HistoSyn: Histomorphology-Focused Pathology Image Synthesis

Chong Yin[1], Siqi Liu[2], Vincent Wai-Sun Wong[3], Pong C. Yuen[1]

[1]Hong Kong Baptist University, Hong Kong
[2]Shenzhen Research Institute of Big Data, Shenzhen
[3]Chinese University of Hong Kong, Hong Kong
{chongyin, pcyuen}@comp.hkbu.edu.hk,siqiliu@sribd.cn,wongv@cuhk.edu.hk

Abstract. Examining pathology images through visual microscopy is widely considered the most reliable method for diagnosing different medical conditions. Although deep learning-based methods show great potential for aiding pathology image analysis, they are hindered by the lack of accessible large-scale annotated data. Large text-to-image models have significantly advanced the synthesis of diverse contexts within natural image analysis, thereby expanding existing datasets. However, the variety of histomorphological features in pathology images, which differ from that of natural images, has been less explored. In this paper, we propose a histomorphology-focused pathology image synthesis (HistoSyn) method. Specifically, HistoSyn constructs instructive textural prompts from spatial and morphological attributes of pathology images. It involves analyzing the intricate patterns and structures found within pathological images and translating these visual details into descriptive prompts. Furthermore, HistoSyn presents new criteria for image quality evaluation focusing on spatial and morphological characteristics which have a stronger correlation to the performance of down-stream tasks. Experiments have demonstrated that our method can achieve a diverse range of high-quality pathology images, with a focus on histomorphological attributes. The code is available at https://github.com/7LFB/HistoSyn.

Keywords: Pathology image synthesis · Histomorphology features.

1 Introduction

Pathology image analysis is widely regarded as the gold standard for diagnosing various medical conditions, such as liver cancer [23] and breast cancer [21], due to its ability to provide visual evidence at the cellular level. The intricate tissue structures in pathology images pose considerable challenges for pathological diagnosis due to variability among observers and the subtle histological differences [10, 12]. In addition, the growing number of patients has significantly burdened the current medical system [29].

AI-based methods emerge as a promising alternative to augment the diagnostic process with more cost-effective, and capable of handling high patient volumes

without compromising diagnostic accuracy [13,25,27,28]. Utilizing deep-learning-based methods necessitates extensive datasets with annotations, which presents a significant challenge in the medical domain. Collecting large datasets for effective training can be quite costly. Recently, the synthesis of images has shown great potential in generating natural images [1, 11, 22]. Generative Adversarial Networks (GANs) [6] have shown promise in generating realistic images [2, 16], but they are also prone to mode collapse and training instability. The latent diffusion model (LDM) [18] is known for generating high-resolution synthetic images a wider range of context based on text prompts.

When applying existing diffusion-based methods to synthetic pathology images, the model may suffer from low diversity and quality due to the distinct difference between nature images and pathology images. The diversity in nature images emphasizes the same subject in different scenes, poses, and views [3, 14, 19]. However, the diversity in pathology images emphasizes the same diagnosis with different spatial distributions among tissue/cellular structures, various histological findings in size, and shape [26, 27]. [15, 17] solely focused on studying morphological attributes indirectly without specifying quantitative attributes. How to effectively enhance and assess the synthesis pathology image with a focus on histomorphology attributes has been rarely explored.

Our main insight is to explore histomorphological attributes in a similar manner to that of pathologists when analyzing pathological images as follows:

- **How to define the diversity in pathology images?** Generating synthetic pathology images requires a process that ensures both diversity and realism in the data. To define diversity in pathology images, one must consider the range of variations in tissue architecture, cell morphology, staining patterns, and disease manifestations that are representative of the real-world variability seen in human tissues. This requires a deep understanding of the pathological conditions being modeled and the associated histomorphological features that are critical for accurate diagnosis.
- **How to evaluate the quality of synthetic pathology image dataset?** The quality evaluation should encompass a diverse range of histomorphological attributes, including but not limited to spatial attributes (e.g., cellular structure, tissue organization), and morphological features (e.g., area). A higher quality synthesized image should closely mirror those of the real dataset with respect to the specific histomorphology attribute distribution.

In this paper, we propose a histomorpholoy-focused pathology image synthesis method (HistoSyn). HistoSyn specifically explores spatial and morphological attributes customized for pathology images and converts these visual cues into prompts for pathology image synthesis. Furthermore, we introduce a new evaluation metric to assess the quality with respect to the spatial and morphology attributes. This new metric aims to quantify the histomorphological attributes that are crucial for accurate diagnosis and provide an interpretable angle of view to evaluate synthetic pathology image dataset. Our contributions are:

- We propose a method for synthesizing pathology images with a focus on histomorphology. The proposed method enhances image synthesis by in-

corporating a wide range of histomorphology attributes depicted through statistical measures.
- We design a new reliable metric to assess the quality of synthetic pathology images based on histomorphology attributes.
- Experiment results demonstrate the effectiveness of our method where the synthesized images help improve the overall classification accuracy. The new metric also shows a stronger correlation to the performance.

2 Proposed Method

Overview. Figure 1 illustrates our proposed histomorphology-focused pathology image synthesis and evaluation method. The image synthesis procedure consists of a histomorphology-focused prompting module and a base Latent Diffusion Model (LDM) [18] including the Variational Autoencoder $\{\mathcal{E}, \mathcal{D}\}$, the U-Net denoiser ϵ_θ, and the text encoder τ_θ. Given a training image x, the training procedure of LDM is achieved by minimizing the true noise ϵ and predicted noise ϵ_θ conditioned on text prompts y:

$$\mathcal{L} = \mathbb{E}_{\mathcal{E}(x),\epsilon\sim\mathcal{N}(0,1),t}[||\epsilon - \epsilon_\theta(z_t, t, \tau_\theta(y))||_2^2] \tag{1}$$

When the model is well-trained, we can generate synthetic pathology images by forwarding a random noise latent variable z_t along with the user-provided conditioned text prompt y:

$$\hat{x} = \mathcal{D}(\epsilon_\theta(z_t, t, \tau_\theta(y))), z_t \sim \mathcal{N}(0, 1) \tag{2}$$

The key to generating high-quality images lies in enriching their diversity through text prompts y which describe the tailored histomorphology attributes.

2.1 Histomorphology-Focused Prompting

In this section, we will explore how to effectively describe the histomorphological attributes specific to pathology images in order to generate more varied and realistic images. Drawing inspiration from pathologists, we analyze spatial and morphological attributes. Given a pathology image x, it is firstly fed into a tissue segmenter $\mathcal{S}eg$ to get main objects (e.g., nuclei, fat droplets) $O = \{o_1, o_2, \ldots, o_N\}$ observed in the pathology images.

$$O = \mathcal{S}eg(x) \tag{3}$$

Then, we perform quantitative analysis focusing on spatial and morphological attributes described by statistical measures.

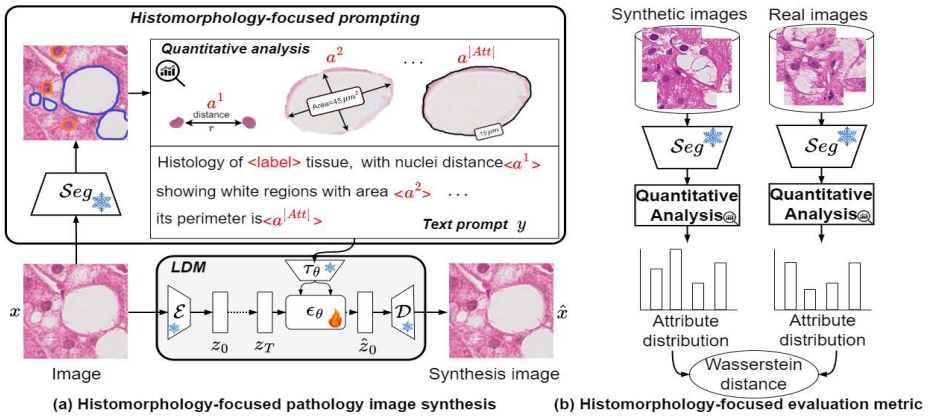

Fig. 1: Overview of the proposed histomorphology-focused pathology image synthesis (a) and evaluation metric (b). It enriches the text prompt by incorporating histomorphology attributes specific to pathology images. The quality evaluation is conducted based on the distribution of its histomorphological attributes.

Spatial attributes. The spatial arrangement of objects in a pathology image is crucial for accurate diagnosis. Based on the K-function [5], which is a statistical measure that assesses the displacement of objects o_i and o_j within a certain distance r, assisting in the analysis of the image:

$$K(r) = \frac{1}{\lambda}\pi r^2 \Big(\sum_{o_i, o_j \in O} [\![d(o_i, o_j) \leq r]\!] \Big)^2 \tag{4}$$

where $[\![\cdot]\!]$ returns 1 if the input is true. The symbol λ represents a constant, and the distance function d calculates the Euclidean distance between two objects. Different choices of objects result in different K-functions. Then l-th kind of spatial attributes a^l of image x depicted by K-function can be expressed by the average distance:

$$a^l = \frac{\sum_{r=0}^{d_{max}} K(r) * r}{\sum_{r=0}^{d_{max}} K(r)} \tag{5}$$

Morphology attributes. The morphological attributes are utilized for measuring and analyzing the shape, size of objects o_i observed in image x. Without loss of generality, we denote ϕ as the k-th kind of attribute calculator. Then, the morphological attribute a^k of image x can be defined as the average attribute across all observed objects:

$$a^k = \frac{1}{N} \sum_{i=1}^{N} (\phi(o_i)) \tag{6}$$

After analyzing the spatial and morphological attributes specific to the objects observed in the pathology images, a comprehensive quantitative assessment of pathology images x can be obtained using attribute sets $Attr = [a^1, \ldots, a^M]$, where M represents the number of attributes being considered.

2.2 Histomorphology-Focused Evaluation Metric

The quality evaluation should delve into the intricate analysis of tissue structures and cellular details, emphasizing the significance of histomorphological examination in diagnosing various diseases. The cell size, shape, and spatial distribution which are indicative of pathological conditions are analyzed and formed as evaluation metric to identify abnormalities.

Given two sets of l-th attribute values, $A = \{a_1^l, a_2^l, \ldots, a_n^l\}$ obtained from the real dataset X, and $B = \{b_1^l, b_2^l, \ldots, b_m^l\}$ obtained from the synthetic dataset $\hat{X}$, we choose the the Wasserstein distance [20] to measure of the discrepancy between the two distributions that these lists represent. we first need to construct the cumulative distribution functions (CDFs) of both lists. Let $F_A(x)$ and $F_B(x)$ be the CDFs of lists A and B, respectively. The Wasserstein distance of order p between A and B is defined as:

$$HistD_l := \left(\int_{-\infty}^{\infty} |F_A(x) - F_B(x)|^p dx \right)^{\frac{1}{p}} \tag{7}$$

When considering N_{attr} kinds of attributes, the overall distance can be calculated simply by taking the average between all attribute distances:

$$\overline{HistD} = \frac{1}{N_{attr}} \sum_{l=1}^{N_{attr}} HistD_l \tag{8}$$

The resulting Wasserstein distance will provide insight into how well the synthetic images replicate the spatial or morphological characteristics of the real images. A smaller Wasserstein distance indicates a closer match between the two distributions and, consequently, a higher quality of the synthetic dataset in mirroring real histomorphology attributes. It enables the assessment of histomorphological attribute diversity and provides an interpretable method for image synthesis.

3 Experimental Results

3.1 Experiment Settings

We assessed the performance using **Liver-NAS** dataset [30], which contains image tiles extracted from WSIs with corresponding labels indicating histological findings. Experiments are conducted using 5-fold cross-validation and the average performance is reported. Only 10% of the training data was used to simulate low data availability in the medical field, with a 1:1 ratio of real to synthetic data. Details of the dataset are in the supplementary materials.

Table 1: Comparison of recognition performance using F1 score.

Methods	Others	Inflammation	Ballooning	Steatosis	Macro-Average
Baseline [18]	82.42±1.49	82.10±2.13	68.80±4.37	94.74±1.22	82.01±1.77
Morph [17]	83.23±1.31	81.99±2.91	79.45±3.77	**96.26**±0.34	85.24±1.50
Ours	**84.11**±2.68	**85.75**±1.47	**81.11**±4.31	95.90±1.36	**86.72**±1.78

Evaluation metrics. For classification experiments, we choose F1 score as our performance metric. To evaluate the quality of synthesized images, we utilized the widely recognized metric Frechet Inception Distance (FID) [9]. Additionally, we assessed the quality through the proposed metric that measures the similarity between the distributions of real and synthetic data, focusing specifically on spatial or morphological attributes.

Training details. We finetune the latent diffusion model (LDM) [19] for pathology image synthesis. All models were obtained from the HuggingFace model repository [24]. We used the checkpoint version 1.5 (runwayml/stable-diffusion-v1-5). We only fine-tuned the denoiser module ϵ_θ while keeping both the VAE and CLIP models frozen. As for the downstream task, we choose ResNet-50 [8] as the backbone for training. We use HoVer-Net [7] for nuclei segmentation. In H&E stained pathology images, white regions indicate the presence of fatty cells or vessels. Two types of objects are considered (nuclei and white regions), with a total of $M = 7$ histomorphological attributes (defined in Supp.), resulting in 7 evaluation metrics $HistD_1, ..., HistD_7$. When fine tune the Latent Diffusion Model (LDM) [18], we employs a learning rate of 5×10^{-5} with maximum 400 steps. The batch size is 4. After training, each prompt is used to generate one image. When training ResNet-50, we adhered to the standard training scheme outlined in [8].

3.2 Comparison with SOTA Methods

We finetune LDM [18] with a vanilla label prompt as the baseline. We also compare our method with Morphology-enriched (Morph) [17], which enriched the prompt with morphology type information learned from k-means clustering.

Histological findings recognition. Table 1 presents the F1 scores for various approaches in identifying histological findings. Our approach achieves the highest macro-average F1 score of 86.72. The 'Baseline' method shows consistent performance but dips to 68.80 in Ballooning. The 'Morph' method improves in the 'Others' category (83.23) but lags in Inflammation and Ballooning, with a macro average of 85.24. Our method excels in Inflammation (85.75) and Ballooning (81.11), and scores 95.90 in Steatosis. This demonstrates the benefits of incorporating quantitative attributes, achieving the highest overall performance.

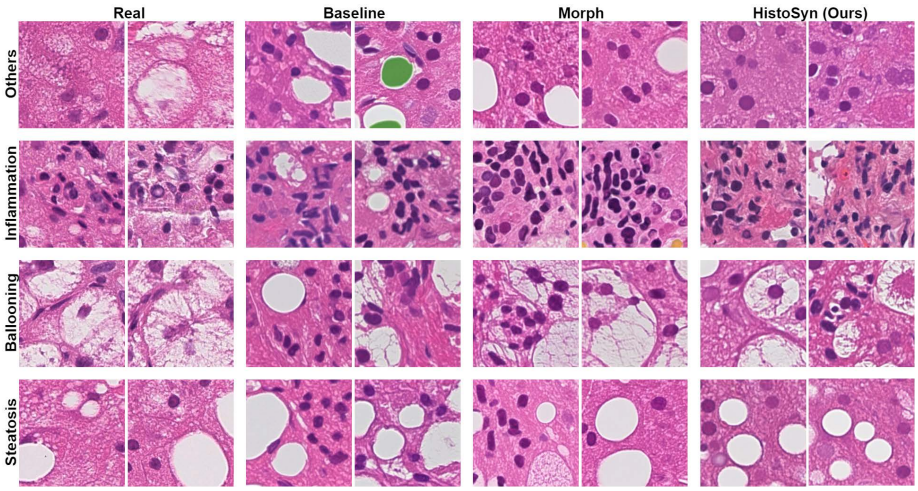

Fig. 2: Our method excels in synthesizing pathology images by capturing diverse histomorphological attributes, such as white region areas and the spatial arrangement of surrounding nuclei, while preserving essential histological findings like inflammation (accumulated nuclei), ballooning (distorted cells), and steatosis (fatty cells) without mixing them in 'Others'.

3.3 Quality Evaluation of Synthesis Image

We use visual inspection and quantitative analysis to determine how faithfully the synthesized imagery represents real datasets.

Quantitative analysis. Table 2 shows that the HistoSyn method outperforms others in both FID and $\overline{HistD}$ scores. HistoSyn achieves an FID of 3.17, significantly lower than the Baseline (4.75) and Morph method (4.23), indicating greater similarity to real images. The mean values for the distribution of real versus synthetic data across the 7 attributes are as follows: 103.6 (real) vs 117.2 (synthetic), 121.1 (real) vs 123.1 (synthetic), 111.2 (real) vs 124.1 (synthetic), 118.1 (real) vs 124.4 (synthetic), 1791.3 (real) vs 1611.4 (synthetic), 0.6 (real) vs 0.7 (synthetic), and 69.0 (real) vs 72.2 (synthetic). For the metric $\overline{HistD}$, which calculates the difference between the distributions, HistoSyn scores 27.55, which is better than the Baseline's 38.08 and the Morph method's 56.44. These results demonstrate HistoSyn's superior ability to generate high-fidelity histological images, closely approximating real image quality.

Visual comparison. Figure 2 demonstrates the synthesis images using different methods. Different methods reveal distinct characteristics. The 'Baseline' images, while somewhat resembling real histology images, lack the nuanced morphological details present in the 'Real' column. The 'Morph' images show an improvement in structural detail, suggesting a more sophisticated approach to

Table 2: Evaluation on the quality of synthetic images.

Methods	FID ↓	$\overline{HistD}$ (Eq 8) ↓
Baseline [18]	4.75±1.17	38.08±14.81
Morph [17]	4.23±0.82	56.44±21.35
HistoSyn (Ours)	**3.17**±0.73	**27.55**±8.90

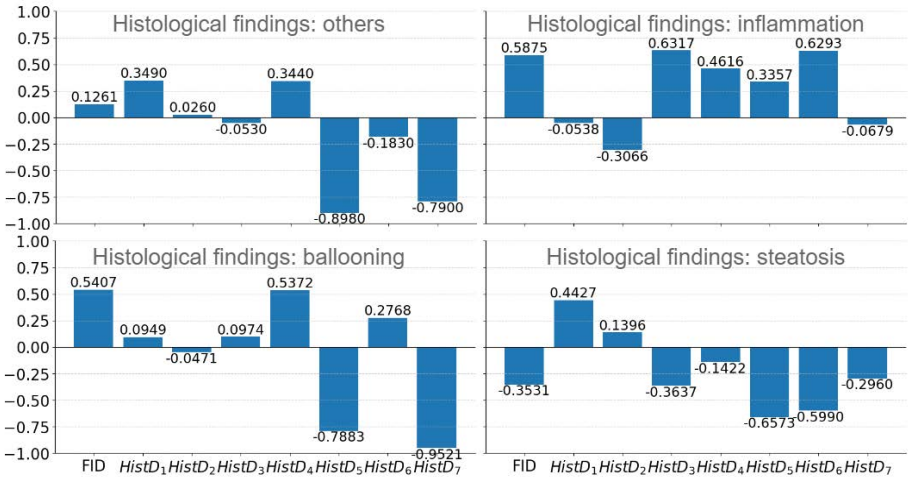

Fig. 3: The Pearson correlation coefficients between various evaluation metrics and the recognition performance of different histological findings.

synthetic image generation that better captures the complexity of biological tissue. Most notably, the 'HistoSys (Ours)' column presents images that are closer to the 'Real' samples, with a clear enhancement in the diversity of the tissue structures depicted.

3.4 Quality Metrics and Performance Correlation.

We evaluate the proposed quality metric and FID to see which better represents the quality of generated images and correlates more strongly with downstream task performance. We use the Pearson correlation coefficient [4] to assess the relationship between quality metrics and classification task performance.

As shown in Figure 3, the proposed metric provides a stronger indication of performance in recognizing histological findings compared to the FID metric. A lower FID, despite positive Pearson correlation, indicates realistic images but doesn't aid downstream tasks, unlike HistD, which helps recognize inflammation and ballooning.

4 Conclusion

In this paper, we propose a histomorphology-focused image synthesis method (HistoSyn) to enable high-quality pathology image generation. HistoSyn enriches the prompts with the histomorphology attributes and guides the synthesis process, ensuring that the generated images maintain important diagnostic features. We also introduce new criteria for image quality evaluation focusing on spatial and morphological characteristics which have a stronger correlation to the classification performance.

Acknowledgments. The first author was supported by the HKBU Studentship and this work was partially supported by the GuangDong Basic and Applied Basic Research Foundation 2023A1515110706.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Alomar, K., Aysel, H.I., Cai, X.: Data augmentation in classification and segmentation: A survey and new strategies. Journal of Imaging **9**(2), 46 (2023)
2. Beji, A., Blaiech, A.G., Said, M., Abdallah, A.B., Bedoui, M.H.: An innovative medical image synthesis based on dual gan deep neural networks for improved segmentation quality. Applied Intelligence **53**(3), 3381–3397 (2023)
3. Chen, W., Hays, J.: Sketchygan: Towards diverse and realistic sketch to image synthesis. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 9416–9425 (2018)
4. Cohen, I., Huang, Y., Chen, J., Benesty, J., Benesty, J., Chen, J., Huang, Y., Cohen, I.: Pearson correlation coefficient. Noise reduction in speech processing pp. 1–4 (2009)
5. Dixon, P.: Ripley's k function (2001)
6. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. Advances in neural information processing systems **27** (2014)
7. Graham, S., Vu, Q.D., Raza, S.E.A., Azam, A., Tsang, Y.W., Kwak, J.T., Rajpoot, N.: Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. Medical Image Analysis p. 101563 (2019)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
9. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium. Advances in neural information processing systems **30** (2017)
10. Juluri, R., Vuppalanchi, R., Olson, J., Ünalp, A., Van Natta, M.L., Cummings, O.W., Tonascia, J., Chalasani, N.: Generalizability of the nonalcoholic steatohepatitis clinical research network histologic scoring system for nonalcoholic fatty liver disease. Journal of clinical gastroenterology **45**(1), 55–58 (2011)
11. Kang, M., Zhu, J.Y., Zhang, R., Park, J., Shechtman, E., Paris, S., Park, T.: Scaling up gans for text-to-image synthesis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10124–10134 (2023)
12. Li, Y.Y., Zheng, T.L., Xiao, S.Y., Wang, P., Yang, W.J., Jiang, L.L., Chen, L.L., Sha, J.C., Jin, Y., Chen, S.D., et al.: Hepatocytic ballooning in non-alcoholic steatohepatitis: Dilemmas and future directions. Liver International (2023)
13. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. Medical image analysis **42**, 60–88 (2017)
14. Men, Y., Mao, Y., Jiang, Y., Ma, W.Y., Lian, Z.: Controllable person image synthesis with attribute-decomposed gan. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5084–5093 (2020)
15. Moghadam, P.A., Van Dalen, S., Martin, K.C., Lennerz, J., Yip, S., Farahani, H., Bashashati, A.: A morphology focused diffusion probabilistic model for synthesis of histopathology images. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 2000–2009 (2023)
16. Nie, D., Trullo, R., Lian, J., Petitjean, C., Ruan, S., Wang, Q., Shen, D.: Medical image synthesis with context-aware generative adversarial networks. In: Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part III 20. pp. 417–425. Springer (2017)

17. Osorio, P., Jimenez-Perez, G., Montalt-Tordera, J., Hooge, J., Duran-Ballester, G., Singh, S., Radbruch, M., Bach, U., Schroeder, S., Siudak, K., et al.: Latent diffusion models with image-derived annotations for enhanced ai-assisted cancer diagnosis in histopathology. arXiv preprint arXiv:2312.09792 (2023)
18. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 10684–10695 (June 2022)
19. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 10684–10695 (2022)
20. Rüschendorf, L.: The wasserstein distance and approximation theorems. Probability Theory and Related Fields **70**(1), 117–129 (1985)
21. Sokolova, A., Johnstone, K., McCart Reed, A., Simpson, P., Lakhani, S.: Hereditary breast cancer: Syndromes, tumour pathology and molecular testing. Histopathology **82**(1), 70–82 (2023)
22. Tao, M., Bao, B.K., Tang, H., Xu, C.: Galip: Generative adversarial clips for text-to-image synthesis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14214–14223 (2023)
23. Trabucco, B., Doherty, K., Gurinas, M., Salakhutdinov, R.: Effective data augmentation with diffusion models. arXiv preprint arXiv:2302.07944 (2023)
24. Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., et al.: Huggingface's transformers: State-of-the-art natural language processing. arXiv preprint arXiv:1910.03771 (2019)
25. Yin, C., Liu, S., Lyu, F., Lu, J., Darkner, S., Wong, V.W.S., Yuen, P.C.: Xfibrosis: Explicit vessel-fiber modeling for fibrosis staging from liver pathology images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11282–11291 (2024)
26. Yin, C., Liu, S., Shao, R., Yuen, P.C.: Focusing on clinically interpretable features: selective attention regularization for liver biopsy image classification. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part V 24. pp. 153–162. Springer (2021)
27. Yin, C., Liu, S., Wong, V.W.S., Yuen, P.C.: Learning sparse interpretable features for nas scoring from liver biopsy images. In: International Joint Conference on Artificial Intelligence (2022), `https://api.semanticscholar.org/CorpusID:250639956`
28. Yin, C., Liu, S., Zhou, K., Wong, V.W.S., Yuen, P.C.: Prompting vision foundation models for pathology image analysis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11292–11301 (2024)
29. Younossi, Z., Anstee, Q.M., Marietti, M., Hardy, T., Henry, L., Eslam, M., George, J., Bugianesi, E.: Global burden of nafld and nash: trends, predictions, risk factors and prevention. Nature reviews Gastroenterology & hepatology **15**(1), 11–20 (2018)
30. Zhou, Y.J., Gao, F., Liu, W.Y., Wong, G.L.H., Mahadeva, S., Raihan Nik Mustapha, N., Wang, X.D., Chan, W.K., Wong, V.W.S., Zheng, M.H.: Screening for compensated advanced chronic liver disease using refined baveno vi elastography cutoffs in asian patients with nonalcoholic fatty liver disease. Alimentary pharmacology & therapeutics **54**(4), 470–480 (2021)