Commun. 14 (1) (2023) 5967.

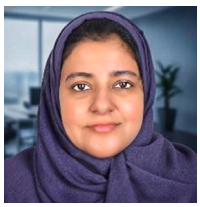
[42] A. Tomczak, S. Ilic, G. Marquardt, T. Engel, N. Navab, S. Albarqouni, Digital staining of white blood cells with confidence estimation, IEEE Trans. Med. Imaging 42 (12) (2023) 3895–3906.
[43] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, B. Catanzaro, High-resolution image synthesis and semantic manipulation with conditional GANs, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 8798–8807.
[44] Y. Choi, Y. Uh, J. Yoo, J.-W. Ha, StarGAN v2: diverse image synthesis for multiple domains, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 8188–8197.
[45] T.M. Abraham, P.C. Costa, C. Filan, Z. Guang, Z. Zhang, S. Neill, et al., Label- and slide-free tissue histology using 3D epi-mode quantitative phase imaging and virtual hematoxylin and eosin staining, Optica 10 (12) (2023) 1605–1618.
[46] L. Shi, Y. Zhang, I.H.M. Wong, C.T.K. Lo, T.T.W. Wong, MulHiST: multiple histological staining for thick biological samples via unsupervised image-to-image translation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2023, pp. 735–744.
[47] E. Min, N. Aimakov, S. Lee, S. Ban, H. Yang, Y. Ahn, et al., Multi-contrast digital histopathology of mouse organs using quantitative phase imaging and virtual staining, Biomed. Opt. Express 14 (5) (2023) 2068–2079.
[48] U. Khan, S. Koivukoski, M. Valkonen, L. Latonen, P. Ruusuvuori, The effect of neural network architecture on virtual H&E staining: systematic assessment of histological feasibility, Patterns 4 (5) (2023).
[49] M.T. Martell, N.J.M. Haven, E.A. McAlister, B.S. Restall, B.D. Cikaluk, R. Mittal, et al., Ultraviolet photoacoustic remote sensing and scattering microscopy for CycleGAN-enabled realistic virtual histology, in: European Conference on Biomedical Optics, Optica Publishing Group, 2023, p. 126310C.
[50] A. Wang, R. Zhu, Unpaired bone marrow smears virtual staining using content and attention-guided generative adversarial networks: UBMSVStain using content and attention-guided generative adversarial networks, in: Proceedings of the 2023 4th International Symposium on Artificial Intelligence for Medicine Science, 2023, pp. 676–681.
[51] S. Koivukoski, U. Khan, P. Ruusuvuori, L. Latonen, Unstained tissue imaging and virtual hematoxylin and eosin staining of histologic whole slide images, Lab. Investig. 103 (5) (2023) 100070.
[52] J. Salido, N. Vallez, L. González-López, O. Deniz, G. Bueno, Comparison of deep learning models for digital H&E staining from unpaired label-free multispectral microscopy images, Comput. Methods Programs Biomed. 235 (2023) 107528.
[53] T. Park, A.A. Efros, R. Zhang, J.-Y. Zhu, Contrastive learning for unpaired image-to-image translation, in: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16, Springer, 2020, pp. 319–345.
[54] T. Fang, Z. Wu, X. Chen, L. Tan, Z. Li, J. Ji, et al., Label-Free virtual peritoneal lavage cytology via deep-learning-assisted single-color stimulated raman scattering microscopy, Adv. Intell. Syst. (2024) 2300689.
[55] R. Koka, L.M. Wake, N.K. Ku, K. Rice, A. LaRocque, E.G. Vidal, et al., Assessment of AI-based computational H&E staining versus chemical H&E staining for primary diagnosis in lymphomas: a brief interim report, J. Clin. Pathol. 78 (3) (2024) 208–211.
[56] P.F. Wong, C. McNeil, Y. Wang, J. Paparian, C. Santori, M. Gutierrez, et al., Clinical-grade validation of an autofluorescence virtual staining system with human experts and a deep learning system for prostate cancer, Mod. Pathol. 37 (11) (2024) 100573.
[57] R. Zhu, H. He, Y. Chen, M. Yi, S. Ran, C. Wang, et al., Deep learning for rapid virtual H&E staining of label-free glioma tissue from hyperspectral images, Comput. Biol. Med. 180 (2024) 108958.
[58] C. Yoon, E. Park, S. Misra, J.Y. Kim, J.W. Baik, K.G. Kim, et al., Deep learning-based virtual staining, segmentation, and classification in label-free photoacoustic histology of human specimens, Light: Sci. Appl. 13 (1) (2024) 226.
[59] M.Z. Asaf, B. Rao, M.U. Akram, S.G. Khawaja, S. Khan, T.M. Truong, et al., Dual contrastive learning based image-to-image translation of unstained skin tissue into virtually stained H&E images, Sci. Rep. 14 (1) (2024) 2335.
[60] J. Fan, N. Zeng, H. He, C. He, S. Liu, H. Ma, Generating bright-field images of stained tissue slices from mueller matrix polarimetric images with CycleGAN using unpaired dataset, J. Innov. Opt. Health Sci. 18 (2) (2024) 2343003.
[61] Z. Chen, I.H.M. Wong, W. Dai, C.T.K. Lo, T.T.W. Wong, Lung cancer diagnosis on virtual histologically stained tissue using weakly supervised learning, Mod. Pathol. 37 (6) (2024) 100487.
[62] M. Boktor, J.E. Tweel, B.R. Ecclestone, J.A. Ye, P. Fieguth, P. Haji Reza, Multi-channel feature extraction for virtual histological staining of photon absorption remote sensing images, Sci. Rep. 14 (1) (2024) 2009.
[63] X. Li, H. Liu, X. Song, C.C. Marboe, B.C. Brott, S.H. Litovsky, et al., Structurally constrained and pathology-aware convolutional transformer generative adversarial network for virtual histology staining of human coronary optical coherence tomography images, J. Biomed. Opt. 29 (3) (2024) 036004.
[64] A.B. Azam, F. Wee, J.P. Väyrynen, W.W.-Y. Yim, Y.Z. Xue, B.L. Chua, et al., Training immunophenotyping deep learning models with the same-section ground truth cell label derivation method improves virtual staining accuracy, Front. Immunol. 15 (2024) 1404640.
[65] X. Yang, B. Bai, Y. Zhang, M. Aydin, Y. Li, S.Y. Selcuk, et al., Virtual birefringence imaging and histological staining of amyloid deposits in label-free tissue using autofluorescence microscopy and deep learning, Nat. Commun. 15 (1) (2024) 7978.
[66] Z. Xu, L. Guo, S. Tsutsui, S. Zhang, A.C. Kot, B. Wen, Digital staining with knowledge distillation: a unified framework for unpaired and paired-but-misaligned data, IEEE Trans. Med. Imaging 44 (9) (2025) 3770–3782.
[67] Y. Li, N. Pillar, T. Liu, G. Ma, Y. Qi, Y. Zhang, et al., Label-free evaluation of lung and heart transplant biopsies using tissue autofluorescence-based virtual staining, BME Front. 6 (2025) 0151.
[68] Y. Zhang, L. Huang, N. Pillar, Y. Li, H. Chen, A. Ozcan, Pixel super-resolved virtual staining of label-free tissue using diffusion models, Nat. Commun. 16 (1) (2025) 5016.
[69] Z. Liu, E. Hirata-Miyasaki, S. Pradeep, J.V. Rahm, C. Foley, T. Chandler, et al., Robust virtual staining of landmark organelles with cytoland, Nat. Mach. Intell. 7 (6) (2025) 901–915.
[70] J. Park, S.-J. Shin, G. Kim, H. Cho, D. Ryu, D. Ahn, et al., Revealing 3D microanatomical structures of unlabeled thick cancer tissues using holotomography and virtual H&E staining, Nat. Commun. 16 (1) (2025) 4781.
[71] C. Mercan, G.C.A.M. Mooij, D. Tellez, J. Lotz, N. Weiss, M. van Gerven, et al., Virtual staining for mitosis detection in breast histopathology, in: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), IEEE, 2020, pp. 1770–1774.
[72] E.A. Burlingame, M. McDonnell, G.F. Schau, G. Thibault, C. Lanciault, T. Morgan, et al., SHIFT: speedy histological-to-immunofluorescent translation of a tumor signature enabled by deep learning, Sci. Rep. 10 (1) (2020) 17507.
[73] R. Wang, P. Song, S. Jiang, C. Yan, J. Zhu, C. Guo, et al., Virtual brightfield and fluorescence staining for fourier ptychography via unsupervised deep learning, Opt. Lett. 45 (19) (2020) 5405–5408.
[74] J.J. Levy, N. Azizgolshani, M.J. Andersen Jr, A. Suriawinata, X. Liu, M. Lisovsky, et al., A large-scale internal validation study of unsupervised virtual trichrome staining technologies on nonalcoholic steatohepatitis liver biopsies, Mod. Pathol. 34 (4) (2021) 808–822.
[75] A. Picon, A. Medela, L.F. Sanchez-Peralta, R. Cicchi, R. Bilbao, D. Alfieri, et al., Autofluorescence image reconstruction and virtual staining for in-vivo optical biopsying, IEEE Access 9 (2021) 32081–32093.
[76] Y.-C. Lo, I.-F. Chung, S.-N. Guo, M.-C. Wen, C.-F. Juang, Cycle-consistent GAN-based stain translation of renal pathology images with glomerulus detection application, Appl. Soft Comput. 98 (2021) 106822.
[77] N. Kaza, A. Ojaghi, P.C. Costa, F.E. Robles, Deep learning-based virtual staining of label-free ultraviolet (UV) microscopy images for hematological analysis, in: Label-Free Biomedical Imaging and Sensing (LBIS) 2021, vol. 11655, SPIE, 2021, pp. 16–24.
[78] A.-K. Thebille, E. Dietrich, M. Klaus, L. Gernhold, M. Lennartz, C. Kuppe, et al., Deep learning-based bias transfer for overcoming laboratory differences of microscopic images, in: Medical Image Understanding and Analysis: 25th Annual Conference, MIUA 2021, Oxford, United Kingdom, July 12–14, 2021, Proceedings 25, Springer, 2021, pp. 322–336.
[79] K. de Haan, Y. Zhang, J.E. Zuckerman, T. Liu, A.E. Sisk, M.F.P. Diaz, et al., Deep learning-based transformation of H&E stained tissues into special stains. NAT commun 12: 4884, 2021.
[80] Y. Hong, Y.J. Heo, B. Kim, D. Lee, S. Ahn, S.Y. Ha, et al., Deep learning-based virtual cytokeratin staining of gastric carcinomas to measure tumor–stroma ratio, Sci. Rep. 11 (1) (2021) 19255.
[81] H. Wieslander, A. Gupta, E. Bergman, E. Hallström, P.J. Harrison, Learning to see colours: biologically relevant virtual staining for adipocyte cell images, PLoS One 16 (10) (2021) e0258546.
[82] A. Lahiani, I. Klaman, N. Navab, S. Albarqouni, E. Klaiman, Seamless virtual whole slide image synthesis and validation using perceptual embedding consistency, IEEE J. Biomed. Health Inform. 25 (2) (2020) 403–411.
[83] S. Liu, B. Zhang, Y. Liu, A. Han, H. Shi, T. Guan, et al., Unpaired stain transfer using pathology-consistent constrained generative adversarial networks, IEEE Trans. Med. Imaging 40 (8) (2021) 1977–1989.
[84] O. Cetin, M. Chen, P. Ziegler, P. Wild, H. Koeppl, Deep learning-based restaining of histopathological images, in: 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, 2022, pp. 1467–1474.
[85] M. Lysik, Z. Swiderska-Chadaj, T. Markiewicz, T. Les, S. Cierniak, M. Lorent, HE-to-PAS histological stain conversion by GAN in renal pathology, in: 2022 International Joint Conference on Neural Networks (IJCNN), IEEE, 2022, pp. 1–7.
[86] G. Zhang, B. Ning, H. Hui, T. Yu, X. Yang, H. Zhang, et al., Image-to-images translation for multiple virtual histological staining of unlabeled human carotid atherosclerotic tissue, Mol. Imaging Biol. (2022) 1–11.
[87] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, J. Choo, StarGAN: unified generative adversarial networks for multi-domain image-to-image translation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 8789–8797.
[88] M. Liu, Y. Ding, M. Xia, X. Liu, E. Ding, W. Zuo, et al., StGAN: a unified selective transfer network for arbitrary image attribute editing, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 3673–3682.
[89] R. Zhang, Y. Cao, Y. Li, Z. Liu, J. Wang, J. He, et al., MVFStain: multiple virtual functional stain histopathology images generation based on specific domain mapping, Med. Image Anal. 80 (2022) 102520.
[90] B. Zeng, Y. Lin, Y. Wang, Y. Chen, J. Dong, X. Li, et al., Semi-supervised PR virtual staining for breast histopathological images, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2022, pp. 232–241.
[91] J. Kim, M. Kim, H. Kang, K. Lee, U-GAT-IT: unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation, arXiv preprint arXiv:1907.10830, 2020.
[92] I. Zingman, S. Frayle, I. Tankoyeu, S. Sukhanov, F. Heinemann, A comparative evaluation of image-to-image translation methods for stain transfer in histopathology, in: Medical Imaging with Deep Learning, PMLR, 2024, pp. 1509–1525.

[93] G. Duan, Y. Cao, W. Guo, L. Cui, Z. Liu, A virtual staining method for immunohistochemical images of breast cancer, in: 2023 16th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), IEEE, 2023, pp. 1–5.
[94] T. Sloboda, L. Hudec, W. Benešová, Editable stain transformation of histological images using unpaired GANs, in: International Conference on Image Analysis and Processing, Springer, 2023, pp. 27–38.
[95] Y. Shen, B. Zhou, Closed-form factorization of latent semantics in GANs, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 1532–1540.
[96] T. Biswas, H. Suzuki, M. Ishikawa, N. Kobayashi, T. Obi, Generation of RGB EVG stained image from hyperspectral H&E stained image using generative adversarial network (GAN), in: Medical Imaging 2023: Digital and Computational Pathology, vol. 12471, SPIE, 2023, pp. 80–89.
[97] T. Biswas, H. Suzuki, M. Ishikawa, N. Kobayashi, T. Obi, Generative adversarial network based digital stain conversion for generating RGB EVG stained image from hyperspectral H&E stained image, J. Biomed. Opt. 28 (5) (2023) 056501.
[98] Y. Ling, W. Tan, B. Yan, Self-supervised digital histopathology image disentanglement for arbitrary domain stain transfer, IEEE Trans. Med. Imaging 42 (12) (2023) 3625–3638.
[99] R. Yan, Q. He, Y. Liu, P. Ye, L. Zhu, S. Shi, et al., Unpaired virtual histological staining using prior-guided generative adversarial networks, Comput. Med. Imaging Graph. 105 (2023) 102185.
[100] X. Liu, B. Li, C. Liu, D. Ta, Virtual fluorescence translation for biological tissue by conditional generative adversarial network, Phenomics 3 (4) (2023) 408–420.
[101] M. Wiedenmann, M. Barch, P.S. Chang, J. Giltnane, T. Risom, A. Zijlstra, An immunofluorescence-guided segmentation model in hematoxylin and eosin images is enabled by tissue artifact correction using a Cycle-Consistent generative adversarial network, Mod. Pathol. 37 (11) (2024) 100591.
[102] S. Zeng, X. Li, Y. Liu, Q. Huang, Y. He, Automatic annotation diagnostic framework for nasopharyngeal carcinoma via pathology–fidelity GAN and prior-driven classification, Bioengineering 11 (7) (2024) 739.
[103] R. Wang, S. Yang, Q. Li, D. Zhong, CytoGAN: unpaired staining transfer by structure preservation for cytopathology image analysis, Comput. Biol. Med. 180 (2024) 108942.
[104] X. Hou, Z. Guan, X. Zhang, X. Hu, S. Zou, C. Liang, et al., Evaluation of tumor budding with virtual panck stains generated by novel multi-model CNN framework, Comput. Methods Programs Biomed. 257 (2024) 108352.
[105] Y. Li, X. Guan, Y. Wang, Y. Zhang, Exploiting supervision information in weakly paired images for IHC virtual staining, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2024, pp. 113–122.
[106] W. Zhang, T.H. Hui, P.Y. Tse, F. Hill, C. Lau, X. Li, High-resolution medical image translation via patch alignment-based bidirectional contrastive learning, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2024, pp. 178–188.
[107] J. Fan, X. Zhang, N. Zeng, S. Liu, H. He, L. Luo, et al., Stain transformation using mueller matrix guided generative adversarial networks, Opt. Lett. 49 (18) (2024) 5135–5138.
[108] X. Guan, Y. Wang, Y. Lin, X. Li, Y. Zhang, Unsupervised multi-domain progressive stain transfer guided by style encoding dictionary, IEEE Trans. Image Process. 33 (2024) 767–779.
[109] J. Li, J. Dong, S. Huang, X. Li, J. Jiang, X. Fan, et al., Virtual immunohistochemistry staining for histological images assisted by weakly-supervised learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 11259–11268.
[110] J. Denholm, A. Hamidinekoo, N. Burlutskiy, L.C. Setyo, I. Zhang, F. Yousefi, et al., Virtual histological staining as a tool for extending renal segmentation across stains, Mod. Pathol. (2025) 100842.
[111] X. Meng, X. Li, X. Wang, A computationally virtual histological staining method to ovarian cancer tissue by deep generative adversarial networks, Comput. Math. Methods Med. 2021 (1) (2021) 4244157.
[112] X. Yang, B. Bai, Y. Zhang, Y. Li, K. de Haan, T. Liu, et al., Virtual stain transfer in histology via cascaded deep neural networks, ACS Photonics 9 (9) (2022) 3134–3143.
[113] J. Vasiljević, F. Feuerhake, C. Wemmert, T. Lampert, HistoStarGAN: a unified approach to stain normalisation, stain transfer and stain invariant segmentation in renal histopathology, Knowl.-Based Syst. 277 (2023) 110780.
[114] C. Franchet, R. Schwob, G. Bataillon, C. Syrykh, S. Péricart, F.-X. Frenois, et al., Bias reduction using combined stain normalization and augmentation for AI-based classification of histological images, Comput. Biol. Med. 171 (2024) 108130.
[115] G. Litjens, P. Bandi, B. Ehteshami Bejnordi, O. Geessink, M. Balkenhol, P. Bult, et al., 1399 H&E-stained sentinel lymph node sections of breast cancer patients: the camelyon dataset, GigaScience 7 (6) (2018) giy065.
[116] N. Kumar, R. Verma, S. Sharma, S. Bhargava, A. Vahadane, A. Sethi, MoNuSeg: multi-organ nuclei segmentation challenge, 2018, https://monuseg.grand-challenge.org/, (Accessed: 22 May 2025).
[117] F. Li, Z. Hu, W. Chen, A. Kak, MIST dataset: multi-stain breast carcinoma histopathology, 2023, https://github.com/openmedlab/Awesome-Medical-Dataset/blob/main/resources/MIST-HER2.md, (Accessed: 1 December 2025).
[118] J.N. Kather, N. Halama, A. Marx, NCT-CRC-HE-100K: colorectal cancer histology dataset, 2018, https://zenodo.org/records/1214456, (Accessed: 1 December 2025).
[119] S. Liu, C. Zhu, F. Xu, X. Jia, Z. Shi, M. Jin, BCI: breast cancer immunohistochemical image generation through pyramid pix2pix, 2023, https://github.com/openmedlab/Awesome-Medical-Dataset/blob/main/resources/BCI.md, (Accessed: 1 December 2025).
[120] X. Xu, S. Kapse, P. Prasanna, Histo-diffusion: a diffusion super-resolution method for digital pathology with comprehensive quality assessment, arXiv preprint arXiv:2408.15218, 2024.
[121] T. Kataria, B. Knudsen, S. Elhabian y, StainDiffuser: multitask dual diffusion model for virtual staining, arXiv preprint arXiv:2403.11340, 2024.
[122] Y. He, Z. Liu, M. Qi, S. Ding, P. Zhang, F. Song, et al., PST-Diff: achieving high-consistency stain transfer by diffusion models with pathological and structural constraints, IEEE Trans. Med. Imaging 43 (10) (2024) 3634–3647.
[123] Z. Yin, B. He, Y. Ying, S. Zhang, P. Yang, Z. Chen, et al., Fast and label-free 3D virtual H&E histology via active phase modulation-assisted dynamic full-field oct, npj Imaging 3 (1) (2025) 12.
[124] Y. Rivenson, H. Wang, Z. Wei, K. de Haan, Y. Zhang, Y. Wu, et al., Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning, Nat. Biomed. Eng. 3 (6) (2019) 466–477.
[125] L. Huang, Y. Li, N. Pillar, T.K. Haran, W.D. Wallace, A. Ozcan, Autonomous quality and hallucination assessment for virtual tissue staining and digital pathology, arXiv preprint arXiv:2404.18458, 2024.
[126] S. Dubey, Y. Chong, B. Knudsen, S. Elhabian y, VIMS: virtual immunohistochemistry multiplex staining via text-to-stain diffusion trained on uniplex stains, in: International Workshop on Machine Learning in Medical Imaging, Springer, 2024, pp. 143–155.
[127] F. Chen, R. Zhang, B. Zheng, Y. Sun, J. He, W. Qin, Pathological semantics-preserving learning for H&E-to-IHC virtual staining, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2024, pp. 384–394.
[128] N. Marini, S. Marchesin, M. Wodzinski, A. Caputo, D. Podareanu, B.C. Guevara, et al., Multimodal representations of biomedical knowledge from limited training whole slide images and reports using deep learning, Med. Image Anal. 97 (2024) 103303.
[129] L.A. Schoenpflug, Y. Nie, F. Sheikhzadeh, V.H. Koelzer, A review on federated learning in computational pathology, Comput. Struct. Biotechnol. J. 23 (2024) 3938–3945.
[130] N. Wagner, M. Fuchs, Y. Tolkach, A. Mukhopadhyay, Federated stain normalization for computational pathology, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2022, pp. 14–23.
[131] H. Guan, P.-T. Yap, A. Bozoki, M. Liu, Federated learning for medical image analysis: a survey, Pattern Recognit. (2024) 110424.
[132] J. Kweon, M. Kim, G. Yun, S. Kwon, J. Yoo, MS-SST: single image reconstruction-based stain-style transfer for multi-domain Hematoxylin & Eosin stained pathology images, IEEE Access 11 (2023) 50090–50097.

Author biography

Rabiah Al-Qudah is an Assistant Professor of Computer Science at Abu Dhabi University and a researcher in Machine Learning and Artificial Intelligence, with a particular emphasis on Computer Vision and smart cities. She holds a PhD and MSc degrees in Computer Science from Concordia University, Canada. She has many years of experience in software and database development. She published several articles in international conferences and reputable journals.

Abubakar Bala received the B.Eng. (Hons.) degree in Computer Engineering from Bayero University Kano, Nigeria, in 2011, the M.Sc. degree in Computer Engineering from King Fahd University of Petroleum and Minerals (KFUPM), Saudi Arabia, in 2015, and the Ph.D. degree in Electrical Engineering from Universiti Teknologi PETRONAS (UTP), Malaysia, in 2022. From 2023 to 2024, he was a Postdoctoral Fellow at the Interdisciplinary Research Centre for Communication Systems and Sensing (IRCCSS), KFUPM. He is currently a Senior Lecturer at the School of Business, Monash University Malaysia. His research interests include optimization, artificial neural networks, computer vision, and metaheuristics, with a recent focus on deep learning for digital pathology.



Mrouj Almuhajri is an Assistant Professor at the Saudi Electronic University, specializing. She holds a PhD and MSc degrees in Computer Science from Concordia University, Canada. Her research focuses on deep learning and computer vision development. She has published multiple peer-reviewed papers in reputable international journals and conferences, contributing to the advancement of application of AI solutions that ensure sustainability.

Khiati Zakaria is an Assistant Professor of Computer Science at Abu Dhabi University and a researcher in Machine Learning and Artificial Intelligence, with a particular emphasis on Natural Language Processing and large language models. He holds a Ph.D. from KAIST, an M.S. from Korea University, and a B.S. in Mathematics and Computer Science from USTHB. His work spans core ML and AI topics, learning algorithms, data collection and curation, model development, benchmarking, and deployment applied especially to NLP problems and LLM development for low-resource languages. He has led academic and industry projects, supervised and mentored students and junior researchers, and collaborated across South Korea, Algeria, and the UAE.

Ching Y. Suen is the Founder and Co-Director of CENPARMI and Concordia Honorary Chair in AI & Pattern Recognition. He received his Ph.D. from the University of British Columbia and his Master's degree from the University of Hong Kong. He has served as Chair of the Department of Computer Science and Associate Dean (Research) of the Faculty of Engineering and Computer Science at Concordia University. Dr. Suen has published 16 books and over 550 papers, many of which have been widely cited and applied in areas such as handwriting recognition, thinning, font analysis, and multiple classifier systems. He has supervised 120 graduate students and hosted around 100 long-term visiting scientists and professors. He is a Fellow of IEEE, IAPR, and the Royal Society of Canada, and the recipient of numerous awards, including the IAPR King-Sun Fu Prize (2020), Elsevier Pattern Recognition Journal Award of Excellence (2016), ICDAR Award (2005), ITAC/NSERC national award (1993), and Concordia's Lifetime Research Achievement Award (2008). He is Emeritus Editor-in-Chief of *Pattern Recognition*, serves on several editorial boards, and is the founder or co-founder of major conferences including ICDAR, IWFHR/ICFHR, ICPRAI, and VI.