




Article

Synthetic Genitourinary Image Synthesis via Generative Adversarial Networks: Enhancing Artificial Intelligence Diagnostic Precision

Derek J. Van Booven ¹ , Cheng-Bang Chen ² , Sheetal Malpani ³, Yasamin Mirzabeigi ³, Maral Mohammadi ⁴, Yujie Wang ², Oleksander N. Kryvenko ³, Sanoj Punnen ⁵ and Himanshu Arora ^{1,4,5,6,*} 

- ¹ John P Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, FL 33136, USA; dvanbooven@med.miami.edu
- ² Department of Industrial and Systems Engineering, University of Miami, Coral Gables, FL 33146, USA; cxc1920@miami.edu (C.-B.C.); yxw509@miami.edu (Y.W.)
- ³ Department of Pathology, Miller School of Medicine, University of Miami, Miami, FL 33136, USA; sheetal.malpani@jhsmiami.org (S.M.); yasamin.mirzabeigi@jhsmiami.org (Y.M.); o.kryvenko@med.miami.edu (O.N.K.)
- ⁴ Department of Pathology, University of Debrecen in Hungary, 4032 Debrecen, Hungary; smaral.kmohammadi@gmail.com
- ⁵ Desai & Sethi Institute of Urology, Miller School of Medicine, University of Miami, Miami, FL 33136, USA; s.punnen@med.miami.edu
- ⁶ The Interdisciplinary Stem Cell Institute, Miller School of Medicine, University of Miami, Miami, FL 33136, USA
- * Correspondence: hxa287@med.miami.edu

Abstract: Introduction: In the realm of computational pathology, the scarcity and restricted diversity of genitourinary (GU) tissue datasets pose significant challenges for training robust diagnostic models. This study explores the potential of Generative Adversarial Networks (GANs) to mitigate these limitations by generating high-quality synthetic images of rare or underrepresented GU tissues. We hypothesized that augmenting the training data of computational pathology models with these GAN-generated images, validated through pathologist evaluation and quantitative similarity measures, would significantly enhance model performance in tasks such as tissue classification, segmentation, and disease detection. Methods: To test this hypothesis, we employed a GAN model to produce synthetic images of eight different GU tissues. The quality of these images was rigorously assessed using a Relative Inception Score (RIS) of 1.27 ± 0.15 and a Fréchet Inception Distance (FID) that stabilized at 120, metrics that reflect the visual and statistical fidelity of the generated images to real histopathological images. Additionally, the synthetic images received an 80% approval rating from board-certified pathologists, further validating their realism and diagnostic utility. We used an alternative Spatial Heterogeneous Recurrence Quantification Analysis (SHRQA) to assess the quality of prostate tissue. This allowed us to make a comparison between original and synthetic data in the context of features, which were further validated by the pathologist's evaluation. Future work will focus on implementing a deep learning model to evaluate the performance of the augmented datasets in tasks such as tissue classification, segmentation, and disease detection. This will provide a more comprehensive understanding of the utility of GAN-generated synthetic images in enhancing computational pathology workflows. Results: This study not only confirms the feasibility of using GANs for data augmentation in medical image analysis but also highlights the critical role of synthetic data in addressing the challenges of dataset scarcity and imbalance. Conclusions: Future work will focus on refining the generative models to produce even more diverse and complex tissue representations, potentially transforming the landscape of medical diagnostics with AI-driven solutions.

Keywords: GAN; prostate cancer; machine learning; pathology; diagnostics; prognostics; digital imaging



Citation: Van Booven, D.J.; Chen, C.-B.; Malpani, S.; Mirzabeigi, Y.; Mohammadi, M.; Wang, Y.; Kryvenko, O.N.; Punnen, S.; Arora, H. Synthetic Genitourinary Image Synthesis via Generative Adversarial Networks: Enhancing Artificial Intelligence Diagnostic Precision. *J. Pers. Med.* **2024**, *14*, 703. <https://doi.org/10.3390/jpm14070703>

Academic Editors: Nicola Maggialetti and Antonio Rosario Pisani

Received: 23 May 2024
Revised: 20 June 2024
Accepted: 24 June 2024
Published: 30 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Artificial intelligence (AI) has revolutionized the medical imaging landscape, offering innovative applications that aid diagnosis and treatment. In diagnostic radiology, deep learning algorithms, such as those developed by Zebra Medical Vision and Aidoc, analyze X-rays and CT scans to detect a range of conditions, providing faster and sometimes more accurate readings than traditional methods [1–6]. In pathology, companies like PathAI use AI to identify patterns in tissue samples, improving cancer diagnoses [7–11]. Similarly, in ophthalmology, tools like IDx-DR for diabetic retinopathy screening autonomously assess retinal images to identify early signs of disease [4,12,13]. In cardiology, AI-powered software like that from Arterys evaluates cardiac MRI and CT scans to provide detailed insights into heart structure and function, aiding in the diagnosis of cardiovascular diseases [14]. Despite these advancements, AI applications are not without concerns. The ‘black box’ nature of many AI systems, where the decision-making process is not transparent, poses challenges to clinical validation and trust. Data privacy and security are also significant issues, as AI models require large datasets for training, potentially exposing sensitive patient information if data are breached or improperly accessed [3,15,16]. Real-world breaches, such as the Anthem Inc. and UCLA Health System breaches, underscore these vulnerabilities. Additionally, algorithmic bias and errors in AI systems necessitate meticulous dataset curation and algorithm training to ensure equitable and accurate medical services.

The adoption of Generative Adversarial Networks (GANs) to generate synthetic data presents a promising solution to these challenges [17–19]. GANs can create realistic medical images, reducing the need to use them and potentially exposing sensitive patient data [17–19]. This method of data augmentation enriches the dataset required for robust AI diagnostic tools and serves as a critical buffer for maintaining patient privacy. In the current study, we utilized GANs for synthetic image generation in genitourinary pathology, highlighting their potential in this context. The GANs underwent rigorous quality control processes, including validation by board-certified pathologists and quantification of image fidelity through Relative Inception Scores and Fréchet Inception Distance, demonstrating high-quality synthetic image production. These images were indistinguishable from real data in many instances, enabling their use in AI diagnostics without the risk associated with actual patient data. By incorporating synthetic data generation via GANs, the healthcare industry can safeguard sensitive patient information, addressing one of the most significant cybersecurity concerns of our time. As we continue to navigate the complexities introduced by AI in healthcare, the role of GANs in cybersecurity becomes increasingly pertinent. They represent a promising path forward, integrating AI into medical practice in a secure, ethical, and conducive manner to patient trust and safety.

2. Methods

2.1. Cohorts Used

We harnessed eight genitourinary tissue types—bladder, cervix, kidney, ovary, prostate, testis, uterus, and vagina—obtained from the Genotype-Tissue Expression (GTEx) database, a comprehensive resource that provides open access to tissue expression data. Additionally, histology images from the cancer genome atlas (TCGA) of 500 individuals representing the adenocarcinoma stage were considered controls. Segmentation was performed using PyHIST, a Python-based histological tool, which processed the images into discrete squares of 64, 128, and 256 pixels. Each segment was curated to contain a minimum of 75% tissue content, a criterion set to minimize regional bias and preserve the representativeness of the histological features.

2.2. Development and Evaluation of a Conditional Generative Adversarial Network

A preliminary conditional Generative Adversarial Network (cGAN) was designed and implemented to assess the performance accuracy of various GAN architectures. The cGAN was developed utilizing Python 3.7.3 and the Tensorflow Keras 2.7.0 package. The generator component of the cGAN comprises three input layers and a single output layer. In parallel,

the discriminator component is configured with analogous input, hidden, and output layers. The cGAN's total parameter count was 7.5 million for each of the evaluated image patterns.

2.3. Implementation and Adaptation of StyleGAN for Tissue Image Analysis

StyleGAN, a progressive generative adversarial network architecture engineered using Python 3.9 and the TensorFlow framework, leveraging the conditional GAN architecture, was used to guide the image synthesis process. This structure allowed the GAN to generate images conditioned on specific tissue types, facilitating targeted image generation. To automate and streamline the process, we employed a bash script tailored for each tissue type that orchestrated the importation of images, their conversion to an RGB color space, and the compilation of these images into a NumPy array. These arrays were then stored as .npy files, ensuring reproducibility and consistency across GAN runs. During the synthetic image generation phase, the generator component of the GAN introduced random noise variables, which were assessed by the discriminator component. This interplay continued iteratively, with the loss graph monitored meticulously until stabilization was observed—a signal to cease the discriminator's assessment and crystallize the synthetic image output. On average, the GAN system required 2.5 h per run, yielding a thousand synthetic images per tissue type. The loss functions—mathematical functions quantifying the error between the generated images and the actual images—were pivotal in guiding the GAN's training. Monitoring these allowed us to fine-tune the GAN's parameters, with an observed convergence of loss functions around the 182nd epoch. This convergence was deemed the optimal stopping point, indicative of the GAN's ability to generate images with minimal discrepancy from the target dataset. The term "loss" here referred specifically to the number of images that were not deemed accurate enough by the GAN, thereby being 'dismissed' during the iterative training process.

2.4. StyleGAN

The StyleGAN implementation was obtained from the NVIDIA Labs Github (<https://github.com/NVLabs/stylegan> accessed on 24 April 2022) and was run with Python 3.9. TensorFlow version 1.12.0 and CUDA version 10.2 were used. The training images were imported into a TFRecords dataset object and stored as a .tfrecords file. Initial training was performed on a V100 Tesla GPU, and it took an average of 2.5 days to complete the first round of training. This trained model was then used as the basis for generating new tissue images. The architecture of StyleGAN was kept exactly as is from the NVIDIA download (<https://arxiv.org/abs/1812.04948> accessed on 24 April 2022). The only parameter that was changed to generate sufficient images was the resolution factor. This resolution factor was set to 256 in order to output the images at a quality that could be inspected manually.

2.5. Quantification Model

To characterize technical and structural variations between synthetic and real images, we utilized Spatial Heterogeneous Recurrence Quantification Analysis (SHRQA), a robust technique capable of measuring complex microstructures based on spatial patterns [20–22]. The SHRQA process, as shown in Supplementary Figure S4, involves six key steps. It begins with the 2D-Discrete Wavelet Transform (2D-DWT) using the Haar wavelet to reveal patterns not visible in the original image [21–26]. Then, each image is transformed into an attribute vector via the space-filling curve (SFC), which importantly preserves the spatial proximity between pixels in the image within the vector. This step is crucial for analyzing the image's geometric recurrence in vector form. A trajectory is formed in state space by projecting this attribute vector, highlighting the image's geometric structure. Through quadtree segmentation, the state space is divided into unique subregions to discern spatial transition patterns [27,28]. An Iterated Function System projection is then applied, converting each attribute vector into a fractal plot that represents recurrence within the fractal topology. Finally, these fractal structures are quantified to illuminate the intricate geometric properties of the image, providing a detailed profile.

2.6. Statistical Calculations

FID was implemented in custom scripts developed in-house. The FID model was pre-trained using Inception V3 weights for transfer learning. In-house code was centered around the FID model and inserted into the StyleGAN to be run during each iteration. Stats were reported at intervals of 1000 and graphed with in-house Python scripts. Reported FID figures represent an inverse relationship between the images; thus, the lower our FID figure, the more similar the images.

PCA analysis was performed by first transforming the images into numerical arrays. Images were separated into normal and synthetic batches. The intensity was calculated (using the R package `imgpalr` and `magick`) as the average of the color of the entire image while keeping the matrix framework (i.e., positional arguments were retained). PCA was conducted using the general `prcomp` function in R, and the plotted results were displayed in `ggplot2`.

2.7. Data Sharing

De-identified participant data will be made available when all primary and secondary endpoints have been met. Any requests for trial data and Supporting Material (data dictionary, protocol, and statistical analysis plan) will be reviewed by the trial management group in the first instance. Only requests that have a methodologically sound proposal and whose proposed use of the data has been approved by the independent trial steering committee will be considered. Proposals should be directed to the corresponding author in the first instance; to gain access, data requestors will need to sign a data access agreement.

3. Results

GAN Model Selection: To evaluate the performance of various GAN architectures and select the most appropriate one, digital histology images were downloaded from the Genotype-Tissue Expression (GTEx) database for the prostate. In 9091, 256×256 image patches were extracted from 599 individuals and divided into training cohorts. Each training cohort was subjected to cGAN, StyleGAN, and dcGAN architectures [29–31]. A total of 200 randomly selected synthetic images generated by each GAN were fed into a generic CNN for classification. The cGAN achieved an accuracy of 36% (72 images were classified correctly), while the StyleGAN and dcGAN demonstrated accuracies of 62.5% (125 correctly classified) and 60% (120 correctly classified), respectively. Although StyleGAN and dcGAN exhibited similar accuracies, the quality of output was more extensive for StyleGAN, which is particularly important considering the less heterogeneity that exists in standard/non-cancer tissue image types.

Image synthesis: Once the GAN was selected, the GTEx database was used to extract digital histology images from eight genitourinary tissue types; 129 images were available for the bladder, 81 for the cervix, 599 for the kidney, 252 for the ovary, 599 for the prostate, 588 for the testis, 234 for the uterus, and 272 for the vagina. Several factors, such as staining protocols, tissue quality, section thickness, tissue folding, and the amount of tissue on the slide, could negatively impact the efficiency of the GAN model in generating high-quality data [32]. To account for this, we conducted pre-processing normalization of the images. Specifically, we selected all the images from all tissue types and evaluated their color distribution by calculating the mean value of RGB colors and normalizing them. Images with an RGB mean intensity value two standard deviations away from the total mean value of all samples were identified as outliers and removed from the dataset. In total, 21 images were discarded due to being outliers. Overall, our pre-processing steps helped to reduce the variability in tissue biopsy images and ensure a more consistent training dataset for the StyleGAN model. Post-processing, these images were used to train the StyleGAN model. The network generator created a total of 200 random synthetic images for each of the tissue types. The patch size of each of these images was set at 5000×5000 to allow sufficient quality for the pathologist's evaluation. These image patches were analyzed using the Adam optimization algorithm. This process helped us find the best iteration value for our model, which was 15,000 iterations. Figure 1A summarizes the steps in the processing and

generation of synthetic images, and Figure 1B and Supplementary Figures S1–S8 showcase the examples of synthetic images generated from eight GU tissues.

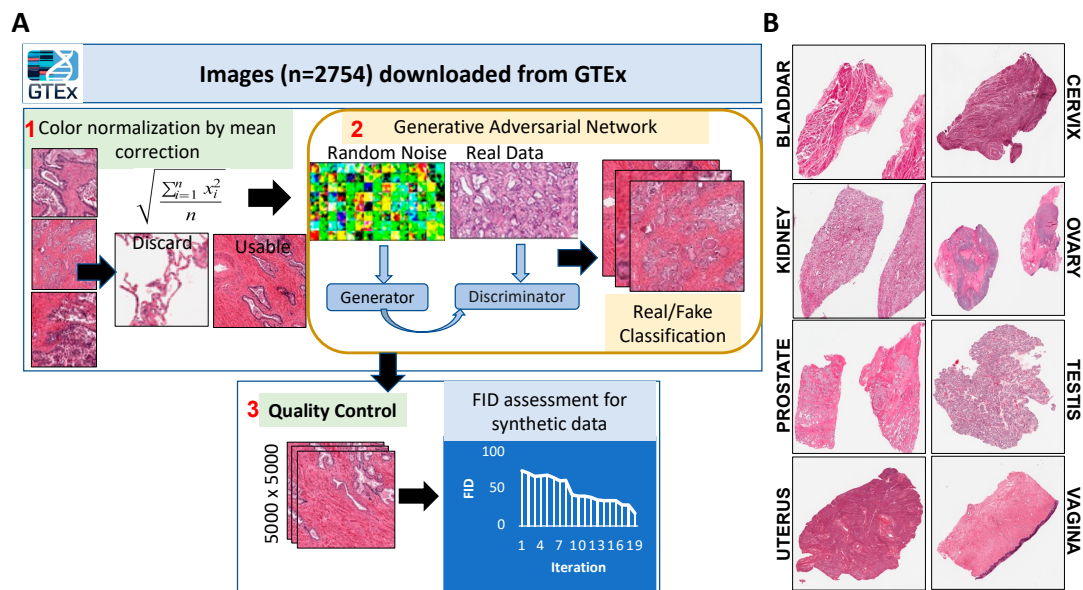


Figure 1. (A) GAN Workflow. Images were normalized, run through the GAN, and then put through QC. (B) Synthetic images were generated for each GU tissue type, respectively.

Next, we applied standard machine learning metrics to evaluate the synthetic images. The Relative Inception Score (RIS) was a primary metric, measuring the clarity and variety of the generated images. A high RIS of 17.2 with a remarkably low standard deviation of 0.15 across different tissue types demonstrated the synthetic images' consistent quality. Furthermore, the Fréchet Inception Distance (FID), a crucial index for GAN performance, was used to compare the distribution of generated images with real images. An FID score that stabilized at 120 indicated that the synthetic images closely mirrored the distribution of the real tissue images, solidifying the efficacy of our GAN model.

Quality Control Through Expert Evaluation: The synthetic images underwent a rigorous review process for quality control. A subset of synthetic prostate images were subjected to detailed visual inspection, focusing on aspects such as sharpness and resolution. This scrutiny was critical to ensuring that the generated images met the high standards required for clinical use. For this, two certified pathologists conducted an independent review of the synthetic image cohorts, where they were provided with a randomized pool of 20 images per tissue type, consisting of a mixture of 15 synthetic and five real images, totaling 160 images. The pathologists were tasked with evaluating the quality of the images and highlighting the concerns they may have for each tissue type. Table 1 summarizes the quality evaluation outcomes of pathology evaluation for all eight tissue types. Supplementary Figures S1–S4 show the 20 images per tissue type shared with the pathologists. Results highlighted an 80% approval rate, signifying a robust endorsement of the synthetic images' clinical utility.

Geometric Analysis of Image Characteristics: To delve deeper into the geometric properties of the synthetic images, we employed Spatial Heterogeneous Recurrence Quantification Analysis (SHRQA). This involved initial image pre-processing, including grayscale conversion, noise reduction, contrast enhancement, and normalization, to reduce the unrelated noise and amplify underlying patterns within the images. Subsequently, each image was transformed into an attribute vector through the application of a Hilbert space-filling curve, a technique that preserves the spatial proximity relationships of the image pixels in a one-dimensional vector. This vectorization facilitated a detailed analysis of the geometric recurrence and structural intricacies within the images. By applying the Iterated Function System projection, we were able to identify and quantify recurrent fractal structures, thereby providing a robust profile of the images' geometric fidelity (Figure 2).

Table 1. Outcomes of the image quality assessment of 20 images per tissue type by two pathologists. The pathologists were subjected to a query of “QC Pass” (P) or “QC Fail” (F) to highlight any concerns they had during the evaluation process.

Image Quality Assessment By Pathologist																				
	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15	I16	I17	I18	I19	I20
BLADDAR	P	P	F	P	P	P	P	P	F	P	P	P	P	P	P	P	F	P	P	P
CERVIX	P	P	F	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P
KIDNEY	F	P	P	P	F	P	P	P	P	F	F	F	P	F	P	P	P	P	P	P
OVARY	P	F	F	P	P	F	P	F	P	P	F	F	F	P	P	P	F	F	F	F
PROSTATE	P	P	P	P	P	P	P	P	F	P	P	F	P	P	P	P	P	P	F	P
TESTIS	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P
UTERUS	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P
VAGINA	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P

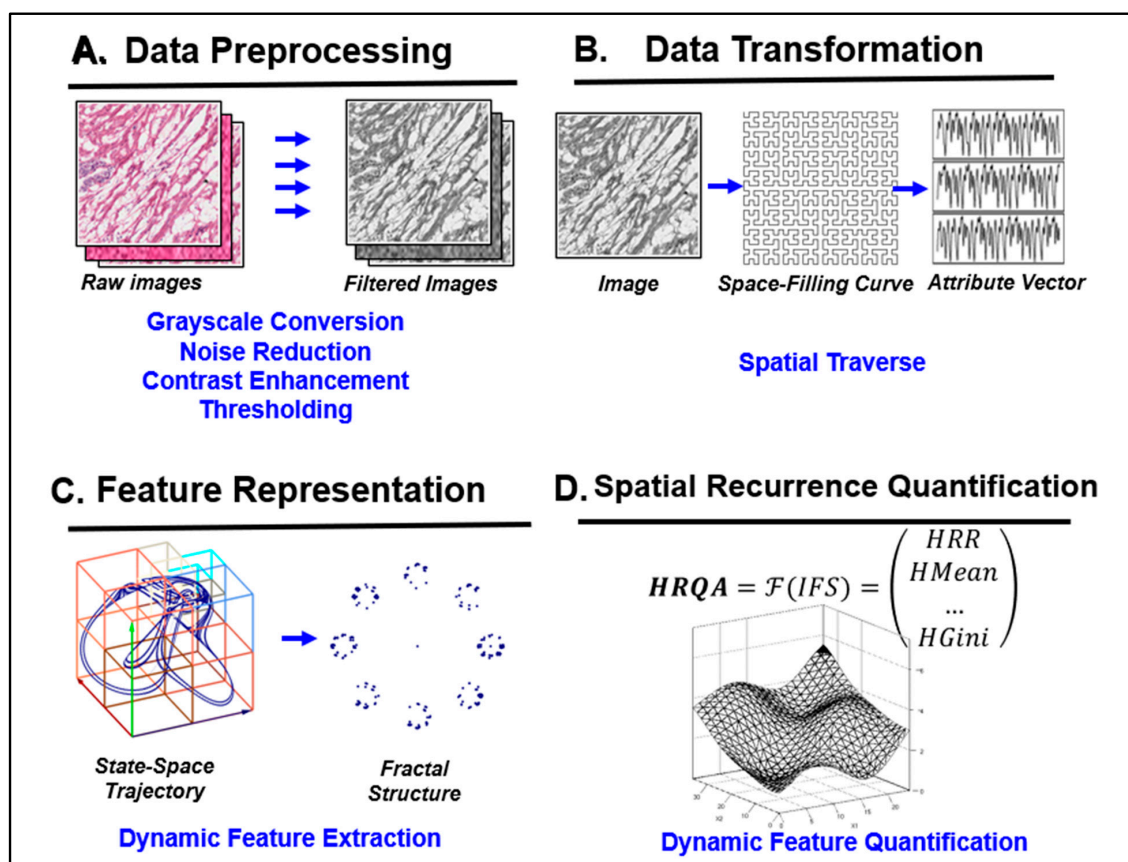


Figure 2. The framework of the Spatial Heterogeneous Recurrence Quantification Analysis (SHRQA). Initially, each image undergoes standard image pre-processing, including grayscale conversion, noise reduction, contrast enhancement, and thresholding. This amplifies intricate patterns and minimizes environmental noise. Subsequently, a space-filling curve transforms each image into an attribute vector, preserving the majority of its proximity information. Through state-space construction, pixel color/intensity transitions form a trajectory in the state space. These transitions are then projected into an Iterated Function System (IFS) to capture complex dynamic properties. The image’s nuanced geometric properties are then mathematically described using recurrence quantification analysis. Ultimately, the extracted spatial recurrence characteristics can be employed to profile images.

The SHRQA method was first applied to examine the spatial recurrence properties of real and synthetic image patches across test tissue data, which was prostate in this specific scenario. Our sample set included an equal number of patches from real and synthetic sources, with a balanced representation of each phenotype. To add an extra layer

of validation, we downloaded histology images from the cancer genome atlas (TCGA) from individuals representing the adenocarcinoma stage. These images, representing different stages of cancer progression (represented by Gleason grade), were randomized.

On all the image types (normal original (NO), normal synthetic (NS), and cancer original (CO)), segmentation was performed using PyHIST, a Python-based histological tool that processed the images into discrete squares of 256 pixels. Each segment was curated to contain a minimum of 90% tissue content, a criterion set to minimize regional bias and preserve the representativeness of the histological features. We analyzed 2000 image patches, each 256×256 pixels, evenly split between real and synthetic. SHRQA quantitatively outlined each patch's microstructures. From an initial extraction of 112 spatial recurrence features per patch, LASSO selected 102 features that were significant to the Gleason pattern. Hotelling's T-squared test, a multivariate extension of the two-sample *t*-test, compared the spatial recurrence attributes of real versus synthetic patches. The resulting *p*-values of 0.4039 signified no significant differences in spatial recurrence properties between NO and NS, but a significant difference was observed between NO and CO ($p = 1.353 \times 10^{-7}$) and NS and CO ($p = 1.759 \times 10^{-7}$), as confirmed by the T-squared tests' *p*-values for each Gleason pattern.

We also employed PCA on the spatial recurrence properties [33–39], visualized using radar charts, revealing that the top five principal components capture 90% of the variability. This allowed us to map the distributions of spatial properties for real and synthetic images across phenotypes, as depicted in Figure 3. Notably, while distributions aligned closely between real and synthetic images (NO–NS), significant differences were evident between NO–CO and NS–CO. These findings across image sections validate the model's efficiency in capturing the geometric intricacies consistent with real images.

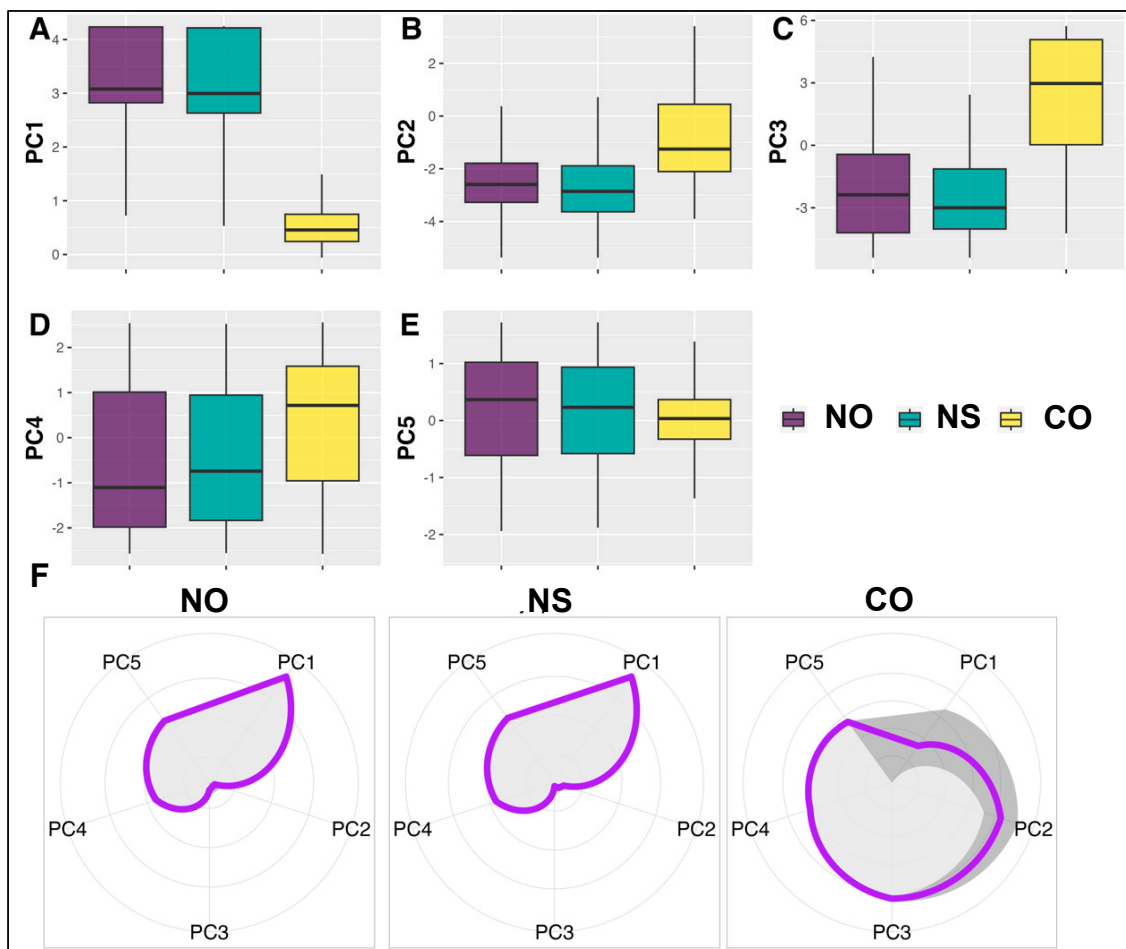


Figure 3. (A–E) The comparison of spatial recurrence properties between normal original (NO), normal synthetic (NS), and cancer original (CO) on the first five PCs (containing >95% of data variability). The

distributions of these five PCs are similar between real and synthetic. (F) The distributions of spatial recurrence properties (in the first five Principal Components (PCs), which contain > 95% of data variability) underlying different patterns for both real and synthetic patches. Note that the purple lines indicate the mean values of each feature, and the gray area shows the 95% confidence interval. Our results indicate that while the distributions of spatial properties are closely aligned between NO and NS, they markedly differ when comparing CO.

4. Discussion and Conclusions

The application of Generative Adversarial Networks (GANs) in producing synthetic medical images, as demonstrated by our research, has significant implications for healthcare. By generating synthetic images that are virtually indistinguishable from real histological samples, GANs provide a powerful tool for training AI systems without the risk of exposing sensitive patient information. This is a key consideration given the notable cybersecurity incidents in recent years, such as the Anthem Inc. and UCLA Health System breaches, which exposed the data of millions. Our study's success in generating high-quality synthetic genitourinary images serves as a proof of concept for the broader application of GANs in medical imaging. By employing this technology, healthcare providers can enhance the robustness of AI diagnostic tools while maintaining stringent data security. For instance, rather than relying on vast databases of patient images, which pose a potential risk if compromised, medical AI applications can be trained using synthetic datasets that carry no privacy concerns.

The practicality of synthetic images generated by GANs is further supported by their performance in standard machine learning metrics and approval by expert pathologists. This dual validation underscores the potential of GANs not only in generating training data but also in providing a buffer against data breaches. As AI continues to permeate the medical field, the ability to create diverse, high-fidelity datasets through GANs becomes increasingly valuable, offering a safeguard against the risks associated with the collection and storage of large-scale patient data.

Looking ahead, the expansion of this methodology to other tissue types and medical conditions could revolutionize the field of medical diagnostics. For example, AI models trained on GAN-generated images could support the early detection of rare diseases without requiring access to potentially sensitive real-world data. Similarly, the generation of synthetic images for rare pathologies could aid in developing diagnostic models where real data are scarce or difficult to obtain due to privacy concerns. However, there are limitations that need to be addressed for the appropriate application of the generated data. For example, to ease the pathology review process, the synthetic images were generated with a large patch size of 5000*5000 pixels. It served the purpose, but on the downside, we had to use the training data with a similar patch size of 5000*5000 pixels, which limited the amount of training data. Secondly, we utilized images representing non-diseased conditions, which had more or less a uniform distribution of features and structures compared to cancer images. This limited the GAN model's ability to generate a vast number of unique synthetic images. Consequently, to perform quantification, we divided the synthetic images into small patch sizes of 256×256 pixels before subjecting them to SHQRA models. This allowed us to perform the feature comparison and quantification successfully. To avoid these issues, an increase in the size of training data and starting with a small patch size, which can be localized within the tissue section, will immensely enhance the efficiency of the model while allowing the evaluation by the pathologists. Another limitation is the time StyleGAN takes to generate the synthetic data, which can limit its widespread application. A potential solution to this is to generate image patches of smaller sizes, which may introduce a reduction in the quality of the data but would significantly increase the model's efficiency. Third, the quantification models utilized in this study may

benefit from assisted learning modules, which will allow feature-specific quantification with respect to each tissue type, unlike its current stage.

In conclusion, the implementation of GANs in digital pathology represents a promising avenue for enhancing both the effectiveness of AI in medical diagnostics and the security of patient data. As healthcare continues to evolve alongside AI, the development of secure, synthetic datasets through GANs will be crucial in mitigating the risks of data breaches while unlocking the potential for more advanced, personalized treatment options.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/jpm14070703/s1>, Figure S1: Showing synthetic images of Bladder. Figure S2: Showing synthetic images of Cervix. Figure S3: Showing synthetic images of Kidney. Figure S4: Showing synthetic images of Ovary. Figure S5: Showing synthetic images of Prostate. Figure S6: Showing synthetic images of Testis. Figure S7: Showing synthetic images of the Uterus. Figure S8: Showing synthetic images of the Vagina.

Author Contributions: Conceptualization, H.A. and D.J.V.B.; methodology, D.J.V.B. and C.-B.C.; software, D.J.V.B., C.-B.C. and Y.W.; validation, D.J.V.B., C.-B.C., S.M., M.M., O.N.K. and S.P.; formal analysis, H.A., Y.M., D.J.V.B. and C.-B.C.; investigation, H.A., D.J.V.B. and C.-B.C.; resources, H.A.; data curation, H.A. and D.J.V.B.; writing—original draft preparation, H.A., D.J.V.B. and C.-B.C.; writing—review and editing, H.A., D.J.V.B., C.-B.C., S.M., M.M., O.N.K., S.P. and Y.W.; visualization, D.J.V.B. and C.-B.C.; supervision, H.A.; project administration, H.A.; funding acquisition, H.A., D.J.V.B. and C.-B.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Scott R. MacKenzie Foundation grant number (AWD-009374). This research was also partially funded by the University of Miami U-LINK grant number PG013269 and Provost Research Award PG012860 and PG015498. S.P. has additional support from NIH/NCI (U01CA239141, 1R01CA272766) and Paps Corps Champions for Cancer Research.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data will be made available upon request to corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

Disclosure of Patent Information: The authors wish to inform you that the technology presented in this study is part of a provisional patent application that has been filed with the United States Patent and Trademark Office (USPTO). The application has been assigned Serial No. 63/598,207 and was filed on 13 November 2023. The patent application is currently pending. Some of the authors of this paper are listed as inventors in the patent application. This patent filing may constitute a potential conflict of interest, and this statement serves to disclose this relationship in the interest of full transparency.

References

1. Ali, H.; Muzammil, M.A.; Dahiya, D.S.; Ali, F.; Yasin, S.; Hanif, W.; Gangwani, M.K.; Aziz, M.; Khalaf, M.; Basuli, D.; et al. Artificial intelligence in gastrointestinal endoscopy: A comprehensive review. *Ann. Gastroenterol.* **2024**, *37*, 133–141. [[CrossRef](#)]
2. Caloro, E.; Gnocchi, G.; Quarrella, C.; Ce, M.; Carrafiello, G.; Cellina, M. Artificial Intelligence in Bone Metastasis Imaging: Recent Progresses from Diagnosis to Treatment—A Narrative Review. *Crit. Rev. Oncog.* **2024**, *29*, 77–90. [[CrossRef](#)]
3. Fabijan, A.; Zawadzka-Fabijan, A.; Fabijan, R.; Zakrzewski, K.; Nowoslawska, E.; Polis, B. Artificial Intelligence in Medical Imaging: Analyzing the Performance of ChatGPT and Microsoft Bing in Scoliosis Detection and Cobb Angle Assessment. *Diagnostics* **2024**, *14*, 773. [[CrossRef](#)]
4. Li, Q.; Tan, J.; Xie, H.; Zhang, X.; Dai, Q.; Li, Z.; Yan, L.L.; Chen, W. Evaluating the accuracy of the Ophthalmologist Robot for multiple blindness-causing eye diseases: A multicentre, prospective study protocol. *BMJ Open* **2024**, *14*, e077859. [[CrossRef](#)] [[PubMed](#)]
5. Vitt, J.R.; Mainali, S. Artificial Intelligence and Machine Learning Applications in Critically Ill Brain Injured Patients. *Semin. Neurol.* **2024**, *44*, 342–356. [[CrossRef](#)]
6. Zhang, P.; Gao, C.; Huang, Y.; Chen, X.; Pan, Z.; Wang, L.; Dong, D.; Li, S.; Qi, X. Artificial intelligence in liver imaging: Methods and applications. *Hepatol. Int.* **2024**, *18*, 422–434. [[CrossRef](#)] [[PubMed](#)]
7. Pinto-Coelho, L. How Artificial Intelligence Is Shaping Medical Imaging Technology: A Survey of Innovations and Applications. *Bioengineering* **2023**, *10*, 1435. [[CrossRef](#)]

8. Prassas, I.; Clarke, B.; Youssef, T.; Phlamon, J.; Dimitrakopoulos, L.; Rofaeil, A.; Yousef, G.M. Computational pathology: An evolving concept. *Clin. Chem. Lab. Med.* **2024**; *online ahead of print*. [[CrossRef](#)]
9. Soliman, A.; Li, Z.; Parwani, A.V. Artificial intelligence's impact on breast cancer pathology: A literature review. *Diagn. Pathol.* **2024**, *19*, 38. [[CrossRef](#)] [[PubMed](#)]
10. Wang, Y.L.; Gao, S.; Xiao, Q.; Li, C.; Grzegorzec, M.; Zhang, Y.Y.; Li, X.H.; Kang, Y.; Liu, F.H.; Huang, D.H.; et al. Role of artificial intelligence in digital pathology for gynecological cancers. *Comput. Struct. Biotechnol. J.* **2024**, *24*, 205–212. [[CrossRef](#)]
11. Yilmaz, F.; Brickman, A.; Najdawi, F.; Yakirevich, E.; Egger, R.; Resnick, M.B. Advancing Artificial Intelligence Integration Into the Pathology Workflow: Exploring Opportunities in Gastrointestinal Tract Biopsies. *Lab. Investig.* **2024**, *104*, 102043. [[CrossRef](#)]
12. Ting, D.S.J.; Foo, V.H.; Yang, L.W.Y.; Sia, J.T.; Ang, M.; Lin, H.; Chodosh, J.; Mehta, J.S.; Ting, D.S.W. Artificial intelligence for anterior segment diseases: Emerging applications in ophthalmology. *Br. J. Ophthalmol.* **2021**, *105*, 158–168. [[CrossRef](#)]
13. Ting, D.S.W.; Peng, L.; Varadarajan, A.V.; Keane, P.A.; Burlina, P.M.; Chiang, M.F.; Schmetterer, L.; Pasquale, L.R.; Bressler, N.M.; Webster, D.R.; et al. Deep learning in ophthalmology: The technical and clinical considerations. *Prog. Retin. Eye Res.* **2019**, *72*, 100759. [[CrossRef](#)]
14. Dey, D.; Slomka, P.J.; Leeson, P.; Comaniciu, D.; Shrestha, S.; Sengupta, P.P.; Marwick, T.H. Artificial Intelligence in Cardiovascular Imaging: JACC State-of-the-Art Review. *J. Am. Coll. Cardiol.* **2019**, *73*, 1317–1335. [[CrossRef](#)]
15. Esmaeilzadeh, P. Challenges and strategies for wide-scale artificial intelligence (AI) deployment in healthcare practices: A perspective for healthcare organizations. *Artif. Intell. Med.* **2024**, *151*, 102861. [[CrossRef](#)] [[PubMed](#)]
16. Huang, Y.; Guo, J.; Chen, W.H.; Lin, H.Y.; Tang, H.; Wang, F.; Xu, H.; Bian, J. A scoping review of fair machine learning techniques when using real-world data. *J. Biomed. Inform.* **2024**, *151*, 104622. [[CrossRef](#)]
17. Cao, P.; Derhaag, J.; Coonen, E.; Brunner, H.; Acharya, G.; Salumets, A.; Zamani Esteki, M. Generative artificial intelligence to produce high-fidelity blastocyst-stage embryo images. *Hum. Reprod.* **2024**, *39*, 1197–1207. [[CrossRef](#)] [[PubMed](#)]
18. Ivanenko, M.; Wanta, D.; Smolik, W.T.; Wroblewski, P.; Midura, M. Generative-Adversarial-Network-Based Image Reconstruction for the Capacitively Coupled Electrical Impedance Tomography of Stroke. *Life* **2024**, *14*, 419. [[CrossRef](#)] [[PubMed](#)]
19. Reddy, S. Generative AI in healthcare: An implementation science informed translational path on application, integration and governance. *Implement. Sci.* **2024**, *19*, 27. [[CrossRef](#)] [[PubMed](#)]
20. Chen, C.B.; Wang, Y.; Fu, X.; Yang, H. Recurrence Network Analysis of Histopathological Images for the Detection of Invasive Ductal Carcinoma in Breast Cancer. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2023**, *20*, 3234–3244. [[CrossRef](#)]
21. Chen, C.B.; Yang, H.; Kumara, S. Recurrence network modeling and analysis of spatial data. *Chaos* **2018**, *28*, 085714. [[CrossRef](#)]
22. Yang, H.; Chen, C.B.; Kumara, S. Heterogeneous recurrence analysis of spatial data. *Chaos* **2020**, *30*, 013119. [[CrossRef](#)] [[PubMed](#)]
23. Zhang, L.; Zhao, Z.; Zhang, Y.; Zhang, S.; Xie, D.; Pu, S.; Mao, H. Efficient Shift Network in Denoising-Friendly Space for Real Noise Removal. In Proceedings of the 2022 IEEE International Conference on Multimedia and Expo (ICME), Taipei, Taiwan, 18–22 July 2022.
24. Raja, L.; Merline, A.; Ganesan, R. Indexing of the discrete globalgrid using linear quadtree. *Int. J. Adv. Inf. Technol.* **2013**, *2*.
25. Kumar, K.; Naga Sai Ram, K.N.; Kiranmai, K.S.S.; Harsha, S. Denoising of Iris Image Using Stationary Wavelet Transform. In Proceedings of the 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, India, 20–21 April 2018; pp. 1232–1237.
26. Wang, Y.; Chen, C.-B. Recurrence Quantification Analysis for Spatial Data. In Proceedings of the IIE Annual Conference, Seattle, WA, USA, 21–24 May 2022; pp. 1–6.
27. Shukla, P.; Verma, A.; Abhishek; Verma, S.; Kumar, M. Interpreting SVM for medical images using Quadtree. *Multimed. Tools Appl.* **2020**, *79*, 29353–29373. [[CrossRef](#)] [[PubMed](#)]
28. Bai, J.; Zhao, X.; Chen, J. Indexing of the discrete global grid using linear quadtree. In Proceedings of the ISPRS Workshop on Service and Application of Spatial Data Infrastructure, XXXVI(4/W6), Hangzhou, China, 14–16 October 2005.
29. Li, X.; Wang, C.; Sheng, Y.; Zhang, J.; Wang, W.; Yin, F.F.; Wu, Q.; Wu, Q.J.; Ge, Y. An artificial intelligence-driven agent for real-time head-and-neck IMRT plan generation using conditional generative adversarial network (cGAN). *Med. Phys.* **2021**, *48*, 2714–2723. [[CrossRef](#)] [[PubMed](#)]
30. Yang, S.; Qiao, K.; Qin, R.; Xie, P.; Shi, S.; Liang, N.; Wang, L.; Chen, J.; Hu, G.; Yan, B. ShapeEditor: A StyleGAN Encoder for Stable and High Fidelity Face Swapping. *Front. Neurobotics* **2021**, *15*, 785808. [[CrossRef](#)]
31. Bian, Y.; Wang, J.; Jun, J.J.; Xie, X.Q. Deep Convolutional Generative Adversarial Network (dcGAN) Models for Screening and Design of Small Molecules Targeting Cannabinoid Receptors. *Mol. Pharm.* **2019**, *16*, 4451–4460. [[CrossRef](#)]
32. Maguluri, G.; Grimble, J.; Caron, A.; Zhu, G.; Krishnamurthy, S.; McWatters, A.; Beamer, G.; Lee, S.Y.; Iftimia, N. Core Needle Biopsy Guidance Based on Tissue Morphology Assessment with AI-OCT Imaging. *Diagnostics* **2023**, *13*, 2276. [[CrossRef](#)]
33. Zou, H.; Hastie, T.; Tibshirani, R. Sparse Principal Component Analysis. *J. Comput. Graph. Stat.* **2006**, *15*, 265–286. [[CrossRef](#)]
34. Song, F.; Guo, Z.; Mei, D. Feature Selection Using Principal Component Analysis. In Proceedings of the 2010 International Conference on System Science, Engineering Design and Manufacturing Informatization, Yichang, China, 12–14 November 2010.
35. Lin, W.; Pixu, S.; Rui, F.; Hongzhe, L. Variable selection in regression with compositional covariates. *Biometrika* **2014**, *101*, 785–797. [[CrossRef](#)]
36. Zou, H.; Hastie, T. Regularization and Variable Selection Via the Elastic Net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2005**, *67*, 301–320. [[CrossRef](#)]

37. Matsui, H.; Konishi, S. Variable selection for functional regression models via the L1 regularization. *Comput. Stat. Data Anal.* **2011**, *55*, 3304–3310. [[CrossRef](#)]
38. Wolfe, P.J.; Godsill, S.J.; Ng, W.-J. Bayesian Variable Selection and Regularization for Time–Frequency Surface Estimation. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2004**, *66*, 575–589. [[CrossRef](#)]
39. Al Sudani, Z.A.; Salem, G.S.A. Evaporation Rate Prediction Using Advanced Machine Learning Models: A Comparative Study. *Adv. Meteorol.* **2022**, *2022*, 1433835. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.