Towards the on-demand Whole Slide Image generation: Prostate patch synthesis through a conditional progressive growing GAN

Alejandro Golfe*, Rocío del Amor*, Adrián Colomer*†, María A. Sales‡, Liria Terradez‡, Valery Naranjo*
†*ValgrAI – Valencian Graduate School and Research Network for Artificial Intelligence*
‡*Anatomical Pathology Service, University Clinical Hospital of Valencia, Valencia, Spain*
**Instituto Universitario de Investigación en Tecnología Centrada en el Ser Humano, HUMAN-tech*,
Universitat Politècnica de València, Valencia, Spain
Email: algolsan@i3b.upv.es

***Abstract*—Prostate cancer is a common disease that affects men, and its diagnosis and prognosis rely on the Gleason scoring system. To automate this process, generative deep learning models can be used to synthesize histopathological tissue patches of non-cancerous and malignant patterns. This work proposes a conditional Progressive Growing GAN to generate synthetic samples by selecting the desired pattern. The model is trained using conditional information about the pattern, and minibatch standard deviation and pixel normalization are used to improve performance and stability. The synthetic samples are assessed using the Frechet Inception Distance (FID). Finally, the proposed framework was applied to the SICAPv2 dataset, and the results showed that data augmentation with our method improved the classification accuracy obtained on this dataset. This demonstrates the effectiveness of the proposed framework as a data augmentation method. Overall, this study provides a promising approach to address the issue of insufficient and unbalanced datasets in prostate cancer diagnosis and prognosis, which can improve the accuracy and reliability of clinical decision-making.**

***Index Terms*—Prostate cancer, Progressive Growing GAN, Conditional GAN, Histology image.**

I. Introduction

Prostate cancer, the second most common cancer in men with 1.4 million new cases in 2020, is diagnosed through prostate biopsy following clinical examination and lab tests. Digital pathology has gained prominence, involving scanning glass slides to create high-resolution whole slide images (WSI). These WSIs are divided into patches to enable Deep Learning (DL) models to analyze them, as shown in Figure 1. The success of artificial intelligence and machine learning solutions combined with this type of data promotes the development of computer vision applications to automate diagnoses, prognoses and disease predictions. Deep learning (DL) approaches have shown potential in many tasks in digital pathology, as mitosis detection [1], tissue classification [2], brain tumor classification [3] and glioma grading [4].

DL models strongly depend on having a well-balanced dataset to achieve optimal performance. However, many classification models struggle with imbalanced data, where some classes have limited samples. This leads to poor performance, particularly for minority classes. To address this, there is a need to develop a generative DL model that can synthesize WSI patches, helping overcome class imbalance and improve classification accuracy.

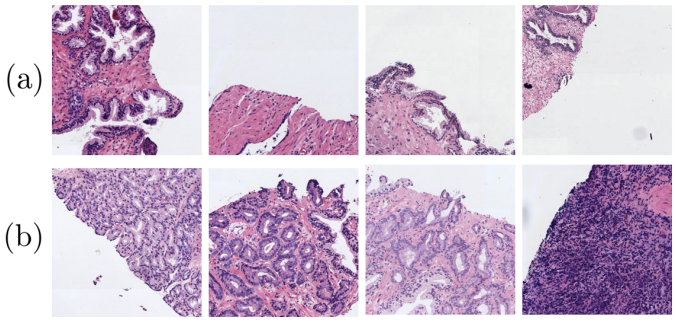

Fig. 1. Patches from SICAPv2 dataset. (a) Non-cancerous; (b) Malignant.

In the following lines, we summarize the main contributions of this work:

- A novel conditional progressive growing GAN framework able to synthesise patches containing non-cancerous or malignant patterns.
- We evaluate the quality of the synthetic data using the Frechet Inception Distance (FID), state-of-the-art metric evaluating image synthesis.
- The proposed method is validated as a data augmentation method.

II. Related Work

Generative models for image synthesis: Generative Adversarial Networks (GANs) were introduced by Goodfellow et al. [5] in 2014. GANs consist of two separate neural networks, the generator G and the discriminator D. The generator G takes a random noise vector z as input and generates synthetic data $G(z)$, while the discriminator D takes the generated data and real images $x \in p_{data}$ as inputs to classify them as real or synthetic. The conditional GAN was later developed by [6], which improved upon the original idea by conditioning the output to the input label. However, previous GAN approaches

have shown issues with stability and convergence. More complex frameworks were proposed to address this, such as Deep Convolutional GAN (DCGAN) [7]. X Chen et al. proposed an alternative approach in [8] named InfoGAN. Rather than using an unstructured noise vector z, the noise vector is decomposed into two parts, one with incompressible noise and the other with structured semantic features. The previously mentioned GAN methods create synthetic samples from random input noise but cannot do the reverse operation. To address this limitation, Donahue et al. proposed the BiGAN model [9]. Finally, Karras et al. [10] introduced the first approach using progressive growth training in their framework, Progressive Growing GAN (ProGAN), which became the central core of their paper.

Generative models for histological image synthesis: In this section, we present the main contributions of the literature on histological image synthesis. In [11], a Deep Convolutional GAN (DCGAN) was used to perform style transfer for synthesizing skin melanoma samples. Other literature examples of this architecture being used for histological image synthesis can be found in [12], [13], where conditional synthesis was performed on the class of cervical cancer samples. This framework was utilized as a data augmentation method and demonstrated an increase in accuracy for classification tasks.

In [14], a ProGAN approach was presented. This framework could synthesize histopathological images of brain tumours and is capable of generating more realistic images. They demonstrated that adding synthetic data to their dataset increases the accuracy of their classification models by approximately 5%. Another work using this architecture was presented by Teramoto et al. in [15]. It should be noted that the last two frameworks presented did not have the capacity for conditional synthesis. Although these papers prove their effectiveness as data augmentation methods, they do not evaluate the quality and realism of the synthetic samples generated. Furthermore, a novel conditional deep learning architecture based on StyleGAN [16], and BigGAN [17] was proposed in [18] for synthesizing colorectal and breast cancer samples.

III. Methods

The proposed framework was based on a conditional progressive growing GAN that was able to synthesize prostate histology non-cancerous and malignant patches. The workflow, which was composed of a generator θ^g and discriminator θ^d, is presented in Figure 2.

CGAN: This work utilizes a conditional generative adversarial network (CGAN) as its methodological core to produce synthetic prostate histology patches with cancerous and non-cancerous patterns. In this sense, the generator model aims to produce synthetic prostate histology non-cancerous and malignant patches. Formally, we denote the random input noise as $Z = \{z_1, ..., z_i, ..., z_C\}$, where z_i is the i-*th* instance obtained from a normal distribution $\mathcal{N}(\mu = 0, \sigma = 1)$ and C is the total number of generated samples. We note as N the total number of patches in the dataset, and for convenience, we give C the same value as N. Furthermore, the generator is provided with the desired cancerous pattern present in the synthetic sample. This pattern is indicated by $P = \{NC, M\}$, where NC represents the non-cancerous pattern, and M represents the malignant one. Accordingly, the synthetic patch generation process can be defined as follows:

$$G = f(Z, P; \theta^g) \tag{1}$$

where $G \in \mathbb{Z}^{m \times n \times 3}$ represents all the generated synthetic prostate histology patches and θ^g the model weights.

The objective of the discriminator model is to classify input patches as fake (0) or real (1). We define $X = \{x_1, ..., x_i, ..., x_N\}$, where x_i represents the i-*th* instance of real prostate histology patches. The input to the discriminator is $B = X \cup G$. The main objective is to predict $\hat{Y}$, which can be expressed as follows:

$$\hat{Y}_b = f(B, P; \theta^d) \tag{2}$$

where θ^d denotes the discriminator model weights.

A. Minibatch standard deviation

Given the inherent tendency of GANs to learn only a subset of the training set, we incorporated the "minibatch discrimination" technique as outlined in [19]. This approach involves computing the standard deviation for each feature in each spatial location across the minibatch, and incorporating this information into the discriminator. While the method does not require any learnable parameters or hyperparameters, we observed improved performance when the information was inserted towards the end of the discriminator.

B. Pixel Normalisation

Sometimes, the competition between the generator and the discriminator can cause the magnitudes of the generated values to become disproportionate. In order to address these issues, we applied feature vector normalization to ensure unit length in the generator after each convolution layer. To accomplish this, we utilized a modification of the "local response normalization" technique proposed in [20]. Equation 3 provides the expression for pixel normalization, where S denotes the number of feature maps, $(a^j_{x,y})$ represents the pixel being normalized, and ϵ denotes a small constant to avoid division by zero.

$$b_{x,y} = \sqrt{\frac{1}{S} \cdot \sum_{j=0}^{S-1} (a^j_{x,y})^2 + \epsilon} \tag{3}$$

C. Loss function

We used the Wasserstein GAN with Gradient Penalty (WGAN-GP) loss function [21] to optimize the performance of our proposal. Initially, the learning rate η, the ρ value which determines the maximum oscillation range of the gradients within [-ρ,ρ], and the batch size n were set. Then, a batch containing both real data $x^{(i)}$ and synthetic data $G(z^{(i)})$ was

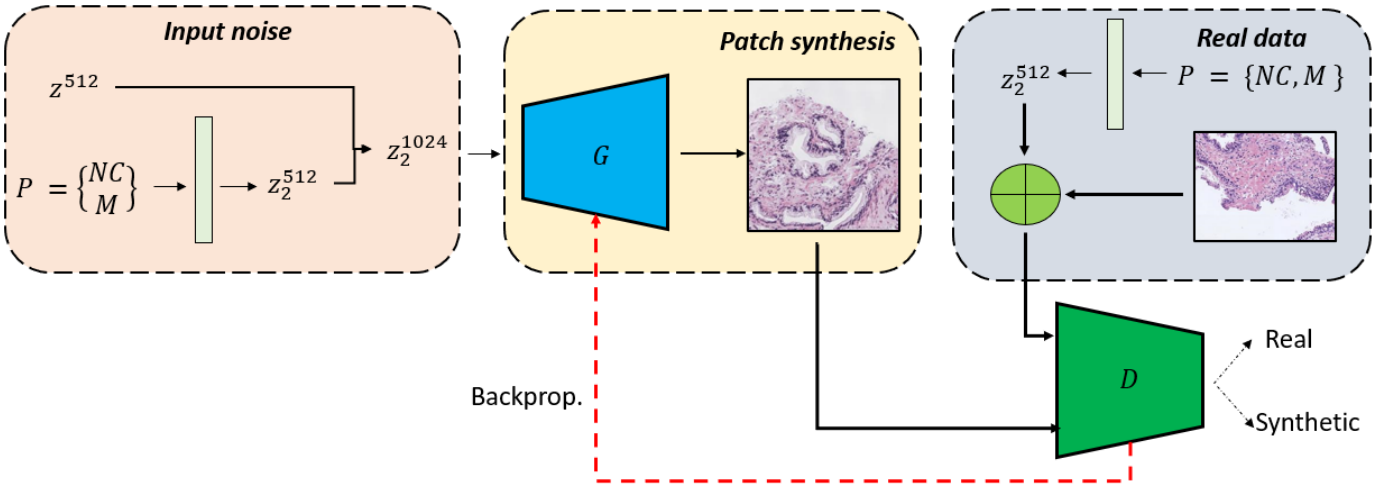

Fig. 2. **Method overview**. To train generator and discriminator models, we utilized histological prostate patches that exhibit both non-cancerous and malignant patterns. In order to effectively learn the morphology and properties of these patterns, we progressively increased the output resolution from 4^2 to 256^2. Both the generator and discriminator models received conditional information obtained by transforming input labels into an embedding vector of fixed size 512.

sampled. The loss function of the discriminator was formulated as Equation

$$loss_D = \frac{1}{n} \cdot \sum_{i=1}^{n} (D(x^{(i)})) - \frac{1}{n} \cdot \sum_{i=1}^{n} (D(G(z^{(i)}))) \quad (4)$$

After computing the gradients, the weights of the discriminator are updated. Then, we define the loss function of the generator as follows:

$$loss_G = \frac{1}{n} \cdot \sum_{i=1}^{n} (D(G(z^{(i)}))) \quad (5)$$

Our method: A progressive training approach was adopted to overcome the difficulty of synthesizing patches with a resolution of 256^2 in the CGAN framework. Increasing the network size gradually while increasing the resolution has several benefits. Starting from 4^2 resolution patches and adding new resolution layers progressively, the network learns high-level features of the image distribution first and then gradually increases the complexity of the details. This method allows the model to learn a simpler problem step by step rather than learning the entire information at once. The Progressive Growing GAN architecture was modified to incorporate conditional information on the cancerous pattern. During training, the generator and the discriminator receive this information without any additional term in the loss function.

IV. Experiments

A. Experimental setting

Datasets: The dataset utilized in this study was previously introduced in [22] and is publicly available at the SICAPv2 dataset website. This dataset represents the most comprehensive public collection of prostate Hematoxylin and Eosin (H&E) biopsies with patch-level annotations of different Gleason grades.

Implementation Details: The optimal combination of hyperparameters for the training process consisted of 100 epochs using the Adam optimizer with a β_1 value of 0 and a β_2 value of 0.99, a learning rate of 0.001, and the WGAN-GP loss function. The batch size was set to 64 for image resolutions ranging from 4^2 to 128^2 and 32 for 256^2. The optimal size for the generator input was determined to be 512.

Evaluation Metrics: To assess the effectiveness of the proposed model, we utilized the Frechet Inception Distance (FID) metric [23]. We assessed the difference between the extracted features from synthetic and real patches using the Inception V3 model, which was trained on the ImageNet dataset [24]. We denote the feature distribution of synthetic and real patches as $\mathcal{N}(\mu, C)$ and $\mathcal{N}(\mu_w, C_w)$, respectively. The FID expression is presented in Equation 6, where Tr corresponds to the trace of the matrix. Note that FID $\in [0, +\infty]$, being 0 the optimal value.

$$FID = \| \mu - \mu_w \|^2 + Tr(C + C_w - 2(C \cdot C_w)^{\frac{1}{2}}) \quad (6)$$

We computed a weighted average, assigning to the non-cancerous class a weight of 0.3656, while for the second class, we assigned a weight of 0.6344. We assigned the weights based on the varying number of samples present in the dataset for each class, which leads to different representations for each one.

B. Results

Quantitative evaluation: Table I presents the weighted FID results for all the frameworks considered in this study, including CGAN, ProGAN, and our method. Our results demonstrate that progressive training using the ProGAN framework offers a significant performance improvement compared to the CGAN approach. However, as noted earlier, the ProGAN architecture lacks conditional synthesis capacity. Our proposed conditional progressive framework provides the most

promising results for the FID metric evaluation. Notably, after applying the staining normalization process suggested in [25], our proposed method offers the best performance, with a 35.2% improvement in the weighted FID metric.

TABLE I
WEIGHTED FID RESULTS FOR ALL THE FRAMEWORKS CONSIDERED IN THIS WORK: CGAN, PROGAN AND OUR METHOD. ADDITIONALLY, WE PROVIDE THE METRICS OBTAINED AFTER THE STAIN NORMALISATION PROCESS.

	FID
CGAN	160.55
ProGAN	126.51
Ours	120.14
Ours + Stain Norm	**77.85**

In Table II, we present the results of the FID metric for each class (benign and malignant)

TABLE II
FID FOR CONDITIONAL EXPERIMENTS. NC FOR NON-CANCEROUS AND M FOR MALIGNANT PATTERNS.

Test		
	NC	M
CGAN	198.45	159.71
Ours	92.14	132.35
Ours + Stain Norm	**88.85**	**73.06**

The results demonstrate that the CGAN architecture underperforms compared to the proposed model, as the network's complexity is insufficient to learn the features of the input images.

Qualitative evaluation: To qualitatively evaluate the proposed method, we generated patches using the CGAN, ProGAN, and our framework. The synthesized patches using the CGAN approach are displayed in Figure 3. While this framework successfully captures colour properties and spatial distribution, it cannot learn morphology and distinct features for non-cancerous and malignant patterns. As a result, it cannot produce complex structures such as glands or nuclei.

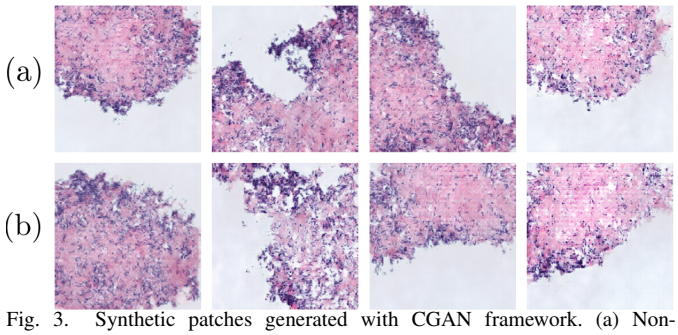

Fig. 3. Synthetic patches generated with CGAN framework. (a) Non-cancerous; (b) Malignant

Figure 4 displays several patches produced using the ProGAN framework. While this approach can generate samples that accurately reflect the morphology and distribution of real patches, it cannot distinguish between non-cancerous and malignant patches.

The results of our framework after the stain normalization post-processing are presented in Figure 5. After introducing

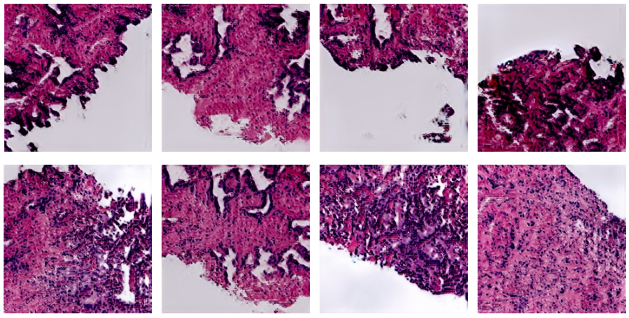
Fig. 4. Synthetic patches generated with the ProGAN framework.

conditional information, synthetic samples exhibited greater homogeneity and coherence with real data as non-cancerous images display well-differentiated glands, while malignant patches demonstrate an increase in gland density and the emergence of irregularities. This indicates that the proposed model can identify and represent each pattern's intrinsic characteristics.

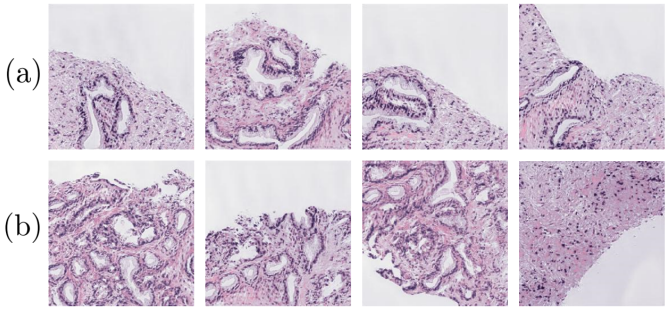

Fig. 5. Synthetic patches generated with our framework. (a) Non-cancerous; (b) Malignant.

Data augmentation strategy validation: This section shows the validation of the proposed method as a data augmentation strategy. For this purpose, we compare the classification model used in [22] trained with SICAPv2 and SICAPv2 augmented with the proposed model. Specifically, the non-cancerous class was augmented by 50%. The results obtained are shown in Table III. The proposed model improves the classification model performance. This fact supports the effectiveness and validity of the proposed work as a patch synthesis method and data augmentation strategy.

TABLE III
RESULTS FOR THE TEST SET FOR THE MODEL PROPOSED IN [22] WITH THE ORIGINAL SICAPV2 DATASET (S) AND UPSAMPLING WITH OUR PROPOSED DATA AUGMENTATION METHOD (S+P). THE METRICS PRESENTED ARE PRECISION, F1-SCORE, COMPUTED PER CLASS, AND GLOBAL ACCURACY.

	Precision		F1-S		ACC	
	S	S+P	S	S+P	S	S+P
NC	0.6506	**0.7067**	0.7073	**0.7233**	-	-
C	**0.8930**	0.8846	0.8542	**0.8752**	-	-
Avg	0.8194	**0.8306**	0.8096	**0.8291**	0.8054	**0.8280**

V. CONCLUSIONS

This research proposes a novel framework for generating high-quality synthetic prostate tissue patches that accurately represent both non-cancerous and malignant tissue types. Our

proposed framework used a conditional Progressive Growing GAN and achieved a significantly better weighted FID metric of 77.85 than the CGAN and ProGAN models, which obtained 160.55 and 120.14, respectively. Furthermore, we trained a classification model using the synthetic images generated by our framework and the SICAPv2 dataset. Our proposed method achieved an increased classification accuracy of 2.3%. These results confirm the effectiveness of our ProstatePatchGAN-generated images for accurately representing both non-cancerous and malignant prostate tissue types and their potential application in various clinical settings.

ACKNOWLEDGEMENT

We gratefully acknowledge the support from the Generalitat Valenciana (GVA) with the donation of the DGX A100 used for this work, action co-financed by the European Union through the Operational Program of the European Regional Development Fund of the Comunitat Valenciana 2014-2020 (IDIFEDER/2020/030). This work has received funding from Horizon 2020, the European Union's Framework Programme for Research and Innovation, under grant agreement No. 860627 (CLARIFY), the Spanish Ministry of Economy and Competitiveness through project PID2019-105142RB-C21 (AI4SKIN) and GVA through projects PROMETEO/2019/109 and INNEST/2021/321 (SAMUEL). The work of Adrián Colomer has been supported by the ValgrAI – Valencian Graduate School and Research Network for Artificial Intelligence & Generalitat Valenciana and Universitat Politècnica de València (PAID-PD-22). Rocío del Amor has been supported by the Spanish Government under FPU Grant (FPU20/05263).

REFERENCES

[1] M. Veta, P. J. Van Diest, S. M. Willems, H. Wang, A. Madabhushi, A. Cruz-Roa, F. Gonzalez, A. B. Larsen, J. S. Vestergaard, A. B. Dahl *et al.*, "Assessment of algorithms for mitosis detection in breast cancer histopathology images," *Medical image analysis*, vol. 20, no. 1, pp. 237–248, 2015.

[2] A. Cruz-Roa, A. Basavanhally, F. González, H. Gilmore, M. Feldman, S. Ganesan, N. Shih, J. Tomaszewski, and A. Madabhushi, "Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks," in *Medical Imaging 2014: Digital Pathology*, vol. 9041. SPIE, 2014, p. 904103.

[3] J. Xu, L. Xiang, Q. Liu, H. Gilmore, J. Wu, J. Tang, and A. Madabhushi, "Stacked sparse autoencoder (ssae) for nuclei detection on breast cancer histopathology images," *IEEE transactions on medical imaging*, vol. 35, no. 1, pp. 119–130, 2015.

[4] M. G. Ertosun and D. L. Rubin, "Automated grading of gliomas using deep learning in digital pathology images: a modular approach with ensemble of convolutional neural networks," in *AMIA annual symposium proceedings*, vol. 2015. American Medical Informatics Association, 2015, p. 1899.

[5] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, 2014.

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

[7] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.

[8] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel, "Infogan: Interpretable representation learning by information maximizing generative adversarial nets," *Advances in neural information processing systems*, vol. 29, 2016.

[9] J. Donahue, P. Krähenbühl, and T. Darrell, "Adversarial feature learning," *arXiv preprint arXiv:1605.09782*, 2016.

[10] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of gans for improved quality, stability, and variation," *arXiv preprint arXiv:1710.10196*, 2017.

[11] A. Mikołajczyk and M. Grochowski, "Data augmentation for improving deep learning in image classification problem," in *2018 international interdisciplinary PhD workshop (IIPhDW)*. IEEE, 2018, pp. 117–122.

[12] Y. Xue, Q. Zhou, J. Ye, L. R. Long, S. Antani, C. Cornwell, Z. Xue, and X. Huang, "Synthetic augmentation and feature-based filtering for improved cervical histopathology image classification," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2019, pp. 387–396.

[13] J. Krause, H. I. Grabsch, M. Kloor, M. Jendrusch, A. Echle, R. D. Buelow, P. Boor, T. Luedde, T. J. Brinker, C. Trautwein *et al.*, "Deep learning detects genetic alterations in cancer histology generated by adversarial networks," *The Journal of pathology*, vol. 254, no. 1, pp. 70–79, 2021.

[14] S. Liu, Z. Shah, A. Sav, C. Russo, S. Berkovsky, Y. Qian, E. Coiera, and A. Di Ieva, "Isocitrate dehydrogenase (idh) status prediction in histopathology images of gliomas using deep learning," *Scientific reports*, vol. 10, no. 1, pp. 1–11, 2020.

[15] A. Teramoto, T. Tsukamoto, A. Yamada, Y. Kiriyama, K. Imaizumi, K. Saito, and H. Fujita, "Deep learning approach to classification of lung cytological images: Two-step training using actual and synthesized images by progressive growing of generative adversarial networks," *PloS one*, vol. 15, no. 3, p. e0229951, 2020.

[16] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 4401–4410.

[17] A. Brock, J. Donahue, and K. Simonyan, "Large scale gan training for high fidelity natural image synthesis," *arXiv preprint arXiv:1809.11096*, 2018.

[18] A. C. Quiros, R. Murray-Smith, and K. Yuan, "Pathologygan: Learning deep representations of cancer tissue," *arXiv preprint arXiv:1907.02644*, 2019.

[19] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen, and X. Chen, "Improved techniques for training gans," in *Advances in Neural Information Processing Systems*, D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, Eds., vol. 29. Curran Associates, Inc., 2016. [Online]. Available: https://proceedings.neurips.cc/paper/2016/file/8a3363abe792db2d8761d6403605aeb7-Paper.pdf

[20] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," *arXiv preprint arXiv:1207.0580*, 2012.

[21] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein gans," *Advances in neural information processing systems*, vol. 30, 2017.

[22] J. Silva-Rodríguez, A. Colomer, M. A. Sales, R. Molina, and V. Naranjo, "Going deeper through the gleason scoring scale: An automatic end-to-end system for histology prostate grading and cribriform pattern detection," *Computer Methods and Programs in Biomedicine*, vol. 195, p. 105637, 2020.

[23] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017. [Online]. Available: https://proceedings.neurips.cc/paper/2017/file/8a1d694707eb0fefe65871369074926d-Paper.pdf

[24] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[25] M. Macenko, M. Niethammer, J. S. Marron, D. Borland, J. T. Woosley, X. Guan, C. Schmitt, and N. E. Thomas, "A method for normalizing histology slides for quantitative analysis," in *2009 IEEE international symposium on biomedical imaging: from nano to macro*. IEEE, 2009, pp. 1107–1110.