Cureus Journal of Computer Science
https://doi.org/10.7759/s44389-025-00020-8

Original Article

Comparative Analysis of Generative and Diffusion Models for Mitosis Synthesis in Histopathology Images

Rahaf B. Alkhadra [1], Pablo D. Santana [2], Taranpreet Rai [3], Kevin Wells [3]

1. *Computer Science/Artificial Intelligence, University of Surrey, Guildford, Surrey, GBR*
2. *Veterinary Pathology Centre, University of Surrey, Guildford, Surrey, GBR*
3. *Centre for Vision Speech and Signal Processing, University of Surrey, Guildford, Surrey, GBR*

Received: May 29, 2025 | Review began: December 15, 2025 | Review ended: January 24, 2026 | Published: January 27, 2026

Abstract

Accurate cancer diagnosis depends on identifying key cellular features, including mitotic figures, which have a major influence on treatment decisions and patient outcomes. Histopathological evaluation remains the clinical gold standard for detecting mitotic figures, but any manual evaluation is laborious and often hindered by large and unpredictable inter-observer variability. The use of Generative Artificial Intelligence (GenAI) continues to gain traction as a viable means of producing synthetic medical images for the purpose of creating data augmentations to feed model training and reduce dependence on manually labelling images. However, the application of GenAI within histopathology has not been thoroughly assessed. In this study, we assess two leading GenAI architectures, Denoising Diffusion Probabilistic Models (DDPM) and StyleGAN3, on their ability to synthesise mitotic figures using the MIDOG++ dataset, which consists of multiple tumour types from various species. Our results show that DDPM captures fine structural morphology very well, and StyleGAN3's color reproduction was superior. Quantitative evaluation using Fréchet Inception Distance (FID), Structural Similarity Index (SSIM), and expert-based Receiver Operating Characteristic analysis showed that DDPM achieved lower FID and higher structural similarity, while both DDPM and StyleGAN3 produced synthetic images that were largely indistinguishable from real samples by expert pathologists (area under the curve ≈ 0.5). Although diffusion models exhibit improved structural fidelity relative to generative adversarial networks-based approaches, they introduce higher computational demands and unresolved questions regarding generalisability across institutions and imaging pipelines. This functional distinction illustrates complementary capabilities that can be judiciously utilised depending upon the augmentation needs of the dataset. By incorporating these high-quality synthetic images into training pipelines, researchers can reduce manual annotation burden, reduce observer variance, and improve the robustness of mitosis detection models. In contrast to prior histopathology studies that have largely focused on GAN-based synthesis, this work presents a systematic comparative evaluation of diffusion models (DDPM) and StyleGAN3 for mitotic figure generation using the MIDOG++ dataset, clarifying their respective strengths and limitations for synthetic data augmentation.

Categories: Advances in Biomedical data analysis, AI applications, Generative AI for healthcare

Keywords: generative ai, digital pathology, diffusion models, gans, mitosis detection, data augmentation, cancer, dogs, cats, dataset

How to cite this article:
Alkhadra R B, Santana P D, Rai T, et al. (January 27, 2026) Comparative Analysis of Generative and Diffusion Models for Mitosis Synthesis in Histopathology Images. Cureus J Comput Sci 3 : es44389-025-00020-8. DOI https://doi.org/10.7759/s44389-025-00020-8

directly correlating with tumour growth and aggressiveness [6,7]. However, current manual mitosis detection practices are inherently subjective, labour-intensive, and susceptible to significant inter-observer variability, leading to inconsistent diagnostic outcomes and potential delays in critical therapeutic decisions [8].

The quantification of mitotic activity in histopathological specimens constitutes a fundamental parameter in tumour grading systems across numerous malignancies, including breast carcinoma, melanoma, and soft tissue sarcomas [9,10]. Conventional mitosis detection protocols require pathologists to manually identify and enumerate mitotic figures within high-power fields (HPFs), typically examining 10-50 HPFs depending on the specific tumour type [11,12]. This process presents several critical limitations including significant time expenditure, particularly in cases with extensive tissue samples, substantial inter-observer variability, cognitive fatigue during prolonged microscopic examination, and variability in tissue processing and staining protocols resulting in clinical consequences [13-15].

Recent advancements in Artificial Intelligence (AI) have provided new avenues to automate and enhance the accuracy of mitosis detection, thereby addressing inherent limitations in manual methods [16-18]. AI-driven approaches, particularly supervised machine learning models including convolutional neural networks (CNNs) and transformer-based architectures, have demonstrated potential but remain heavily reliant on the availability of large, diverse, and high-quality labelled datasets [19]. Multiple studies have reported detection accuracies exceeding 85% in controlled settings, representing significant improvement over manual detection in terms of both speed and consistency [20-22]. Nevertheless, these supervised methodologies encounter substantial implementation barriers, primarily centred around dataset requirements. The synthesis of comprehensive histopathological training datasets presents multifaceted challenges: (1) acquisition of diverse tissue samples representing various tumour types, grades, and patient demographics; (2) labour-intensive labelling by multiple expert pathologists to establish ground truth; (3) significant ethical and regulatory considerations regarding patient consent and data privacy; and (4) technical variability across different laboratory settings, imaging platforms, and staining protocols [23-27]. New strategies capable of synthesising realistic, diverse, and privacy-preserving histopathological images are urgently required to overcome these limitations and accelerate the clinical implementation of AI-assisted diagnostic systems.

Generative AI models, specifically Generative Adversarial Networks (GANs) and diffusion-based models, offer promising solutions by enabling the synthesis of highly realistic histopathological images while preserving patient privacy and facilitating dataset expansion. Among these, StyleGAN3 and Denoising Diffusion Probabilistic Models (DDPM) represent cutting-edge generative techniques with complementary strengths [28,29]. StyleGAN3 excels in producing high-quality images rapidly, characterised by realistic colour distribution, texture fidelity, and staining pattern representation, making it suitable for histopathological applications that rely heavily on accurate colour interpretations and pattern recognition [30]. For example, advances in these frameworks have been used for the discovery of new cancer stages [31]. Conversely, DDPM is renowned for its structural realism, fine-grained detail preservation, and training stability, making it highly suitable for synthesising images requiring precise anatomical and structural accuracy [32,33]. The iterative denoising process inherent to DDPM facilitates exceptional detail resolution, particularly advantageous for representing the nuclear morphology and chromatin patterns characteristic of mitotic figures [34].

Herein, we introduce a comprehensive evaluation of DDPM and StyleGAN3 models for generating synthetic histopathological images containing mitotic figures. The study uses MIDOG++, which is a large multi-domain mitotic figure dataset comprising 503 histological specimens across seven tumour types from human and canine tissue, with 11,937 expert-annotated mitotic figures, and is explicitly designed to capture severe domain shifts arising from heterogeneous tumour morphology, slide preparation protocols, and whole-slide imaging scanners across multiple laboratories [35]. The central challenge addressed in this work is the robustness of generative models to clinically relevant domain shifts in mitotic figure appearance, which are underpinned by variability in tumour morphology, staining protocols, slide preparation, and whole-slide imaging scanners across institutions and species. Such variability is known to degrade the performance and generalisation of deep learning models, motivating a direct comparison of GAN-based and diffusion-based architectures to assess their relative capacity to capture structural and appearance-level diversity under these conditions. This multifaceted dataset enables rigorous assessment of generative model performance across heterogeneous imaging conditions, closely approximating real-world clinical variability. Through systematic experimental evaluation, we quantitatively assess the quality and clinical utility of synthetic images using established metrics including Fréchet Inception Distance (FID), Structural Similarity Index (SSIM), and Mean Squared Error (MSE).

How to cite this article:
Alkhadra R B, Santana P D, Rai T, et al. (January 27, 2026) Comparative Analysis of Generative and Diffusion Models for Mitosis Synthesis in Histopathology Images. Cureus J Comput Sci 3 : es44389-025-00020-8. DOI https://doi.org/10.7759/s44389-025-00020-8

Furthermore, qualitative assessment by expert pathologists utilising Receiver Operating Characteristic (ROC) analysis provides critical insights regarding the clinical indistinguishability of synthetic images. Our research bridges a critical knowledge gap, providing actionable insights for the optimal application and integration of generative AI models into clinical diagnostics and training datasets. These findings will establish a new understanding for developing hybrid generative strategies combining the strengths of multiple generative models, potentially transforming computational pathology and cancer diagnostics.

Materials And Methods

Experimental setup

Workflow

The study followed a five-stage experimental framework: (i) Data Preparation, (ii) Model Selection, (iii) Model Training and Image Generation, (iv) Image Evaluation, and (v) Results Analysis. Figure *1* illustrates the full process.

FIGURE 1: Experimental framework followed throughout this manuscript.

Data Preparation

We used the MIDOG++ multi-domain mitotic-figure dataset comprising 503 whole-slide images spanning seven tumour types across human and canine specimens, with ~12,000 expert-confirmed mitotic annotations. Slides were scanned at 40× equivalent resolution.

Hardware and Software

All models were trained on a single NVIDIA A100 GPU (40 GB HBM2e, 450 W TDP) housed in a SLURM-managed cluster. Code was written in PyTorch 2.2 with CUDA 12.4; diffusion training used the Hugging Face-diffusers library, and GAN training used the official StyleGAN3-T implementation.

Model selection

We implemented and evaluated two distinct approaches to synthetic mitotic figure generation in histopathology images: StyleGAN3 and DDPM. These state-of-the-art generative models were selected based on their demonstrated capabilities in producing high-quality, realistic images and their complementary architectural approaches to image synthesis [36,37].

StyleGAN3

We utilised the translation-equivariant generator variant of StyleGAN3 (StyleGAN-T) for its optimal balance between generation quality and computational efficiency. This variant was selected due to its favourable performance characteristics: generation time under 1 second per image and second-highest Zero-shot FID score among comparative models. The translation equivariance property enhances the model's ability to generate consistent mitotic figures regardless of their position within the image space, crucial for histopathological applications.

The model architecture follows the original StyleGAN3 configuration with specific adaptations for histopathological image generation (Figure *2*). A batch size of 8 was implemented to balance training stability and GPU memory constraints. The adaptive discriminator augmentation (ADA) mechanism was configured with a strength parameter (gamma) of 8.2, based on recommendations from the original StyleGAN3 paper [38]. This parameter plays a crucial role in preventing overfitting and maintaining stability during the adversarial training process, particularly important when working with limited datasets of specialised histopathological features.

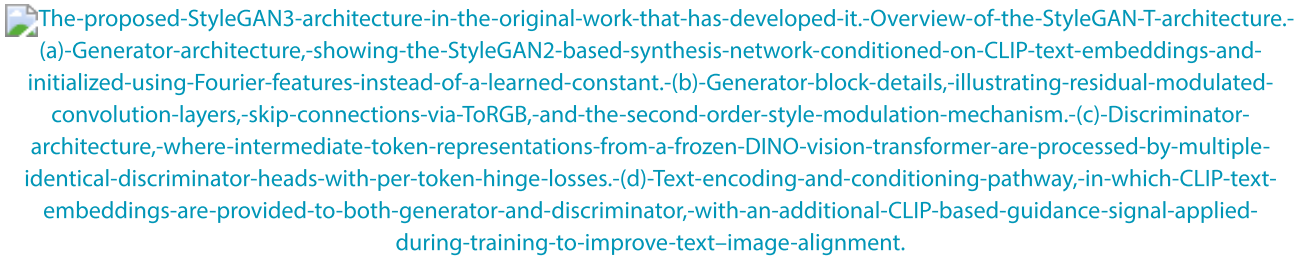

FIGURE 2: The proposed StyleGAN3 architecture in the original work that has developed it. Overview of the StyleGAN-T architecture. (a) Generator architecture, showing the StyleGAN2-based synthesis network conditioned on CLIP text embeddings and initialized using Fourier features instead of a learned constant. (b) Generator block details, illustrating residual modulated convolution layers, skip connections via ToRGB, and the second-order style modulation mechanism. (c) Discriminator architecture, where intermediate token representations from a frozen DINO vision transformer are processed by multiple identical discriminator heads with per-token hinge losses. (d) Text encoding and conditioning pathway, in which CLIP text embeddings are provided to both generator and discriminator, with an additional CLIP-based guidance signal applied during training to improve text–image alignment.

Denoising Diffusion Probabilistic Models

The DDPM implementation utilised a U-Net architecture, specifically configured for the denoising process critical to diffusion models (Figure *3*) The model was designed to process images at 256×256 pixel resolution with a batch size of 8. The U-Net structure featured a symmetrical configuration of down-sampling and up-sampling blocks, with channel counts increasing from 128 to 512, enabling the capture of both fine-grained details and broader contextual information essential for realistic mitotic figure generation.

For optimisation, AdamW optimiser was used with weight decay for improved generalisation, coupled with a cosine learning rate scheduler starting at 1e-4 and incorporating a 500-step warm-up period [39]. This adaptive learning rate approach stabilised the early training stages and optimised convergence. The noise scheduling process utilised 1000 timesteps to ensure a fine-grained denoising process, enabling the model to learn the underlying structure of mitotic figures at various scales of detail. The loss function was formulated as an MSE between predicted noise and applied noise, enabling the model to learn the reverse process of gradually denoising images. Attention mechanisms were integrated in both down-sampling and up-sampling layers, implementing spatial self-attention to improve the model's capability in capturing long-range dependencies, crucial for maintaining global coherence of the generated mitotic figures within their histological context.

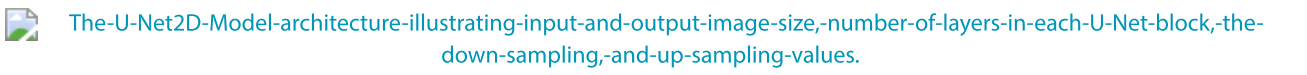

FIGURE 3: The U-Net2D Model architecture illustrating input and output image size, number of layers in each U-Net block, the down-sampling, and up-sampling values.

Dataset

Dataset Characteristics

Our study utilised the MIDOG++ dataset, a comprehensive collection of histopathology images originally developed for mitotic figure detection, as shown in Figure *4* [35]. The dataset contains 503 images representing seven different tumour types from both human and canine cancer, with approximately 12,000 annotated mitotic figures. The diversity of this dataset, encompassing multiple tumour types and species, provides an ideal foundation for training and evaluating generative models in the context of synthesising mitotic figures in varied histopathological contexts [35].

FIGURE 4: Samples of extracted patches from real images in the MIDOG++ dataset.

Data Preprocessing

To prepare the dataset for model training, a patch-based extraction approach was implemented to isolate regions of interest (ROI) containing mitotic figures from the original histopathological images. The selected patch size was set to 256 × 256 pixels, carefully determined to achieve an optimal trade-off between capturing the mitotic figures and excluding irrelevant features. To ensure comprehensive coverage and prevent loss of important features near patch boundaries, an overlapping extraction strategy with a stride of 128 pixels (50% overlap) was employed. This method increased the likelihood of capturing mitotic figures near the edges of patches while providing diverse perspectives of the same mitotic figure [40]. As a result, the extraction process generated 88,777 patches, each containing at least one mitotic figure.

Following extraction, data augmentation techniques were applied to enhance diversity within the training dataset. Methods such as Random Horizontal Flip were incorporated to introduce variability, improving the model's invariance to mitotic figure orientation. Additionally, normalisation was included in the preprocessing pipeline to standardise pixel values across the dataset. This comprehensive preprocessing framework ensured a diverse and robust dataset suitable for model training and evaluation. Samples of the extracted patches and the applied preprocessing techniques are presented in Figure *5* [35].

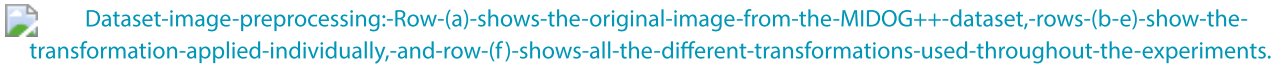

FIGURE 5: Dataset image preprocessing: Row (a) shows the original image from the MIDOG++ dataset, rows (b-e) show the transformation applied individually, and row (f) shows all the different transformations used throughout the experiments.

Computational resources

Both models were trained on NVIDIA A100 GPUs with 40GB VRAM, providing the necessary computational capacity for these architectures. The high memory bandwidth and tensor core capabilities of the A100 were particularly advantageous for training sophisticated generative models on high-resolution histopathology images.

Evaluation metrics

To comprehensively assess model performance in synthesising mitotic figures, we employed multiple evaluation metrics designed to measure both structural similarity and pixel-level differences between generated and ground truth images. It is important to note that while metrics such as FID, SSIM, and MSE provide quantitative insight into distributional similarity and perceptual fidelity, they do not directly capture clinically meaningful attributes specific to mitotic figures. In particular, high image realism does not necessarily imply improved performance in downstream detection or grading tasks. These metrics are therefore used in this study as proxies for structural and visual quality rather than as indicators of diagnostic effectiveness.

FID Score

The FID score was utilised to assess the statistical similarity between the feature distributions of real and synthesised images. This metric calculates the distance between feature representations extracted from a pre-trained Inception-v3 model. Lower FID scores indicate that the synthesised image feature distribution more closely resembles that of real images. We calculated the FID score between equal sets of real histopathology patches containing mitotic figures and generated patches from each model.

Structural Similarity Index

The SSIM metric evaluates the structural similarity between generated and real images by considering luminance, contrast, and structure components. SSIM values range from -1 to 1, with values closer to 1 indicating higher similarity. We computed the average SSIM between generated patches and their closest real counterparts using two approaches:

· Between grayscale versions of generated and ground truth images

· Between the original-coloured images and ground truth

This dual approach allowed for assessment of structural similarities independent of potential colour variations.

Mean Squared Error

MSE provides a pixel-level comparison between generated and real images, offering insights into the overall fidelity of the generated content. While not always indicative of perceptual quality, it complements the other metrics by quantifying absolute differences in pixel values.

ROC Study

To incorporate human perception assessment, we conducted an ROC study evaluating the indistinguishability of generated images from real ones. Expert pathologists were presented with a mix of real and generated images and asked to classify them using a Likert scale rating system. Based on these results, ROC curves were plotted and the Area Under the Curve (AUC) was calculated, providing insights into how convincing the generated images appeared to expert human observers [33]. By employing this comprehensive evaluation framework, we aimed to provide a multifaceted assessment of our generative models' performance (Figure *6*). This approach allowed us to evaluate not only the technical quality and diversity of the generated images but also their clinical relevance and potential utility in real-world pathology applications.

(a)-The-working-mechanism-of-the-ROC-framework-and-(b)-the-graphical-user-interface-of-a-website-presented-to-the-pathologists-presenting-the-questionnaire-with-the-images-to-be-classified.

FIGURE 6: (a) The working mechanism of the ROC framework and (b) the graphical user interface of a website presented to the pathologists presenting the questionnaire with the images to be classified.

ROC, Receiver Operating Characteristic

Results

Experimental setup

All experiments were executed on a single NVIDIA A100 GPU with 40 GB HBM2e memory in a SLURM-managed cluster. Model training and evaluation were implemented in PyTorch 2.2 with CUDA 12.4. The diffusion-based models were built using the Hugging Face diffusers library, while StyleGAN3 training followed the official StyleGAN3-T implementation. Model-specific hyperparameters were optimised for stable convergence. For StyleGAN3-T, a batch size of 8 and an ADA gamma value of 8.2 were used, aligning with the recommended configuration to prevent overfitting. For DDPM, a U-Net2D architecture with channel depths ranging from 128 to 512 was employed, trained for 100 epochs using the AdamW optimiser with a 1e-4 learning rate and cosine decay schedule. The noise scheduler consisted of 1,000 timesteps. Attention layers were integrated into the U-Net to preserve spatial dependencies during denoising.

Evaluation was performed using four complementary metrics: FID to assess distributional similarity, SSIM to evaluate perceptual closeness, MSE for pixel-level fidelity, and ROC analysis to assess perceptual realism via expert pathologist review. Random seed settings were fixed at 42 to ensure deterministic training runs. All code and scripts used for preprocessing, training, and evaluation are available upon request for reproducibility.

Moreover, the MIDOG++ dataset was examined to gain insights into the distribution of data across two primary dimensions: species (human and canine) and cancer types (various subcategories). Upon analysing the figures, a substantial imbalance in the dataset becomes apparent. Figure *7* visually depicts the distribution of images across these categories, revealing uneven representation among species and cancer types. For instance, some cancer types, such as

breast carcinoma or melanoma, may dominate in terms of image counts, while others, like neuroendocrine tumours or soft tissue sarcomas, are underrepresented. Similarly, the distribution between species, human and canine, may also show a bias, with one species significantly outweighing the other in image prevalence. These disparities emphasise the challenges posed by unbalanced datasets, particularly when dealing with intricate medical imaging tasks.

This imbalance carries significant implications for downstream applications, including synthetic data generation using GANs and DDPMs. Unbalanced datasets have been known to adversely affect the training and output quality of these generative models. A common phenomenon observed under such circumstances is mode collapse, where the generative models focus predominantly on overrepresented classes, failing to capture the diversity of underrepresented categories. This limitation leads to synthetic images that reflect the biases present in the original dataset, ultimately reducing the variability and comprehensiveness of synthesised mitotic figures. Such outcomes are critical as they directly influence the evaluation metrics and interpretability of models reliant on these synthesised figures. Consequently, the observed imbalance is expected to impact not only the diversity of synthetic outputs but also the performance and generalisability of machine learning models trained on this dataset.

The relevance of addressing this imbalance extends beyond synthetic data generation. Inaccuracies in representation may also influence diagnostic tools and automated image analysis systems built using the MIDOG++ dataset. Uneven performance across categories undermines the reliability of these tools in clinical practice, where consistent and unbiased results are essential for patient care. Mitigating these challenges requires proactive strategies such as data augmentation, class re-sampling, and the implementation of weighted loss functions during model training. By enhancing the representation of underrepresented classes or reducing the dominance of overrepresented ones, these methods can improve the robustness of generative models and ensure equitable performance across categories. Addressing these issues also paves the way for more diverse and accurate datasets, fostering advancements in cancer pathology and automated diagnostic technologies. A balanced dataset, supported by innovative methodologies, is pivotal for achieving comprehensive insights into cancer subtypes while promoting fairness and reliability in medical imaging applications.

Moreover, our comparative evaluation revealed that DDPM is more suitable for applications requiring structural detail, such as classifier training or morphological analysis, while StyleGAN3 excels in generating visually realistic images with superior colour fidelity. Thus, DDPM may be better for technical validation, whereas StyleGAN3 is preferred for visual realism and rapid dataset expansion. Thus, DDPM may be better suited for technical validation and analytic tasks, whereas StyleGAN3 provides an efficient solution for expanding training datasets where visual appeal and diversity are priorities. While the models were evaluated only on the MIDOG++ dataset in this study, future work will assess their generalisability to other datasets, such as the CAMELYON and TUPAC collections. Exploring performance across different data distributions will be essential to confirm the broader applicability of our findings and to refine strategies for integrating synthetic data into diverse histopathological pipelines.

In addition to evaluating model performance, the experiments were monitored for failure modes commonly associated with generative model training, such as overfitting and mode collapse. Throughout our experiments, we did not observe significant signs of mode collapse in either the DDPM or StyleGAN3 models. Synthetic outputs retained diversity across mitotic figure morphology and colour representation, even across multiple training epochs. It is attributed to the diverse composition of the MIDOG++ dataset of several species and cancer types, as well as to the architect of diffusion model and the latest StyleGAN3 design, which are known to mitigate overfitting tendencies compared to earlier GAN variants. Specifically, due to the improved training stability mechanisms, such as StyleGAN3's alias-free design and the intrinsic denoising process of diffusion models, both of which promote better feature disentanglement and discourage the memorisation of specific training examples, thereby reducing the risk of overfitting and mode collapse during image synthesis.


(a)-Distribution-of-the-MIDOG++-dataset-cases-by-species-and-tumour-type-and-(b)-distribution-of-total-cases-sorted-by-tumour-type-and-species.

FIGURE 7: (a) Distribution of the MIDOG++ dataset cases by species and tumour type and (b) distribution of total cases sorted by tumour type and species.

StyleGAN3-T pre-trained model optimisation

StyleGAN3-T is an advanced Alias-Free GAN architecture designed to overcome the "texture sticking" limitation commonly observed in conventional GANs [41]. This limitation arises when generated image features correlate with absolute pixel coordinates rather than with the objects they belong to, resulting in unnatural textures [41]. By addressing this issue, StyleGAN3-T enables the synthesis of mitotic figures that naturally represent their structures without constraints tied to pixel regions. This is particularly relevant for medical imaging tasks, where preserving the integrity and variability of features is crucial for downstream analysis.

The pre-trained StyleGAN3-T model was utilised to investigate the effects of dataset subset sizes on key evaluation metrics, namely FID, SSIM, and MSE. To systematically evaluate these effects, subset sizes of 10K, 20K, 30K, and 40K samples were selected, labelled as experiments 1a-1d. The model's inherent feature correlation capabilities suggested that mitotic figure images across species and cancer types would exhibit minimal variability. This hypothesis was confirmed through visual assessments, where generated images displayed limited variation across experiments. While this consistency indicates the model's robustness, it also highlights challenges in capturing the full diversity present in biological samples. A limitation identified during the experiments was the domain mismatch between the pre-trained StyleGAN3-T model and the target mitotic figure dataset. This mismatch stemmed can be attributed to distinct changes in features, distribution, and labelling between the pre-training domain and the specific requirements of this application [42]. Such inconsistencies can hinder the transfer learning process, potentially compromising the quality of the generated images.

Quantitative findings across subset sizes, detailed in Table *1* and Figure *7*, reveal progressive improvements in evaluation metrics with increasing subset size. Visual assessments (Figure *8*) further support these findings, showcasing noticeable trends in the generated images. For instance, colour scheme accuracy improved as the subset size increased, with fewer instances of abnormally light images compared to conventional stained tissue samples. Additionally, earlier subsets often generated excessive mitotic figures (represented as black features) with limited variety. These challenges were progressively mitigated as the subset size grew, suggesting the model benefits significantly from larger training datasets. Moreover, image quality also improved consistently with larger subset sizes, aligning with NVIDIA's original recommendations of 50-100K images for optimal performance in StyleGAN-based applications [43]. The results suggest that the current subset sizes, while informative, may not fully meet the data requirements for achieving optimal diversity and realism in the generated mitotic figures. This underscores the critical need for larger datasets and more comprehensive training methodologies to enhance the model's output.

The use of StyleGAN3-T in this context highlights both the potential and the limitations of Alias-Free GANs in medical imaging. While promising results were achieved, the identified domain mismatch and data requirements point to areas where further research and development are essential for maximising the utility of this architecture in synthesising accurate and diverse mitotic figures for advanced medical imaging analyses.

Experiment	FID	SSIM	MSE
1a	38.297	0.093	105.581
1b	31.805	0.091	105.735
1c	31.211	0.0938	105.647
1d	34.821	0.933	105.660

TABLE 1: Experiment 1: Optimisation and evaluation of pre-trained StyleGAN3-T model using different sizes subset.

FID, Fréchet Inception Distance; MSE, Mean Squared Error; SSIM, Structural Similarity Index

(a)-Comparison-of-FID,-SSIM,-and-MSE-scores-across-experiments-for-the-StyleGAN3-model-and-(b)-comparison-of-FID,-SSIM,-and-MSE-metrics-across-different-experiments-performed-using-the-StyleGAN3-model.

FIGURE 8: (a) Comparison of FID, SSIM, and MSE scores across experiments for the StyleGAN3 model and (b) comparison of FID, SSIM, and MSE metrics across different experiments performed using the StyleGAN3 model.

FID, Fréchet Inception Distance; MSE, Mean Squared Error; SSIM, Structural Similarity Index

DDPM model optimisation

Model Baseline Establishment

Experiments 2a and 2b served to establish baseline performance for DDPMs utilising the UNet2D architecture. These experiments were specifically designed to inform the optimisation strategy for subsequent analyses. Both models were trained on a subset of 2,500 images over 100 epochs to examine the impact of normalisation on synthesised outputs. The synthesised results demonstrated improvements in image quality following normalisation. These enhancements were further validated through quantitative metrics, including reductions in FID scores and increases in SSIM values, as detailed in Table *2* and Figure *9*.

Interestingly, the MSE metric displayed a marginal increase, indicative of heightened pixel-level differences. This finding appeared to contradict the observed visual improvements, where normalised outputs showed enhanced structural and textural quality. To address this inconsistency, a modified-MSE metric was introduced, tailored specifically for grayscale images. The modified-MSE aimed to isolate structural accuracy from colour information, allowing for a more targeted evaluation of synthesised mitotic figures. By implementing this adjusted metric, the evaluation framework became more aligned with the visual quality improvements, offering greater reliability for subsequent analyses.

The findings from these experiments underscored the importance of normalisation in enhancing DDPM performance. Normalisation techniques were shown to not only improve image synthesis visually but also impact key metrics that influence downstream applications. As a result, normalisation was incorporated into all subsequent experiments, forming a fundamental part of the optimisation strategy for DDPMs in this context. This reflects a critical step in ensuring robust model performance and addressing challenges inherent in medical imaging tasks, such as maintaining structural accuracy while accommodating data variability. The use of these tailored evaluation methods and architectural adjustments highlights the necessity of aligning model outputs with specific domain requirements. In this case, achieving accurate and diverse synthesis of mitotic figures necessitated iterative refinement through normalisation and metric customisation. These strategies pave the way for further enhancements in DDPM applications, fostering more

reliable and generalisable outcomes in medical imaging research. By resolving discrepancies between visual assessments and quantitative metrics, the framework established in Experiments 1.1 and 1.2 provides a strong foundation for advancing generative modelling techniques in this domain.

Metric	FID	SSIM	MSE
2a	276.545	0.0341	5641.925
2a	-	0.03	0.064
2b	128.306	0.0631	7597.894
2b	-	0.068	0.121

TABLE 2: Experiment 2: (a) Training over 100 epochs with no normalisation on 2.5K and (b) training over 100 epochs with random normalisation on 2.5K.

FID, Fréchet Inception Distance; MSE, Mean Squared Error; SSIM, Structural Similarity Index

Synthesised-images-using-the-pre-trained-StyleGAN3.-The-experiments-generate-four-sets-of-images-with-increasing-subset-sizes.-These-are-(a)-10k,-(b)-20k,-(c)-30k,-and-(d)-40k.

FIGURE 9: Synthesised images using the pre-trained StyleGAN3. The experiments generate four sets of images with increasing subset sizes. These are (a) 10k, (b) 20k, (c) 30k, and (d) 40k.

Data Sample Scaling and Parameter Tuning

Building upon baseline experiments, subsequent trials aimed to refine the pipeline for mitotic figure synthesis by gradually increasing the training subset size while optimising preprocessing techniques and model architecture parameters. Experiment 3 expanded the training subset to 5,000 images, continuing the trend of decreasing FID scores and increasing SSIM values, as reported in Table *3* and Figure *9*. Notably, the modified MSE metric for grayscale images decreased, signalling improved structural accuracy despite reductions in colour fidelity. To address challenges related to colour representation, experiment 4a introduced custom normalisation parameters, calculated from dataset-specific mean and standard deviation values. Training with a 10K image subset using these parameters resulted in improved FID scores. However, SSIM and MSE metrics for colour images remained relatively unchanged, and visual assessments showed no substantial enhancements. To explore further improvements, experiment 4b applied colour jitter augmentation to the same 10K subset, achieving continued reductions in FID scores. Additionally, manual evaluations demonstrated a 44-image increase in hue representation accuracy with the inclusion of colour jitter augmentation.

Experiment 5a maintained the use of custom normalisation and colour jitter preprocessing while increasing the subset size to 20K images. Results revealed further improvements in FID scores, though SSIM values slightly decreased for both grayscale and colour images, as detailed in Table *3* and Figures *9*-*10*. To assess the impact of batch size, Experiment 5b replicated Experiment 4b conditions but doubled the batch size from 8 to 16. While larger batch sizes resulted in reduced image quality, evident through higher FID, lower SSIM, and increased MSE metrics (Table *3* and Figure *9*), visual evaluations indicated enhanced colour hue representation (Figure *10a*, c, e). Despite this trade-off, the larger batch configuration was retained for subsequent trials to accommodate computational demands of processing larger subsets while preserving observed visual enhancements.

Experiment 5c maintained the 20K subset while simplifying blockout channel complexity to improve generalisation capability, reduce memory requirements, and accelerate training. Although visual assessments indicated marginal improvements, quantitative metrics pointed to decreased image quality, as reported in Table *3* and Figure *10*. The final Experiment 6a applied the optimised pipeline to a 40K image subset, yielding improved FID scores and visually enhanced cellular structures despite the metric limitations associated with larger batch sizes and modified blockout channels. These findings, coupled with visual observations from Figures *10*, underscore the necessity of expert pathologist validation for specialised histopathological image assessment, as conventional metrics may inadequately capture structural accuracy. To ensure rigorous evaluation, the synthesised results will undergo ROC analysis, where expert pathologists will classify images as synthetic or real. The resulting true positive (TP) and false positive (FP) rates will be used to plot the ROC curve, with the AUC serving as a measure of discrimination ability. In this unique context, lower AUC values (<0.5) are preferred, as they indicate that synthetic images effectively deceive expert evaluators.

Experiment	Images	MSE	FID	SSIM
3a	Colour	114.339	0.072	9666.829
3a	Greyscale	-	0.081	0.100
4a	Colour	106.286	0.0648	9520.910
4a	Greyscale	-	0.065	0.108
4b	Colour	91.43	0.064	9816.089
4b	Greyscale	-	0.064	0.111
5a	Colour	86.49	0.044	10729.624
5a	Greyscale	-	0.052	0.111
5b	Colour	93.60	0.041	10687.211
5b	Greyscale	-	0.049	0.111
6a	Colour	85.433	0.043	10374.123
6a	Greyscale	-	0.049	0.114

TABLE 3: Experiment 3a: Training on 5K, Experiment 4a: Training on 10K subset with data custom normalisation, and 4b: Training on 10K subset with data custom normalisation and colour jitter, Experiment 5a: Training on 20K subset with data custom normalisation and colour jitter, 5b: Training on 20K subset with data custom normalisation and colour jitter and increased batch size, and Experiment 6a: Training on 40K subset with data custom normalisation and colour jitter and increased batch size and modified blockout channel.

FID, Fréchet Inception Distance; MSE, Mean Squared Error; SSIM, Structural Similarity Index

How to cite this article:
Alkhadra R B, Santana P D, Rai T, et al. (January 27, 2026) Comparative Analysis of Generative and Diffusion Models for Mitosis Synthesis in Histopathology Images. Cureus J Comput Sci 3 : es44389-025-00020-8. DOI https://doi.org/10.7759/s44389-025-00020-8


Comparison-of-SSIM-and-MSE-scores-across-experiments-for-grayscale-and-coloured-image-synthesis:-(a,b)-SSIM-scores,-(c,d)-MSE-scores,-and-(e)-comparison-of-FID-score-across-all-experiments-for-coloured-image-synthesis.

FIGURE 10: Comparison of SSIM and MSE scores across experiments for grayscale and coloured image synthesis: (a,b) SSIM scores, (c,d) MSE scores, and (e) comparison of FID score across all experiments for coloured image synthesis.

FID, Fréchet Inception Distance; MSE, Mean Squared Error; SSIM, Structural Similarity Index

ROC study

An ROC study was performed to evaluate the perceptual realism of synthetic images of mitotic figures made by two generative models (StyleGAN3 and DDPM). A mixed set of 30 images (10 real, 10 from StyleGAN3, and 10 from DDPM) was shown to an expert pathologist after they were randomised to minimise ordering effects. The expert was instructed to classify each image as either real or generated. The expert's responses were used to generate ROC curves and compute area under the curve (AUC) metrics. In the first ROC curve (Figure *11*), it compares the expert's ability to distinguish between DDPM generated images and non-DDPM images and GAN generated images and non-GAN images. The AUCs were 0.54 for DDPM versus non-DDPM and 0.58 for GAN versus non-GAN, indicating that the expert had only limited ability to distinguish the generated images from real images. The second ROC curve (Figure *12*) compares real images against all generated images combined. In this case, the expert had an AUC of 0.38.

The ROC study revealed that synthetic mitotic figures created by StyleGAN3 and DDPM models were almost indistinguishable from real images according to expert evaluation, indicating a powerful ability of the models to reproduce important histopathological features (Figure *13*). The AUC values near 0.5 further suggest that the generated samples exhibited enough morphological and textural variance to preclude reliable discrimination even by expert eye. Interestingly, while GAN-generated images produced a slightly higher AUC than DDPM-generated images, the minimal difference did not reflect a consistent qualitative preference, leading to possible conclusions that any separability is not interpreted systematically at the visual level. The significantly lower AUC (0.38) in the combined real versus synthetic classification analyses further bolsters the evidence that the synthetic images, if mixed as in the visual context of the current investigation, did not systematically differ from real samples in anything recognisable.

The equal representations of StyleGAN3 and DDPM outputs along with random samples from a held-out validation set should minimise the possibility of either model artefacts or outlier selection skewing the results. Our results highlight the strong reliability of both generative models, achieving not only surface visual realism but also deeper fidelity to structure that passes expert inspection. Further our results suggest synthetic data from these models could be included in training pipelines without significant artefactual bias and could potentially provide more data diversity and improve generalisation of downstream ML models. The similar expert perception of both models also suggests a convergence on some level of generative model capability in which advancements in architecture, such as diffusion-based denoising procedures and alias-free layers, cumulatively overcome previous obstacles observed in synthetic histopathology. Hence, this set of results offers strong encouragement for the growing uptake of generative augmentation strategies in clinical computational pathology.


Synthesised-images-using-DDPM-model,-according-to-the-respective-experiments-(a-h).

FIGURE 11: Synthesised images using DDPM model, according to the respective experiments (a-h).

DDPM, Denoising Diffusion Probabilistic Models

ROC-curves-comparing-the-expert-pathologist's-ability-to-distinguish-DDPM-generated-images-from-non-DDPM-images-and-GAN-generated-images-from-non-GAN-images.-The-area-under-the-curve-(AUC)-was-0.54-for-DDPM-versus-non-DDPM-and-0.58-for-GAN-versus-non-GAN,-indicating-limited-discriminability-and-performance-close-to-random-guessing-(AUC-=-0.5).

FIGURE 12: ROC curves comparing the expert pathologist's ability to distinguish DDPM-generated images from non-DDPM images and GAN-generated images from non-GAN images. The area under the curve (AUC) was 0.54 for DDPM versus non-DDPM and 0.58 for GAN versus non-GAN, indicating limited discriminability and performance close to random guessing (AUC = 0.5).

AUC, Area Under the Curve; DDPM, Denoising Diffusion Probabilistic Models; GAN, Generative Adversarial Network; ROC, Receiver Operating Characteristic

ROC-curve-evaluating-the-expert-pathologist's-ability-to-differentiate-real-images-from-all-generated-images-combined.-The-obtained-AUC-of-0.38-suggests-the-expert-was-generally-unable-to-reliably-distinguish-synthetic-images-from-real-ones,-further-supporting-the-visual-realism-of-the-generated-mitotic-figures.

FIGURE 13: ROC curve evaluating the expert pathologist's ability to differentiate real images from all generated images combined. The obtained AUC of 0.38 suggests the expert was generally unable to reliably distinguish synthetic images from real ones, further supporting the visual realism of the generated mitotic figures.

AUC, Area Under the Curve; ROC, Receiver Operating Characteristic

Discussion

Experimental analysis

The experimental performance of the proposed generative framework was analysed with consideration to the experimental set-up including evaluation protocol and dataset characteristics. All experiments were conducted under controlled computational conditions using a single NVIDIA A100 GPU, ensuring consistency across training runs and eliminating experimental variability. Deterministic behaviour was enforced through fixed random seeds, allowing observed performance differences to be attributed to model architecture and learning dynamics rather than stochastic effects.

During training, both DDPM and StyleGAN3 were able to achieve convergence under comparable constraints. StyleGAN3 benefited from adaptive discriminator augmentation, which experimentally reduced overfitting tendencies during adversarial training. DDPM, in contrast, relied on iterative denoising across many timesteps, supported by an attention-enhanced U-Net architecture. This architectural distinction influenced convergence behaviour, with DDPM exhibiting slower but more stable optimisation, while StyleGAN3 converged more rapidly with less emphasis on fine structural refinement. Continuous monitoring during training did not reveal instability, divergence, or collapse in either model, indicating that the selected hyperparameters were appropriate for the MIDOG++ dataset.

Evaluation metrics provided additional insights, where distribution measured using FID, and perceptual structure preservation, measured via SSIM, consistently favoured DDPM, indicating closer alignment with real mitotic morphology. This agrees with recent work comparing denoising diffusion models with GANs highlighting them as promising alternative to produce high-quality medical images with enhanced diversity and less artifacts [44]. MSE values were comparable between models, confirming that pixel-level fidelity alone was insufficient to differentiate generative quality in a diagnostically meaningful manner. In addition to these evaluation methods, the ROC analysis validated that both systems achieved high perceptual realism, as expert pathologists were unable to distinguish synthetic images from real samples. Therefore, the agreement between quantitative metrics and expert evaluation supports the robustness of the experimental assessment framework.

Moreover, the dataset composition played a role in shaping experimental outcomes. Analysis of the MIDOG++ dataset revealed pronounced imbalance across tumour types and species, with dominant categories contributing disproportionately to training samples. This imbalance constrained generative diversity, particularly for rare tumour classes, and likely influenced metric sensitivity. Despite this limitation, neither DDPM nor StyleGAN3 exhibited pronounced mode collapse across training epochs. This behaviour can be attributed to architectural advances in both approaches, including denoising-driven feature disentanglement in diffusion models and the alias-free design of StyleGAN3, which experimentally reduced memorisation and over-specialisation. The experimental analysis demonstrates that observed performance differences between DDPM and StyleGAN3 are architecturally driven and reproducible under controlled conditions. DDPM consistently favoured tasks requiring accurate spatial and morphological representation, whereas StyleGAN3 prioritised training efficiency and colour consistency. These findings emerge directly from the experimental configuration, evaluation metrics, and dataset properties, providing a grounded basis for comparative interpretation.

Limitations and ethical considerations

The experimental conclusions of this study are constrained by several factors that shape the interpretation of the results and highlight areas requiring further methodological refinement. The MIDOG++ dataset exhibits pronounced imbalance across tumour types and species, which may bias generative diversity toward overrepresented classes and limit the faithful synthesis of rare morphologies. Such imbalance can influence both quantitative evaluation metrics and the perceived robustness of the generated images, particularly when assessing generalisability beyond the training distribution.

When assessing realism, the current framework primarily emphasises quantitative similarity metrics and expert-based ROC analysis. While these assessments demonstrate visual plausibility, they do not fully capture diagnostic utility within clinical workflows or the effect of synthetic data on downstream tasks such as mitosis detection, grading, or prognostic modelling. The absence of longitudinal and task-driven validation restricts insight into how synthetic images influence model performance over extended training and deployment scenarios.

Ethically, synthetic histopathology data reduces reliance on patient-derived samples and supports privacy preservation, insufficient transparency in dataset composition and generation protocols may risk reinforcing existing diagnostic biases. Clear documentation of data sources, class distributions, and intended application contexts remains essential to ensure equitable and responsible use. Addressing these constraints is necessary to support reliable integration of generative models into computational pathology and to strengthen confidence for their applicability in real-life applications.

Future work

Future experimental work should extend validation beyond the MIDOG++ dataset to rigorously assess model generalisability across heterogeneous histopathology collections, such as the CAncer MEtastases in Lymph nOdes challeNge (CAMELYON) datasets. CAMELYON introduces substantial domain variation, including differences in tissue origin, whole-slide imaging resolution, scanner hardware, and staining protocols, which provides a stringent testbed for evaluating the robustness of generative models trained on mitotic figure-centric datasets. Cross-dataset evaluation will therefore be essential to determine whether the structural advantages observed for diffusion models on MIDOG++ persist when confronted with metastasis-focused annotations, large-scale whole-slide contexts, and broader tumour heterogeneity. Such analysis will help distinguish architecture-driven performance gains from dataset-specific effects and will clarify the extent to which diffusion-based synthesis can generalise across institutions and acquisition pipelines. Demonstrating consistent performance under these conditions would address the aforementioned limitations and strengthen confidence in the applicability of diffusion models for synthetic data generation in real-world computational pathology workflows, where variability in staining, tumour morphology, and imaging conditions is a significant challenge.

How to cite this article:
Alkhadra R B, Santana P D, Rai T, et al. (January 27, 2026) Comparative Analysis of Generative and Diffusion Models for Mitosis Synthesis in Histopathology Images. Cureus J Comput Sci 3 : es44389-025-00020-8. DOI https://doi.org/10.7759/s44389-025-00020-8

Conclusions

This research explored the potential of generative AI, specifically DDPM and StyleGAN3, for synthesising clinically realistic histopathology images containing mitotic figures. DDPM was selected for its ability to preserve fine structural detail, which is critical for histopathological interpretation, while StyleGAN3 was employed for its strength in reproducing biologically realistic colour characteristics. Both models were trained and evaluated using the MIDOG++ dataset, which encompasses multiple tumour types and species and reflects the domain heterogeneity that challenges data-driven diagnostic systems, particularly in the context of limited annotated data. The experimental findings indicate that DDPM generates high-fidelity mitotic figures with consistent morphological structure, making it well suited for applications that prioritise structural accuracy in training data for diagnostic models. In contrast, StyleGAN3 demonstrates superior colour fidelity and computational efficiency, supporting its use in scenarios where visual realism and rapid data generation are required. Quantitative evaluation using FID, SSIM, and MSE, together with expert-based visual assessment, confirms that both approaches can produce synthetic images of sufficient quality to be plausibly integrated into histopathological workflows.

Importantly, these findings highlight a fundamental trade-off between diffusion-based and GAN-based generative models. While diffusion models offer advantages in structural fidelity and robustness under heterogeneous conditions, they incur higher computational cost compared with GAN-based approaches, and their generalisability across institutions, scanners, and staining protocols remains to be systematically established. Conversely, GAN-based models provide faster training and image synthesis with lower computational demands but may be more sensitive to domain variability. An important consideration for future work is the impact of synthetic mitotic figures on downstream mitosis detection performance. While this study demonstrates that both DDPM- and StyleGAN3-generated images achieve high structural and visual realism, their effect on training supervised detection models remains to be quantitatively established. In practice, improvements in image realism do not necessarily translate to gains in detection accuracy, particularly under domain shift. Evaluating whether synthetic data generated by diffusion models or GANs improves robustness, sensitivity, and generalisation of mitosis detectors across tumour types and institutions will therefore be a critical next step. Such task-level validation will be essential to determine the practical utility of generative augmentation beyond perceptual quality alone.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Rahaf B. Alkhadra, Pablo D. Santana, Kevin Wells, Taranpreet Rai

Acquisition, analysis, or interpretation of data: Rahaf B. Alkhadra, Pablo D. Santana, Kevin Wells, Taranpreet Rai

Drafting of the manuscript: Rahaf B. Alkhadra

Critical review of the manuscript for important intellectual content: Rahaf B. Alkhadra, Pablo D. Santana, Kevin Wells, Taranpreet Rai

Supervision: Kevin Wells, Taranpreet Rai

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

References

1. Hassanpour SH, Dehghani M: Review of cancer from perspective of molecular. Journal of Cancer Research and Practice. 2017, 4:127-129. 10.1016/j.jcrpr.2017.07.001
2. Cancer Research UK: Why is early cancer diagnosis important?. https://www.cancerresearchuk.org/about-cancer/spot-cancer-early/why-is-early-diagnosis-important.
3. Nofallah S, Wu W, Liu K, Ghezloo F, Elmore JG, Shapiro LG: Automated analysis of whole slide digital skin biopsy images. Frontiers in Artificial Intelligence. 2022, 5:1005086. 10.3389/frai.2022.1005086
4. Institute of Medicine (US) Committee on Cancer Control in Low- and Middle-Income Countries, Sloan FA, Gelband H: Cancer Control Opportunities in Low- and Middle-Income Countries. National Academies Press, Washington, DC; 2007.
5. Yoo D, Divard G, Raynaud M, et al.: A machine learning-driven virtual biopsy system for kidney transplant patients. Nature Communications. 2024, 15:554. 10.1038/s41467-023-44595-z
6. Cooper GM: The Cell: A Molecular Approach. 2nd edition. Sinauer Associates, Sunderland, MA; 2000.
7. Dominguez-Brauer C, Thu KL, Mason JM, Blaser H, Bray MR, Mak TW: Targeting mitosis in cancer: Emerging strategies. Molecular Cell. 2015, 60:524-536. 10.1016/j.molcel.2015.11.006
8. Jahanifar M, Shephard A, Zamanitajeddin N, Graham S, Raza SEA, Minhas F, Rajpoot N: Mitosis detection, fast and slow: Robust and efficient detection of mitotic figures. Medical Image Analysis. 2024, 94:103132. 10.1016/j.media.2024.103132
9. Aljehani MR, Alamri FH, Elyas MEK, Almohammadi AS, Alanazi ASA, Alharbi MA: The importance of histopathological evaluation in cancer diagnosis and treatment. International Journal of Health Sciences. 2023, 7:3614-3623. 10.53730/ijhs.v7ns1.15270
10. Ibrahim A, Lashen A, Toss M, Mihai R, Rakha E: Assessment of mitotic activity in breast cancer: revisited in the digital pathology era. Journal of Clinical Pathology. 2022, 75:365-372. 10.1136/jclinpath-2021-207742
11. Pan X, Lu Y, Lan R, Liu Z, Qin Z, Wang H, Liu Z: Mitosis detection techniques in H&E stained breast cancer pathological images: A comprehensive review. Computers & Electrical Engineering. 2021, 91:107038. 10.1016/j.compeleceng.2021.107038
12. Ibrahim A, Lashen AG, Katayama A, Mihai R, Ball G, Toss MS, Rakha EA: Defining the area of mitoses counting in invasive breast cancer using whole slide image. Modern Pathology. 2022, 35:739-748. 10.1038/s41379-021-00981-w
13. Smits LJH, van Lieshout AS, Bosker RJI, et al.: Clinical consequences of diagnostic variability in the histopathological evaluation of early rectal cancer. European Journal of Surgical Oncology. 2023, 49:1291-1297. 10.1016/j.ejso.2023.02.008
14. Khatab Z, Hanna K, Rofaeil A, Wang C, Maung R, Yousef GM: Pathologist workload, burnout, and wellness: connecting the dots. Critical reviews in clinical laboratory sciences. 2023, 61:254-274. 10.1080/10408363.2023.2285284
15. Johnston PW, Fioratou E, Flin R: Non-technical skills in histopathology: definition and discussion. Histopathology. 2011, 59:359-367. 10.1111/j.1365-2559.2010.03710.x
16. Dlamini Z, Francies FZ, Hull R, Marima R: Artificial intelligence (AI) and big data in cancer and precision oncology. Computational and Structural Biotechnology Journal. 2020, 18:2300-2311. 10.1016/j.csbj.2020.08.019
17. Khalifa M, Albadawy M: Artificial intelligence for clinical prediction: Exploring key domains and essential functions. Computer Methods and Programs in Biomedicine Update. 2024, 5:100148. 10.1016/j.cmpbup.2024.100148
18. Rai T, Morisi A, Bacci B, et al.: Keeping pathologists in the loop and an adaptive F1-score threshold method for mitosis detection in canine perivascular wall tumours. Cancers. 2024, 16:644. 10.3390/cancers16030644
19. Li X: CNN-based cancer image diagnosis: current progress and future directions. Applied and Computational Engineering. 2024, 106:7-12. 10.54254/2755-2721/106/20241262
20. Sahu A, Das PK, Meher S: High accuracy hybrid CNN classifiers for breast cancer detection using mammogram and ultrasound datasets. Biomedical Signal Processing and Control. 2023, 80:104292. 10.1016/j.bspc.2022.104292
21. Tembhurne JV, Hebbar N, Patil HY, Diwan T: Skin cancer detection using ensemble of machine learning and deep learning techniques. Multimedia Tools and Applications. 2023, 82:27501-27524. 10.1007/s11042-023-14697-3
22. Joshua ESN, Bhattacharyya D, Chakkravarthy M, Byun Y-C: 3D CNN with visual insights for early detection of lung cancer using gradient-weighted class activation. Journal of Healthcare Engineering. 2021, 2021:6695518. 10.1155/2021/6695518
23. Kashyap A, Rapsomaniki MA, Barros V, et al.: Quantification of tumor heterogeneity: from data acquisition to metric generation. Trends in Biotechnology. 2022, 40:647-676. 10.1016/j.tibtech.2021.11.006

How to cite this article:
Alkhadra R B, Santana P D, Rai T, et al. (January 27, 2026) Comparative Analysis of Generative and Diffusion Models for Mitosis Synthesis in Histopathology Images. Cureus J Comput Sci 3 : es44389-025-00020-8. DOI https://doi.org/10.7759/s44389-025-00020-8

24. Chen RJ, Wang JJ, Williamson DF, et al.: Algorithmic fairness in artificial intelligence for medicine and healthcare. Nature Biomedical Engineering. 2023, 7:719-742. 10.1038/s41551-023-01056-8

25. Perez-Lopez R, Ghaffari Laleh N, Mahmood F, Kather JN: A guide to artificial intelligence for cancer researchers. Nature Reviews Cancer. 2024, 24:427-441. 10.1038/s41568-024-00694-7

26. McGraw D, Mandl KD: Privacy protections to encourage use of health-relevant digital data in a learning health system. npj Digital Medicine. 2021, 4:2. 10.1038/s41746-020-00362-8

27. Bai B, Yang X, Li Y, Zhang Y, Pillar N, Ozcan A: Deep learning-enabled virtual histological staining of biological samples. Light: Science & Applications. 2023, 12:57. 10.1038/s41377-023-01104-7

28. Song S, Liang Y, Wu J, Lai Y-K, Qin Y: Feature proliferation -- the "cancer" in StyleGAN and its treatments. arXiv. 2023, 10.48550/arXiv.2310.08921

29. Jiang H, Imran M, Zhang T, Zhou Y, Liang M, Gong K, Shao W: Fast-DDPM: Fast denoising diffusion probabilistic models for medical image-to-image generation. arXiv. 2025, 10.48550/arXiv.2405.14802

30. Zheng Z, Wang M, Fan C, et al.: Light&fast generative adversarial network for high-fidelity CT image synthesis of liver tumor . Computer Methods and Programs in Biomedicine. 2024, 254:108252. 10.1016/j.cmpb.2024.108252

31. Dede R, Bilgin G: Deep learning based discovery of new breast cancer stages: StyleGAN3 and Swin Transformer approach. 2024 Medical Technologies Congress (TIPTEKNO), Mugla, Turkiye. 2024, 1-4. 10.1109/TIPTEKNO63488.2024.10755469

32. Alblwi A, Makkawy S, Barner KE: D-DDPM: deep denoising diffusion probabilistic models for lesion segmentation and data generation in ultrasound imaging. IEEE Access. 2025, 13:41194-41209. 10.1109/ACCESS.2025.3548128

33. Rai T, Gola C, Hernández M, et al.: Evaluating diffusion model generated synthetic histopathology image data against authentic digital pathology images. Medical Imaging 2024: Digital and Computational Pathology. 2024, 12933:1-7. 10.1117/12.3006285

34. Song T, Wen R, Zhang L: RoughSet-DDPM: An image super-resolution method based on rough set denoising diffusion probability model. Technical Gazette. 2024, 31:162-170. 10.17559/TV-20230717000808

35. Aubreville M, Wilm F, Stathonikos N, et al.: A comprehensive multi-domain dataset for mitotic figure detection. Scientific Data. 2023, 10:484. 10.1038/s41597-023-02327-4

36. Li YA, Han C, Mesgarani N: StyleTTS: A style-based generative model for natural and diverse text-to-speech synthesis. arXiv. 2023, 10.48550/arXiv.2205.15439

37. Zhang S, Liu L, Li G, Du Y, Wu X, Song Z, Li X: Diffusion model-based image generative method for quality monitoring of direct grain harvesting. Computers and Electronics in Agriculture. 2025, 233:110130. 10.1016/j.compag.2025.110130

38. Sauer A, Karras T, Laine S, Geiger A, Aila T: StyleGAN-T Unlocking the power of GANs for fast large-scale text-to-image synthesis. arXiv. 2023, 10.48550/arXiv.2301.09515

39. Simeon G, De Fabritiis G: TensorNet: cartesian tensor representations for efficient learning of molecular potentials. NIPS '23: Proceedings of the 37th International Conference on Neural Information Processing Systems. 2023, 37334-37353.

40. Ibrahim A, Jahanifar M, Wahab N, et al.: Artificial intelligence-based mitosis scoring in breast cancer: clinical application. Modern Pathology. 2024, 37:100416. 10.1016/j.modpat.2023.100416

41. Habeeb D, Alhassani AH, Abdullah LN, Der CS, Alasadi LKQ: Advancements and challenges: a comprehensive review of GAN-based models for the mitigation of small dataset and texture sticking issues in fake license plate recognition. Engineering, Technology & Applied Science Research. 2024, 14:18401-18408. 10.48084/etasr.8870

42. Fu J, Li S, Jiang Y, et al.: StyleGAN-Human: a data-centric odyssey of human generation. Computer Vision - ECCV 2022. Avidan S, Brostow G, Cissé M, Farinella GM, Hassner T (ed): Springer, Cham; 2022. 13676:1-19. 10.1007/978-3-031-19787-1_1

43. Salian I: NVIDIA Research Achieves AI Training Breakthrough Using Limited Datasets. (2020). Accessed: December 7, 2025: https://blogs.nvidia.com/blog/neurips-research-limited-data-gan.

44. Müller-Franzes G, Niehues JM, Khader F, et al.: A multimodal comparison of latent denoising diffusion probabilistic models and generative adversarial networks for medical image synthesis. Scientific Reports. 2023, 13:12098. 10.1038/s41598-023-39278-0

How to cite this article:
Alkhadra R B, Santana P D, Rai T, et al. (January 27, 2026) Comparative Analysis of Generative and Diffusion Models for Mitosis Synthesis in Histopathology Images. Cureus J Comput Sci 3 : es44389-025-00020-8. DOI https://doi.org/10.7759/s44389-025-00020-8