

https://dspace.org/

# Synthetic cytology image generation to augment teaching and quality assurance in pathology

**McAlpine, Ewen David**

**2024**

https://hdl.handle.net/10539/37570

Downloaded from a DSpace repository. https://dspace.org/

# SYNTHETIC CYTOLOGY IMAGE GENERATION TO AUGMENT TEACHING AND QUALITY ASSURANCE IN PATHOLOGY

EWEN DAVID McALPINE

MBBCh (Wits), FCPath (SA) ANAT, MMed (Wits)

A thesis submitted in fulfilment of the requirements for the degree of

## DOCTOR OF PHILOSOPHY

in field of

## ANATOMICAL PATHOLOGY

at the

## UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG




## 2023

*We make our world significant by the courage of our questions*

*and the depth of our answers*

Carl Sagan

# STUDENT DECLARATION

I, Ewen David McAlpine, declare that this Thesis is my own, unaided work. It is being submitted for the *Degree of Doctor of Philosophy* at the University of the Witwatersrand, Johannesburg.

It has not been submitted before for any degree or examination at any other University.

Signed:  on this 24th day of May 2023 at Fourways, Johannesburg.

# DEDICATION

Firstly, to my family and friends who have patiently listened to me think out loud about this work time and time again and who have understood (hopefully) when I have been unable to spend time with them to complete this work.

Secondly, to all my pathology teachers and trainers who have guided me along my professional journey – both as a trainee and as a pathologist.

# PRESENTATIONS AND PUBLICATIONS

*Publications arising during the course of this Doctor of Philosophy*

Publications make * are included in the Integrated Narrative of this corpus of work.

## GENERAL BACKGROUND

**McAlpine E**, Michelow P (2020). The cytopathologist's role in developing and evaluating artificial intelligence in cytopathology practice. *Cytopathology*. 2020 31(5): 385-392. doi: 10.1111/cyt.12799

**McAlpine E**, Michelow P (2020). Implementing Deep Learning Algorithms in Anatomical Pathology using open-source Deep Learning Libraries. *Advances in Anatomic Pathology*. 27(4): 260-268. doi: 10.1097/PAP.0000000000000265.

**McAlpine E**, Pantanowitz L, Michelow P (2021). Challenges Developing Deep Learning Algorithms in Cytology. *Acta Cytologica.* 65:301–309. doi: 10.1159/000510991

## LITERATURE REVIEW

*** McAlpine E**, Michelow P, Celik T (2022). The utility of unsupervised machine learning in Anatomic Pathology. *American Journal of Clinical Pathology*. 2022 January. 157(1):5-14. doi:10.1093/ajcp/aqab085

## RESEARCH ARTICLES

*** McAlpine E**, Michelow P, Celik T (2022). The dynamics of pathology dataset creation using urine cytology as an example. *Acta Cytologica*. 2022 January. 66:46–54. doi: 10.1159/000519273

*** McAlpine E**, Michelow P, Liebenberg E, Celik T (2022). Is it real or not? Towards AI-based realistic synthetic cytology image generation to augment teaching and quality assurance in pathology. *Journal of the American Society of Cytopathology*. 2022 May – June. 11:123-132. doi: 10.1016/j.jasc.2022.02.001

*** McAlpine E**, Michelow P, Liebenberg E, Celik T (2022). Are synthetic cytology images ready for prime time? A comparative assessment of real and synthetic urine cytology images. *Journal of the American Society of Cytopathology*. 2023. March – April 12:126-135. doi: 10.1016/j.jasc.2022.10.001

*Presentations given during the course of this Doctor of Philosophy*

**January 2022** – South African Division of the International Academy of Pathology (virtual congress). An Introduction to Artificial Intelligence.

**August 2022** - American Society of Cytopathology Podcast CytoPathPod

https://cytopathpod.podbean.com/e/jascarticle-is-itrealor-not-towardartificialintelligence-basedrealisticsyntheticcytology-imagegenerationto-augment-teachingandquality-assurance-inpath/

# ABSTRACT


## INTRODUCTION

Urine cytology offers rapid and relatively inexpensive screening for the detection of high-grade urothelial neoplasia in patients with haematuria. In our setting of a public sector laboratory in South Africa, however, there is a paucity of such specimens with which to train cytotechnologists and cytopathologists. Advancements in Generative Adversarial Networks present a potential solution to this problem by allowing for the generation of synthetic urine cytology images to supplement teaching and training. We illustrate an end-to-end machine learning model – from dataset creation to testing synthetic images in pathology personnel – to assess this technology in a real-world setting.

## METHODS

Two hundred and fourteen urine cytology slides were digitised and processed to construct a morphologically balanced dataset containing examples of benign, atypical and malignant urine cytology images. This dataset was used to train a StyleGAN3 model to generate synthetic urine cytology images. These synthetically generated images were then tested in a group of pathology personnel – both pathologists and trainees – to assess whether a difference between real and synthetic urine cytology images exists. Diagnostic error rate and subject image assessment were tested.

## RESULTS

StyleGAN3 was able to generate a wide morphological diversity of realistically appearing benign, atypical and malignant urine cytology images. When testing how these synthetic images were perceived by pathology personnel, there was no significant difference in diagnostic error rate, subjective image quality or inclusion of synthetic images in a cytology teaching set.

## DISCUSSION

This work presents a proof-of-concept illustration of the feasibility of the use of synthetic cytology images to supplement pathology teaching when real examples may be difficult to obtain. Furthermore, this work presents important insights into the dynamics of pathology dataset creation and discusses the use of synthetic data in health education and the ethical and legal issues that arise with the use of synthetic patient data.

## CONCLUSION

Our work demonstrates that realistic, morphologically diverse urine cytology images can be generated using existing GANs technology and that human observers find such synthetic data visually acceptable. Additionally, our data indicate that there is no significant difference in synthetic data in terms of subjective image quality or diagnostic classification as determined by pathology personnel.

# ACKNOWLEDGEMENTS

Invariably, projects of this nature and scope require the commitment and assistance of many individuals to realise. I would like to express my sincere gratitude to the following people:

- My supervisors, Pamela Michelow and Turgay Celik, for their guidance and patience throughout the process of this PhD.
- Eric Liebenberg for his continued maintenance of the infrastructure used in this project and assistance with StyleGAN experiments.
- Paul Diessel for his assistance in developing the online image survey system utilised in this research.
- 3F Scientific for their assistance with digitising the slides used in this research.
- National Health Laboratory Service and Lancet Laboratories for providing the urine cytology slides used in this project.
- Faculty of Health Sciences Research Equipment Grants of the University of the Witwatersrand and NVIDIA Academic Hardware Grant which partially funded the IT infrastructure utilised in this PhD.

# TABLE OF CONTENTS

STUDENT DECLARATION ..... iii

DEDICATION ..... iv

PRESENTATIONS AND PUBLICATIONS ..... v

GENERAL BACKGROUND ..... v

LITERATURE REVIEW ..... v

RESEARCH ARTICLES ..... v

ABSTRACT ..... vii

ACKNOWLEDGEMENTS ..... ix

TABLE OF CONTENTS ..... x

LIST OF FIGURES ..... xii

LIST OF TABLES ..... xiii

NOMENCLATURE AND ABBREVIATIONS ..... xiv

CHAPTER 1: INTRODUCTION ..... 1

CHAPTER 2: AIMS AND OBJECTIVES ..... 5

CHAPTER 3: OVERVIEW OF THE PRESENTED CORPUS OF WORK ..... 6

3.1 THE UTILITY OF UNSUPERVISED MACHINE LEARNING IN ANATOMIC PATHOLOGY ..... 6

3.2 THE DYNAMICS OF PATHOLOGY DATASET CREATION USING URINE CYTOLOGY AS AN EXAMPLE ..... 7

3.3 IS IT REAL OR NOT? TOWARD ARTIFICIAL INTELLIGENCE-BASED REALISTIC SYNTHETIC CYTOLOGY IMAGE GENERATION TO AUGMENT TEACHING AND QUALITY ASSURANCE IN PATHOLOGY ..... 9

3.4 ARE SYNTHETIC CYTOLOGY IMAGES READY FOR PRIME TIME? A COMPARATIVE ASSESSMENT OF REAL AND SYNTHETIC URINE CYTOLOGY IMAGES ..... 11

CHAPTER 4: INTEGRATED NARRATIVE ..... 13

4.1 PUBLISHED LITERATURE REVIEW – UNSUPERVISED LEARNING ..... 13

4.2 OVERVIEW OF RESEARCH AREA....14
4.3 ADDITIONAL METHODOLOGY....15
4.4 SYNTHESIS AND DISCUSSION....18
CHAPTER 5: REFERENCES....47
CHAPTER 6: PUBLISHED RESEARCH....55
6.1: THE DYNAMICS OF PATHOLOGY DATASET CREATION USING URINE CYTOLOGY AS AN EXAMPLE....55
6.2: IS IT REAL OR NOT? TOWARDS AI-BASED REALISTIC SYNTHETIC CYTOLOGY IMAGE GENERATION TO AUGMENT TEACHING AND QUALITY ASSURANCE IN PATHOLOGY.....56
6.3: ARE SYNTHETIC CYTOLOGY IMAGES READY FOR PRIME TIME? A COMPARATIVE ASSESSMENT OF REAL AND SYNTHETIC URINE CYTOLOGY IMAGES.....67
CHAPTER 7: SUMMARY AND CONCLUSIONS....78
APPENDICES....79
APPENDIX 1: STUDENT AND CO-AUTHORS DECLARATION....79
APPENDIX 2: ETHICS CERTIFICATE....84
APPENDIX 3: TURNITIN REPORT....85
APPENDIX 4: JOURNAL PERMISSIONS....86

# LIST OF FIGURES

Figure 1: (A) Conceptualising the relationship between artificial intelligence, machine learning and deep learning. (B) Types of machine learning. From McAlpine et al[9], used with permission. .....2
Figure 2: A visual summary of the TPS criteria applied in this work. From McAlpine et al[22], used with permission. .....20
Figure 3: An illustration of a GANs model, comprising a Generator that learns to generate a synthetic image from random noise, and a Discriminator that learns to distinguish these synthetic images from real images. From McAlpine et al[10], used with permission. .....24
Figure 4: Image labelling time by diagnostic category. B, benign; A, atypical; M, malignant; S, squamous; R, reject. From McAlpine et al[27], used with permission. .....31
Figure 5: Image labelling time for images before and after the median relative label time. From McAlpine et al[27], used with permission. .....33
Figure 6: An illustration of the difference in image size/resolution between the images generated in Article #3 and Article #4. .....35
Figure 7: A heatmap comparing the p-values for the 12 individual observers (rows 0 through 11), for error rates, inclusion in a teaching set and subjective image quality score. From McAlpine et al[22], used with permission. .....38
Figure 8:StyleGAN3 initiation results. .....43
Figure 9: StyleGAN3 results at 20kimgs. .....44
Figure 10: StyleGAN3 results at 100kimgs .....45
Figure 11: StyleGAN3 results at 3,410kimgs .....46

# LIST OF TABLES

Table 1: Summary of the QA dataset and associated level of agreement. From McAlpine et al[27], used with permission. .....32
Table 2: Summary of the level of agreement between urothelial categories before and after the median relative label time. From McAlpine et al[27], used with permission. .....33

# NOMENCLATURE AND ABBREVIATIONS

| | |
|---|---|
| AI | Artificial intelligence |
| AUC | Area under the curve |
| DP | Digital Pathology (including Digital Cytology) |
| FISH | Fluorescence in situ hybridisation |
| GANs | Generative Adversarial Networks |
| GradCAM | Gradient-weighted Class Activation Mapping |
| N/C | Nuclear to cytoplasmic ratio |
| px | Pixel |
| QA | Quality Assurance |
| Registrar | Postgraduate medical specialist trainee, synonymous with the term “Resident” |
| ROC | Receiver operating characteristic |
| TPS | The Paris System for Reporting Urine Cytology (1st edition) |
| WSI | Whole Slide Image |

# CHAPTER 1: INTRODUCTION

The application of artificial intelligence (AI) to pathology is considered to be the next revolution in this domain of medicine – following on from the major changes in practice related to the introduction of immunohistochemistry and molecular techniques in anatomical pathology[1]. This “third revolution” in pathology dovetails with the rise of digital pathology (specifically whole slide scanners) and digital pathology is regarded as a necessary initial step in the process of introducing AI to pathology diagnostics[2].

Broadly speaking, the term artificial intelligence is used to refer to any technology that allows computers to mimic human behaviour. Machine learning is regarded as a subset of artificial intelligence and makes use of algorithms that learn without rules being explicitly programmed. Deep learning refers to techniques that involve using large amounts of data to train deep neural networks[3]. Figure 1 illustrates the broad concepts of AI, machine learning and deep learning, as well as the broad categories of supervised, unsupervised and reinforcement learning.

The advent of whole slide scanners, which digitise pathology slides, has allowed for the development of pathology-specific AI algorithms[4]. Moreover, AI has the ability to learn from the enormous quantities of data generated in the process of patient care, with the ultimate goal of improving such care by refining diagnostics, classification, and prognostication of disease[4]. Digital pathology has revolutionised daily pathology practice in large parts of the world[4], however, this technology remains largely unadopted for routine diagnostics in our local context. This is expected to change in coming years, most likely with the private sector adopting digital pathology first followed by the public/state sector.

Hanna et al[5], neatly group the use cases for machine learning in pathology into three broad categories:

i. Tools that help pathologists in their goal to deliver safer patient care by improving diagnostic accuracy,
ii. Tools that improve productivity, and
iii. Tools that are used to detect novel digital biomarkers.

Furthermore, these authors indicate important advantages that digital pathology offers. The shift to a digital workflow produces distinct productivity advantages, including, rapid and

easy access to previous pathology specimens and digital consultation[5]. Africa as a continent has not been spared from the global shortage of pathologists and a worldwide trend of declining pathologist numbers[6–8]. Digital pathology has the potential to address this, at least partially, by allowing for the provision of anatomical pathology services to areas of our continent that are currently underserviced. Furthermore, AI may offer a solution to the shortage of pathologists in the long term – by both making existing pathologists more productive and by automating some of the diagnostic burden.

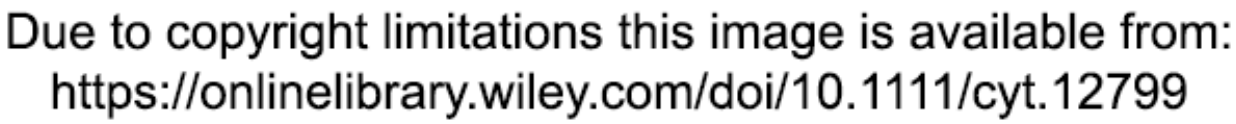



Figure 1: (A) Conceptualising the relationship between artificial intelligence, machine learning and deep learning. (B) Types of machine learning. From McAlpine et al[9], used with permission.

This corpus of work presents an end-to-end machine learning project that aimed to illustrate the potential solution that Generative Adversarial Networks (GANs) offer to a problem encountered in our local pathology setting. The motivation behind this project is set out in our work entitled “Is it real or not? Towards AI-based realistic synthetic cytology image generation to augment teaching and quality assurance in pathology”[10] (**Article #3**). Briefly, the paucity of urine cytology specimens representing urothelial neoplasia in our setting of a

public sector laboratory negatively impacts teaching, training and on-going Quality Assurance (QA). GANs offer a unique solution to this challenge by using a pair of deep neural networks that, once trained, can generate unique and realistic examples of the data used to train them (termed *synthetic data*). This thesis presents a use case of applying GANs technology to cytology to solve the challenge set out above and addresses important questions relating to the use of this technology in a teaching/training context – such as how these images are perceived by pathology personnel.

Implementing this technology in cytology in general and, our local setting in particular presented a number of challenges. Digital pathology, including cytology, is not routinely used in either of the two laboratories from which the material was drawn.

i. The cytologists involved in this project had no real-world experience in digital cytology or the processing of whole slide images (WSIs).
ii. The considerable computing resources required to train GANs were not readily available in our teaching department prior to this project.
iii. GANs require large amounts of data for training. This data was not available, nor was it feasible to label a dataset of the size usually required for GANs training – tens of thousands or hundreds of thousands of cytology images. This is, in part, due to a severe shortage of pathology personnel and a high diagnostic workload.
iv. Significant morphological diversity exists in all pathology specimens (including urine cytology) – even within the same diagnostic category. This relates to factors such as staining variation, slide background and variation in cellular and non-cellular elements.

What is presented in this thesis are four articles that follow this process from conception to testing real and synthetic urine cytology images in a cohort of pathology personnel. Chapter 2 presents the overall aims and objective of this study, Chapter 3 presents an overview of the published work, Chapter 4 comprises an integrated narrative/synthesis and discussion of the work (including a published literature review) and Chapter 6 presents the remaining three of the four published works included in this thesis.

The following publications are included:

**Article #1:**

*The Utility of Unsupervised Machine Learning in Anatomic Pathology*[11], published in the American Journal of Clinical Pathology, is a review of the current literature pertaining to unsupervised machine learning in pathology (including GANs).

**Article #2:**

*The Dynamics of Pathology Dataset Creation Using Urine Cytology as an Example*[12], published in Acta Cytologica, presents our findings from analysing the dynamics of annotating a modest urine cytology dataset.

**Article #3:**

*Is it real or not? Toward artificial intelligence-based realistic synthetic cytology image generation to augment teaching and quality assurance in pathology*[10], published in the Journal of the American Society of Cytopathology, presents a proof-of-concept illustration of using GANs to create realistic synthetic malignant urine cytology images using modest computing resources and with very limited data.

**Article #4:**

*Are synthetic cytology images ready for prime time? A comparative assessment of real and synthetic urine cytology images*[13], published in the Journal of the American Society of Cytopathology, further develops this to produce higher resolution benign, atypical and malignant urine cytology images and tests these images on a cohort of pathology personnel.

While this project was undertaken to address a problem that is potentially unique to our local setting, the general concept can be more widely implemented for pathology teaching and learning, QA, proficiency testing and, even potentially, examination. Furthermore, theoretical future uses cases for GANs technology in pathology are presented later in this work. Research into the application of GANs technology in pathology began seriously in 2018 and much research in this domain remains to be undertaken.

# CHAPTER 2: AIMS AND OBJECTIVES

**AIM:**

The aim of the research contained in this thesis was to utilise machine learning technology (specifically, GANs) to generate synthetic urine cytology images for training and quality assurance purposes.

To this end, the following objectives were set:

**OBJECTIVES:**

1. Construct a labelled dataset of high-magnification urine cytology images inclusive of benign, atypical, and malignant urothelial cells.
2. Assess the dynamics of pathology dataset creation – including QA measures such as labelling agreement.
3. Create a dataset of highly realistic benign, atypical and malignant synthetic urine cytology images using GANs.
4. Assess the subjective evaluation of real and synthetic images and diagnostic error rates between real and synthetic images by pathology personnel.

# CHAPTER 3: OVERVIEW OF THE PRESENTED CORPUS OF WORK

## 3.1 THE UTILITY OF UNSUPERVISED MACHINE LEARNING IN ANATOMIC PATHOLOGY

**American Journal of Clinical Pathology. 2022 January;157(1):5-14.**

Published: 24 July 2021

### ABSTRACT

Due to copyright restrictions, this abstract is available from:
https://academic.oup.com/ajcp/article/157/1/5/6327583

### KEY CONTRIBUTIONS

- The article makes the point that one of the major barriers to the development of supervised machine learning in pathology is the relative scarcity of large, high-quality labelled pathology datasets.
- Data annotation is an onerous task and unsupervised learning offers a potential solution to this problem.
- This review introduces the concept of unsupervised machine learning to the general pathology community. The principal position taken in this article is the idea of utilising unsupervised techniques such as clustering, generative adversarial networks (GANs) and autoencoders to improve the performance and generalisability of supervised machine learning algorithms by leveraging the potential of large volumes of unlabelled pathology data.
- It further qualifies that pathology dataset annotation may be difficult to undertake in light of dwindling pathologist numbers and increasing diagnostic workload.
- The review presents relevant literature relating to the use of clustering, GANs and autoencoders in pathology.
- Specifically relating to GANs, this article discusses the use of generative models in stain normalisation, stain transfer and the creation of synthetic data in pathology (as performed in this work).

## 3.2 THE DYNAMICS OF PATHOLOGY DATASET CREATION USING URINE CYTOLOGY AS AN EXAMPLE

**Acta Cytologica. 2022 January-February;66(1):46-54.**

Published: 18 October 2021

### ABSTRACT

Introduction

Dataset creation is one of the first tasks required for training AI algorithms but is underestimated in pathology. High-quality data are essential for training algorithms and data should be labelled accurately and include sufficient morphological diversity. The dynamics and challenges of labelling a urine cytology dataset using The Paris System (TPS) criteria are presented.

Methods

2,454 images were labelled by pathologist consensus via video conferencing over a 14-day period. During the labelling sessions, the dynamics of the labelling process were recorded. Quality assurance images were randomly selected from images labelled in previous sessions within this study and randomly distributed throughout new labelling sessions. To assess the effect of time on the labelling process, the labelled set of images was split into 2 groups according to the median relative label time and the time taken to label images and intersession agreement were assessed.

Results

Labelling sessions ranged from 24m 11s to 41m 06s in length, with a median of 33m 47s. The majority of the 2,454 images were labelled as benign urothelial cells, with atypical and malignant urothelial cells more sparsely represented. The time taken to label individual images ranged from 1s to 42s with a median of 2.9s. Labelling times differed significantly among categories, with the median label time for the atypical urothelial category being 7.2s, followed by the malignant urothelial category at 3.8s and the benign urothelial category at 2.9s. The overall intersession agreement for quality assurance images was substantial. The level of agreement differed among classes of urothelial cells - benign and malignant urothelial cell classes showed almost perfect agreement and the atypical urothelial cell class

showed moderate agreement. Image labelling times seemed to speed up, and there was no evidence of worsening of intersession agreement with session time.

Discussion/conclusion

Important aspects of pathology dataset creation are presented, illustrating the significant resources required for labelling a large dataset. We present evidence that the time taken to categorise urine cytology images varies by diagnosis/class. The known challenges relating to the reproducibility of the AUC (atypical) category in TPS when compared to the NHGUC (benign) or HGUC (malignant) categories is also confirmed.

## KEY CONTRIBUTIONS

- This article presents important insights into the process of dataset creation and annotation in pathology – one of the first steps in a machine-learning workflow.
- The annotation of pathology datasets is a laborious and time-consuming process. This work aimed to quantify this process and address issues such as whether labelling time differs among diagnostic classes and whether visual fatigue develops in the labelling process – factors that will affect the time taken to label and the quality of a dataset.
- As stated in the article, pathology datasets need to account for the morphological diversity inherent to pathology specimens – both cytology and histology. We make use of K-means clustering with cosine similarity to select images for labelling – a form of active learning called diversity sampling[14] – in an attempt to address this question of morphological diversity.
- Further explanation of what this concept of morphological diversity relates to is presented in **Article #3** of this work and relates not only to diagnostic classes (benign, atypical and malignant) but also to factors such as slide background and morphological patterns of cellular and non-cellular components.
- We also assimilate suggestions for dataset creation from various sources and present important factors that need to be considered when creating a pathology dataset.
- In addition, we present an example of how quality assurance may be instituted in the annotation process.
- Finally, we confirm the known challenges of the reproducibility of TPS – specifically in the atypical category – even when applied by expert cytologists in unison.

## 3.3 IS IT REAL OR NOT? TOWARD ARTIFICIAL INTELLIGENCE-BASED REALISTIC SYNTHETIC CYTOLOGY IMAGE GENERATION TO AUGMENT TEACHING AND QUALITY ASSURANCE IN PATHOLOGY

**Journal of the American Society of Cytopathology. 2022 May-June;11(3):123-132.**
Published: 8 February 2022

**ABSTRACT**

Introduction

Urine cytology offers a rapid and relatively inexpensive method to diagnose urothelial neoplasia. In our setting of a public sector laboratory in South Africa, urothelial neoplasia is rare, compromising pathology training in this specific aspect of cytology. Artificial intelligence-based synthetic image generation-specifically the use of generative adversarial networks (GANs)-offers a solution to this problem.

Materials and methods

A limited, but morphologically diverse, dataset of 1000 malignant urothelial cytology images was used to train a StyleGAN3 model to create completely novel, synthetic examples of malignant urine cytology using computer resources within reach of most pathology departments worldwide.

Results

We have presented the results of our trained GAN model, which was able to generate realistic, morphologically diverse examples of malignant urine cytology images when trained using a modest dataset. Although the trained model is capable of generating realistic images, we have also presented examples for which unrealistic and artifactual images were generated-illustrating the need for manual curation when using this technology in a training context.

Conclusions

We have presented a proof-of-concept illustration of creating synthetic malignant urine cytology images using machine learning technology to augment cytology training when real-world examples are sparse. We have shown that despite significant morphologic diversity in terms of staining variations, slide background, variations in the diagnostic malignant cellular

elements, the presence of other nondiagnostic cellular elements, and artifacts, visually acceptable and varied results are achievable using limited data and computing resources.

## KEY CONTRIBUTIONS

- This key proof-of-concept article describes and illustrates the process of generating synthetic malignant urine cytology images – specifically to address a need for teaching material in our local setting.
- The work presents the central motivation for this entire project – the relative lack of examples of urothelial neoplasia in urine cytology in our public sector laboratory. The vast majority of our laboratory's malignant urine cytology specimens represent squamous cell carcinomas of cervical origin (submitted for staging purposes).
- Urothelial neoplasia is not an uncommon malignancy in South Africa and cytology personnel need to be able to identify high-grade urothelial cell neoplasia on urine cytology.
- We present a potential solution to the lack of training material using GANs to generate synthetic cytology images. Moreover, we illustrate how this can be accomplished with limited computing resources.
- We used a specific iteration of GANs to achieve our results – StyleGAN3 – because of its success in working with limited datasets due to Adaptive Discriminator Augmentation (ADA) technology.
- We further illustrate that StyleGAN3 is able to generate a wide range of morphological patterns (i.e., does not suffer from mode collapse).
- In this work, we also discuss the use of synthetic healthcare data in medical education and research and the novel ethical issue relating to this emerging technology.
- The predominant limitation of this work is that we limited the GANs to generating malignant examples only and at a resolution half of what we originally intended. This was due to limited computing resources (a single 12Gb GPU). This was rectified in the final work (**Article #4**) presented below, where we generated benign, atypical and malignant examples at 512x512px when additional computing resources became available.

## 3.4 ARE SYNTHETIC CYTOLOGY IMAGES READY FOR PRIME TIME? A COMPARATIVE ASSESSMENT OF REAL AND SYNTHETIC URINE CYTOLOGY IMAGES

**Journal of the American Society of Cytopathology.**

Article in press. Published online: 7 October 2022

**ABSTRACT**

Introduction

The use of synthetic data in pathology has to date predominantly been augmenting existing pathology data to improve supervised machine learning algorithms. We present an alternative use case – using synthetic images to augment cytology training when the availability of real-world examples is limited. Moreover, we compare the assessment of real and synthetic urine cytology images by pathology personnel to explore the usefulness of this technology in a real-world setting.

Materials and Methods

Synthetic urine cytology images were generated using a custom-trained conditional StyleGAN3 model. A morphologically balanced 60-image dataset of real and synthetic urine cytology images was created for an online image survey system to allow for the assessment of the differences in visual perception between real and synthetic urine cytology images by pathology personnel.

Results

12 participants were recruited to answer the 60-image survey. The study population had a median age of 36.5 years and a median of 5 years of pathology experience. There was no significant difference in diagnostic error rates between real and synthetic images, nor was there a significant difference between subjective image quality scores between real and synthetic images when assessed on an individual observer basis.

Conclusion

The ability of Generative Adversarial Networks (GANs) technology to generate highly realistic urine cytology images was demonstrated. Furthermore, there was no difference in

how pathology personnel perceived the subjective quality of synthetic images, nor was there a difference in diagnostic error rates between real and synthetic urine cytology images. This has important implications for the application of GANs technology to cytology teaching and learning.

**KEY CONTRIBUTIONS:**

- In this work, we demonstrated the ability of StyleGAN3 to generate highly realistic 512x512px images of benign, atypical and malignant urine cytology images.
- The developments in this article were made possible due to two factors: a larger dataset of labelled images and a larger, faster GPU (and NVIDIA RTX 6000).
- We then present the findings of testing these synthetic images on a cohort of pathology personnel. We did this by presenting participants with a self-conducted image survey in which they were presented with both real and synthetic urine cytology images and asked to place the image in a diagnostic category, provide a subjective quality score for each image and state whether they would include the image in a cytology teaching set. Participants were blinded to the fact that the survey included synthetic images.
- When comparing the responses for real and synthetic images on an individual observer basis (i.e., comparing each observer to themselves) we demonstrated that there was no significant difference in diagnostic error rate, subjective quality score or inclusion in a teaching set between real and synthetic images for any of the observers.
- This article also includes a discussion of the literature relating to how the perception of synthetic pathology images has been tested in the past – with some studies opting to use a traditional Turing Test, while others utilised methods more akin to our article – albeit without blinding participants to the fact that the test dataset contained synthetic images.
- Importantly, this work also touches on an important consideration for the use of synthetic images in pathology teaching – how the size/resolution of these images affects the ability of pathologists to detect synthetic images. The literature suggests that the larger the image, the easier it is for a pathologist to detect real from synthetic examples. Most studies to date have used low-resolution images (below 512x512px), more recently, however, higher-resolution images are being tested.

# CHAPTER 4: INTEGRATED NARRATIVE

## 4.1 PUBLISHED LITERATURE REVIEW – UNSUPERVISED LEARNING

The Utility of Unsupervised Machine Learning in Anatomic Pathology

**McAlpine ED**, Michelow P, Celik T. The Utility of Unsupervised Machine Learning in Anatomic Pathology. *Am J Clin Pathol*. 2022;157(1):5-14. doi:10.1093/ajcp/aqab085

This article has been published in the *Americal Journal of Clinical Pathology* (Oxford University Press).

Due to copyright limitations this article can be located at the following website address:

https://academic.oup.com/ajcp/article/157/1/5/6327583

## 4.2 OVERVIEW OF RESEARCH AREA

Since 2014, when Goodfellow et al[15], introduced the concept of GANs, much effort has been directed to improving the quality, size and diversity of synthetic images generated by GANs. The question of whether this technology could be used to address a challenge encountered in our local pathology setting – that of a paucity of examples of certain urine cytology specimens – is addressed in this work.

While the use of GANs technology in medicine in general and pathology, in particular, is not new, to the best of our knowledge, this is the first time such technology has been applied to:

i. Cytology specifically, and
ii. To address a paucity of teaching material rather than use synthetic data to improve the performance and generalisability of supervised machine learning algorithms as has been performed in the literature.

We present an end-to-end machine learning project to solve the problem of sparse examples of urothelial neoplasia on urine cytology specimens in our teaching department.

**Article #1** presents a review of the current literature relating to GANs (and other unsupervised techniques) in pathology, **Article #2** addresses the dynamics of pathology dataset annotation/labelling, **Article #3** presents a proof-of-concept illustration of the generation of low to medium resolution synthetic malignant urine cytology images and **Article #4** presents the culmination of this work in which full-resolution benign, atypical and malignant urine cytology images were generated and tested on a cohort of pathology personnel.

## 4.3 ADDITIONAL METHODOLOGY

### ETHICS

Ethical approval for this study was obtained from the Human Research Ethics Committee (Medical) of the University of the Witwatersrand (Certificate number: M190604, Appendix 2).

Informed consent was obtained from all participants/image observers in **Article #4**.

### STUDY SETTING

This research was conducted in the Division of Anatomical Pathology, University of the Witwatersrand, Johannesburg using urine cytology specimens from both the division itself and a private pathology laboratory, Lancet Laboratories, Johannesburg.

### CASE RETRIEVAL AND SELECTION

A total of 214 WSIs of urine cytology slides were included in this project. These specimens were Papanicolaou-stained voided urine samples. 142 cases (66.4%) were obtained from Lancet Laboratories and 72 cases (33.6%) from National Health Laboratory Service, Johannesburg. Lancet Laboratories is a large private laboratory and National Health Laboratory Service is a large public sector laboratory, both situated in Johannesburg, an urban city in South Africa. It is estimated that just under 20% of South Africans utilise the private sector and the remainder of the population is serviced by the public sector[16]. Most of the cases included in this dataset are from 2018 and 2017, however, following an amended ethics application, cases from 2020 and 2019 were also included, as were cases from our department's teaching collection dating back as far back as 1988. This was done specifically to increase the diversity of examples in the dataset and increase the number of malignant examples in the dataset.

### SUMMARY OF STUDY POPULATION

The study population comprised 187 patients in total (multiple slides from the same patients are included in the 214 total WSIs above). 104 (55.6%) of the patients were male and 82 (43.9%) were female. Gender was unknown for a single case.

The median age of the patients included in this study was 59 years, with a range of 21 to 90 years and an interquartile range (IQR) of 24.

### DIGITISATION OF CYTOLOGY SLIDES

The urine cytology slides were digitised into WSIs using a Pannoramic 250 digital scanner (3DHISTECH Ltd., Budapest, Hungary) at 400x magnification with 7 Z-stacks combined into an extended focus single layer image. The WSIs were converted from 3DHISTECH's propriety MRXS format into BigTIFF format using 3DHISTECH's free conversion utility. These images were tiled into 1024x1024px image patches.

**OBJECT DETECTION AND CROPPING OF IMAGE PATCHES**

An object detector, trained on a custom dataset, was used to localise and crop individual cells and cell groups/clusters. The Python package *Detecto* was used[17] using the methodology described by the candidate in *Implementing Deep Learning Algorithms in Anatomic Pathology Using Open-source Deep Learning Libraries*[18]. This package makes use of the Faster Region-based Convolutional Neural Networks (R-CNN) architecture[19] and leverages transfer learning to speed up training and improve model accuracy on smaller, novel datasets.

The object detector was trained on 1111 labelled 1024x1024px images (1000 training and 111 validation images) derived from 25 WSIs over 20 epochs. Individual cells and cell clusters were annotated using the open-source programme LabelImg[20]. The trained object detector was assessed using an unseen test set of 120 images. The detector detected 367 true positive objects, 37 false positive objects and missed 41 objects of interest (false negatives). Resulting in a precision of 0.91, a recall (sensitivity) of 0.9 and an F-score of 0.9.

The trained object detector was then used to localise cells and cell clusters and 512x512px image patches were cropped around these localised objects.

**FINAL GANs TRAINING SET**

Our morphologically unbalanced image set comprised 7,320 benign urothelial images, 3,367 atypical urothelial images, and 3,059 malignant urothelial images. The 7,320 benign urothelial images were derived from 174 of the 214 slides, atypical urothelial images from 120 slides, and malignant urothelial images from 84 slides. The reason behind this is that malignant examples include a spectrum of benign and atypical cells in addition to malignant cells within the same sample and, similarly, atypical cases include both atypical and benign urothelial cells.

As indicated in **Article #4**, we applied K-means clustering with cosine similarity to balance the set, resulting in a final GANs training set of 3,240 benign urothelial images, 2,400 atypical urothelial images and 1,820 malignant urothelial images.

## MACHINE LEARNING MODELS

Principally, two machine learning algorithms were utilised in the research presented in this thesis:

1. K-means clustering with cosine similarity

   Image features were extracted from images by a pretrained Resnet50 model. These image features were clustered using K-means clustering using cosine similarity as the distance measures, rather than the more usual Euclidean distance. In contrast to Euclidean distance, which can perform poorly with high dimensional data, cosine similarity tends to perform better with high dimensional data – especially sparse data[14].

2. StyleGAN3

   The original implementation of StyleGAN3, created by Nvidia, was utilised in both **Article #3** and **Article #4**. The code for this model is available for download from https://github.com/NVlabs/stylegan3 and is made freely available by Nvidia for non-commercial purposes.

   In both articles, the translation equivariance configuration of StyleGAN3 (StyleGAN3-T), rather than the rotation equivariance configuration (StyleGAN3-R), was utilised with a batch size of 8 and an R1 regularization weight (gamma) of 0.5. All other training parameters were set at default values.

   The model used in **Article #4** was conditional to allow for the generation of the different classes of urothelial images (benign, atypical and malignant), while the model used in **Article #3** was trained in an unconditional manner.

   As discussed below, both models made use of Adaptive Discriminator Augmentation (ADA) which is the default mode in StyleGAN3.

   The objective evaluation metric utilised while training our StyleGAN3 models was the Fréchet inception distance (FID), a commonly used metric introduced in 2017. The FID is a robust metric that is sensitive to noise and, specifically, intraclass mode collapse[21,22] (as discussed in detail in **Article #3**).

## 4.4 SYNTHESIS AND DISCUSSION

This section of the thesis includes an integration and synthesis of our end-to-end machine learning project and its associated publications. It includes an overview of each of the published articles, the methodologies used and the results obtained in each research article, and a supplementary literature commentary on the topic. This integrated narrative begins with a brief discussion of the epidemiology of bladder cancer and The Paris System for Reporting Urine Cytology. The section then expands on machine learning in urine cytology followed by a discussion relating to the utility of unsupervised machine learning in anatomical pathology – including an exploration of GANs. Following this, data set creation and annotation are discussed. Finally, the generation of synthetic urine cytology images is presented, culminating in the testing of these images by pathology personnel.

Bladder cancer is the 10$^{th}$ most common cancer reported worldwide[23] and, in South Africa specifically, bladder cancer is the 9$^{th}$ most common cancer detected in men and the 17$^{th}$ most common cancer among women[24]. Urine cytology offers a rapid and cost-effective screening tool for the initial workup of high-risk patients with haematuria[25]. However, issues with sensitivity, diagnostic accuracy, and high interobserver variability present challenges in this domain of cytology. The Paris System (TPS) for reporting urine cytology, introduced in 2016[25] and updated in 2022[26], aims to address these concerns regarding urine cytology by defining morphological criteria for diagnostic categories and standardising the reporting of urine cytology to make it universally acceptable to the cytology community. In our local setting of a public sector pathology laboratory, most neoplastic urine cytology specimens submitted to our laboratory are squamous cell carcinomas of cervical origin. The result of this is that our trainees (both registrars and cytotechnologists) have limited exposure to urine cytology for urothelial neoplasia, similarly, creating sets of urine cytology for QA testing is also challenging. While the use of teaching collections has been the mainstay of how this lack of material has been addressed in the past, the development of generative machine learning models (such as GANs) offers an alternative solution to this problem by allowing for the generation of entirely synthetic images. While this is not the first time such algorithms have been applied to medical image generation, the direct application of this technology to cytopathology is, to the best of our knowledge, novel. Developments in GANs research have made the generation of high-resolution, believable images possible in recent years and several studies have alluded to the usefulness of GANs in medical education, however, several questions arise when applying these techniques to cytopathology and medical

education at large. Most studies on the topic make use of a simple *Turing Test* to assess the success of GANs at generating believable pathology images. A Turing Test, named after British mathematician Alan Turing, requires a human observer (e.g., a pathologist) to determine whether an image is either real or synthetic. While this is useful in certain settings, a more detailed analysis of the suitability of synthetic images for pathology educational purposes is required to validate this approach in a "real-world" setting. The motivation behind this body of work was to use GANs technology to address the paucity of teaching/QA material relating to urine cytology in our local setting and to test this data on pathology practitioners. What is presented in this body of work is an end-to-end machine learning project that begins with dataset creation and culminates with the testing of synthetic images, created by a custom-trained GANs, on pathology personnel to determine their believability and utility.

*The Paris System (TPS) for reporting urine cytology*

As mentioned, TPS was introduced to standardise the reporting of urine cytology and prioritises the detection of high-grade urothelial carcinoma on urine cytology.

TPS, in its initial 2016 form[25], has seven diagnostic categories: nondiagnostic/unsatisfactory, negative for high-grade urothelial carcinoma (NHGUC), atypical urothelial carcinoma (AUC), suspicious for high-grade urothelial carcinoma (SHGUC), high-grade urothelial carcinoma (HGUC), low-grade urothelial neoplasm (LGUN), and other. A recently published update on TPS[26] (termed TPS 2.0) presents some changes to the original TPS, but the basic morphological criteria used in the initial iteration of TPS and our work, remain unaffected to any significant degree, as do criteria for specimen adequacy. TPS 2.0 retains the initial goal of TPS, to detect high-grade urothelial carcinoma. Notable updates to this system include:

i. Incorporation of the LGUN category in the NHGUC category – based on the principle that TPS is designed to detect high-grade lesions.
ii. The de-emphasis of specific N/C ratios, although these cut-offs remain unaltered, and use the terms "increased" for N/C ≥ 0.5 and "high" for N/C ≥ 0.7. Importantly, N/C remains the starting point for morphological evaluation.
iii. Introduction of criteria for the assessment of upper urinary tract cytology using strict morphological criteria.

iv. Incorporation of an Atypical Squamous Cells (ASC) category to address squamous lesions in urine cytology.

v. Further clarification on dealing with degenerative cellular changes in urine cytology specimens.

Our dataset was created using the original TPS criteria which, as stated above, have been carried through to TPS 2.0. In our project, images that contained benign elements only were assigned to the *Benign* class (corresponding with the NHGUC category), images containing mildly atypical urothelial cells (defined as N/C ratio greater than 0.5 and one or more of nuclear hyperchromasia, irregular nuclear contours or irregular coarse chromatin) were assigned to the *Atypical* class (corresponding to the AUC category) and images containing malignant urothelial cells (defined as N/C > 0.7 with hyperchromasia and nuclear irregularity and/or coarse chromatin) were classified as *Malignant* (combined SHGUC and HGUC categories). A visual summary of TPS criteria used in this work is presented in Figure 2. This illustration was also presented to image participants taking part in the research presented in **Article #4** as a visual aid.

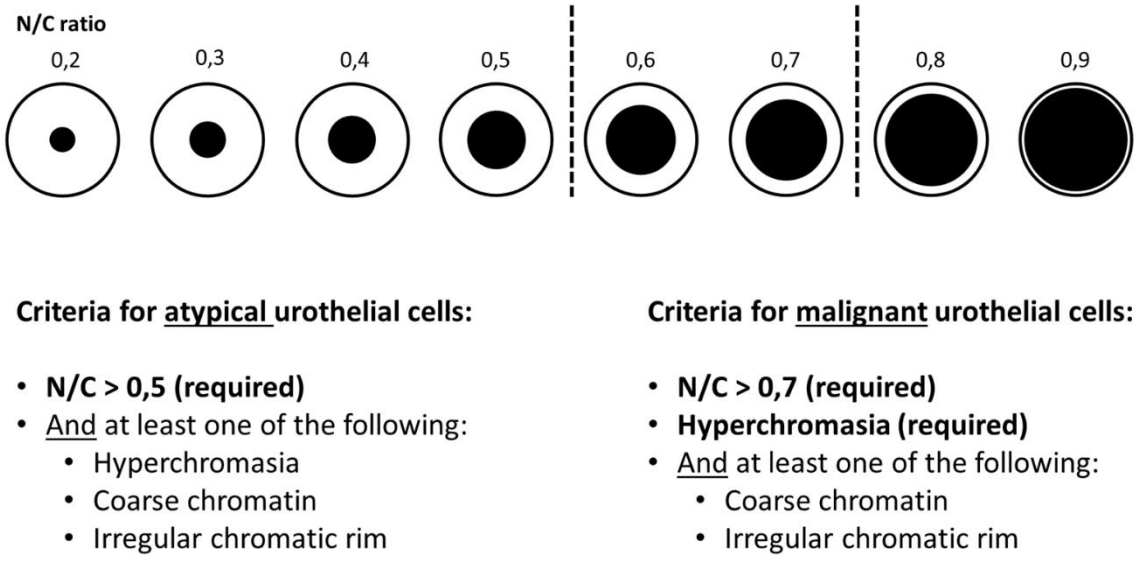



Figure 2: A visual summary of the TPS criteria applied in this work. From McAlpine et al[13], used with permission.

While challenges such as intra- and inter-observer variability remain, TPS has demonstrated improvements in specificity, positive predictive value, and diagnostic accuracy in this branch of cytology[27] – specifically in the negative for high-grade urothelial carcinoma (NHGUC) and

high-grade urothelial carcinoma (HGUC) categories[28]. The intermediate diagnostic category of atypical urothelial cells (AUC) remains a challenge when it comes to reproducibility[28–30] – a fact supported by our own work[12]. In **Article #2**, we demonstrate higher agreement (determined by kappa coefficient) in benign and malignant urothelial categories compared to the atypical one[12].

*Ancillary techniques in urine cytology*

While urine cytology remains a commonly used screening test, various ancillary investigations have been developed to supplement it. Bladder cancer poses an enormous financial burden on healthcare resources due to the need for lifelong follow-up of patients given high rates of recurrence and development of new primary cancers[31]. Both slide-based and slide-free ancillary methods have been proposed and include FISH-based techniques such as UroVision® - a test to detect homozygous deletion of chromosome 9p21. Other proposed techniques include both laboratory-based tests and point-of-care tests for protein products, immunohistochemical panels and, more recently, mutation analysis and next-generation sequencing[31]. While these techniques are likely to become available in more affluent settings – including the private sector in South Africa – for low-income settings across the African continent, urine cytology is likely to play a significant role in patient screening for a number of years to come. This will require that pathology personnel are adequately trained to accurately classify these specimens in the future. This, in turn, will require the exposure of such personnel to adequate and varied urine cytology specimens during their initial training and ongoing professional development. While we do not propose that synthetic images will replace real-world samples in this process, we hope that such images will augment teaching and training material going forward.

*Machine learning in urine cytology*

Supervised learning techniques such as object detection and image classification algorithms hold immense promise for daily practice in cytopathology. The hope is that such algorithms will lessen the increasing diagnostic burden placed on cytology personnel due to increased workloads and diminishing practitioner numbers[6–8].

Research into the use of supervised machine learning in urine cytopathology is ongoing and several publications support its utility in this branch of pathology. In a study by Sanghvi et al[32], the authors developed a workflow that used automated image analysis to predict TPS categories in urine cytology WSIs. This study made use of multiple convolutional neural networks to predict N/C ratio, the presence of nuclear hyperchromasia, coarse chromatin

and nuclear irregularity, cell degradation and overall risk of malignancy. They combined these cell-level features with slide-level features (quantity of urothelial cells, number of atypical cells and number of cell clusters) to train a classifier to predict a TPS category. The authors achieved a ROC AUC of 0.88, a sensitivity of almost 80% and a specificity of almost 85% for detecting HGUC. Replicating an existing cytology classification system, as was done in this case, has the added benefit of allowing for better interpretability by cytologists.

Similarly, Nojima et al[33], developed a deep learning algorithm aimed at detecting HGUC (the main aim of TPS) and achieved an ROC AUC of approximately 0.99. Furthermore, the authors demonstrated the ability of their deep learning system to predict the presence of stromal invasion on corresponding tissue biopsies – something not considered possible with cytology specimens. Using gradient-weighted class activation mapping (GradCAM) to visualise how their algorithm predicted stromal invasion, the authors indicate that “nuclear colour” (presumably equating to staining pattern) predicted the presence or absence of invasive malignancy.

Kaneko et al[34], showed a similarly high ROC AUC with their deep learning model (average AUC of 0.99). In addition, their model achieved higher accuracy than three cytopathologists. Importantly, this study used urine specimens from patients with confirmed bladder cancer (i.e., not in a standard screening setting). Additionally, while urothelial cells were labelled as either benign, atypical or malignant, the authors trained a binary classifier that designated cases as either negative (only benign cells) or positive (atypical and malignant cells), rather than trying to replicate TPS as done by Sanghvi et al[32].

Very recently, Ou et al[35], demonstrated the use of AI-assisted diagnostics in urine cytology by creating a deep learning model that identified atypical urothelial cells to assist cytology personnel in categorising urine cytology WSI’s in terms of TPS. Their algorithm identified cells with “low-risk” atypia (N/C 0.5-0.7) and “high-risk” atypia (N/C > 0.7) and presented the algorithm's findings to human observers, resulting in a panel of two cytotechnologists and a cytopathologist performing close to the level of an expert urine cytology panel.

Interestingly, a proprietary system called VisioCyst® (VitaDX International, Rennes, France), combines both fluorescence imaging and AI to improve the detection of urothelial carcinoma – spanning the use of both ancillary methods, as presented above, and machine learning. A multicentre study to validate this system showed greatly improved sensitivity for the detection of both low- and high-grade tumours using this system compared to conventional cytology[36].

Of particular note is that research into machine learning in urine cytology (and cytology in general) has focused on supervised tasks such as classification/diagnosis. This body of work is the first to explore the utility of GANs to generate synthetic cytology images to augment teaching and training.

### Article #1 – The Utility of Unsupervised Machine Learning in Anatomic Pathology[11]

This review focuses on three unsupervised machine learning methods and how these have mainly been used to improve supervised algorithms in pathology. This discussion will focus specifically on two of these three categories: clustering and GANs – both used extensively in this work.

*Generative Adversarial Networks*

Generative Adversarial Networks are a group of machine learning models that produce synthetic versions of the data on which they are trained. In general, they comprise two networks – a generator model and a discriminator model that compete during training. The generator creates synthetic data while the discriminator is presented with examples of both real synthetic images and attempts to detect the fake data. Figure 3, originally from **Article #3**, illustrates the GANs training process where the discriminator attempts to classify an image as either real or synthetic. The classification error of the discriminator is propagated back to both the discriminator and the generator, with the generator attempting to maximise the discriminator's error and the discriminator attempting to minimise its own error[22].

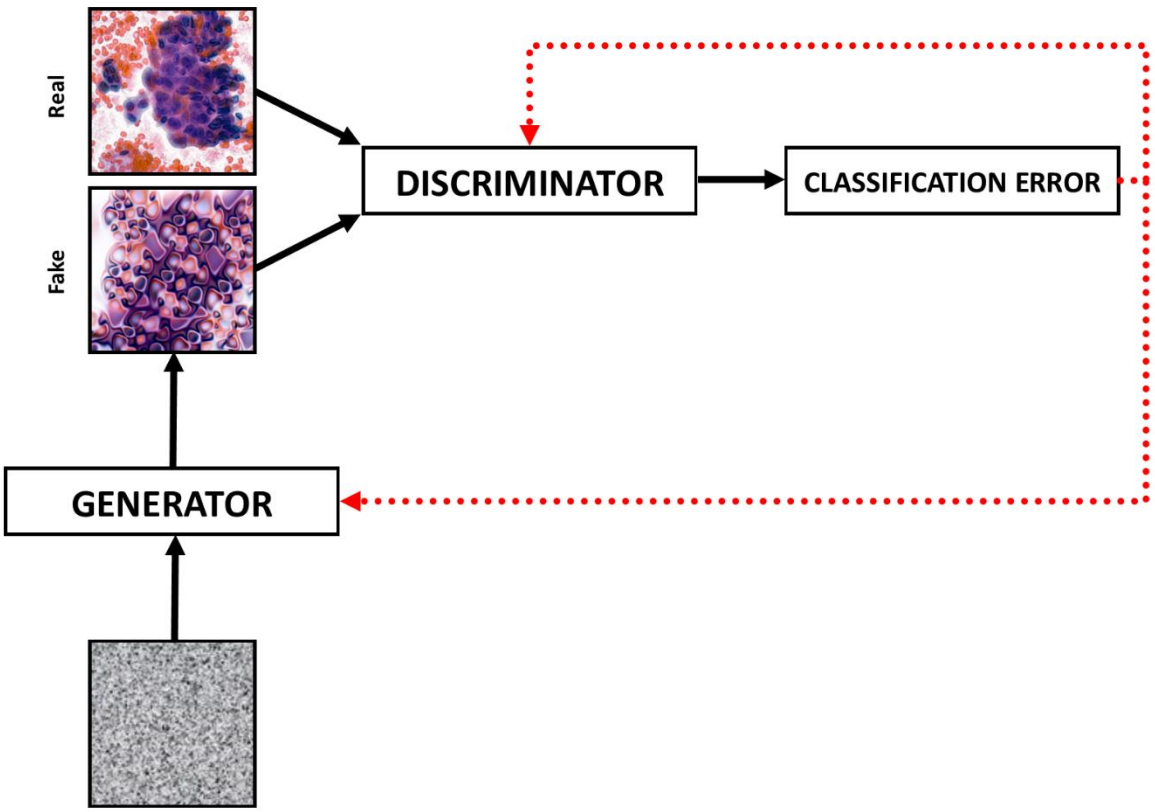



Figure 3: An illustration of a GANs model, comprising a Generator that learns to generate a synthetic image from random noise, and a Discriminator that learns to distinguish these synthetic images from real images. From McAlpine et al[10], used with permission.

To date, the use of GANs in pathology has revolved around augmenting data to improve the performance of supervised machine learning algorithms. Supervised algorithms are sensitive to the variation in staining seen across different laboratories[37–40]. Furthermore, colour variation is created by different whole slide scanners[37]. GANs have emerged as an important method of colour/stain normalisation as a pre-processing step in supervised algorithm training[37–40]. Additionally, creating synthetic histochemical and immunohistochemical stains from an existing stain (termed stain transfer) has been proposed in the literature – both to replace physical staining in a diagnostic setting[41,42] and to improve supervised learning algorithms[43].

Creating synthetic pathology images using GANs has also focussed mainly on improving the performance and generalisability of supervised algorithms[44–49] rather than the proposal in this corpus of work – utilising such images for teaching, training and QA – although, synthetic medical image generation has previously been reported in the fields of

ophthalmology and radiology[50–52]. Beers et al[52], used GANs (specifically PGGAN, discussed in further detail below) to generate synthetic fundoscopy and glioma magnetic resonance images (MRIs). Fundoscopy images were generated at a resolution of 512x512px, while the MRIs were generated at a lower resolution of 256x256px. The authors of this work indicate that, while some unrealistic images were created, the PGGAN was able to generate believable and diverse images. Artefact was seen, however, including overly distinct edges and a checker-board artefact in fundoscopy images. More recently, Segal et al[51], presented a demonstration of the development of a multi-class GANs, also using the PGGAN architecture, to create synthetic chest radiographs (X-rays) at a high resolution of 1024x1024px. Furthermore, the authors conducted a Turing Test to evaluate the ability of radiology staff (both radiologists and radiology registrars) to detect synthetic images. Real images were identified correctly 73% of the time and synthetic images were incorrectly identified as real 61% of the time. Of note is the fact that, although images were generated at a resolution of 1024x1024px, images were presented to respondents at a resolution of 260x260px[51].

Other more recent developments in the application of GANs to pathology include improving the speed of image acquisition in WSI scanning and GANs-based stain normalisation and improving the speed and accuracy of automatic mitoses counting in breast cancer. Fanous et al[53], demonstrated the ability of GANs to remove motion blur generated by moving the stage of the scanning microscopy rapidly during the image acquisition process – although, this was not tested on current commercial whole slide scanners but on a specific static microscope. Kang et al[54], improved the speed of stain transfer by first training a StainGAN[38] to normalise slides, using the trained generator to normalise images and then using these GANs-altered images to train a stain normalisation deep learning model – demonstrating a 40-fold improvement in the speed of normalising a slide. MiNuGAN (Mitoses and Nuclear Segmentation GAN), developed by Razavi et al[55], uses a conditional GANs architecture to segment nuclei and mitoses simultaneously, producing better results than the traditional U-net architecture used for image segmentation.

**Article #1** also presents an explanation of CycleGAN and argues why this specific GANs architecture is of particular interest to anatomical pathology. CycleGAN allows for unpaired image-to-image transfer meaning that one does not require to have paired images (e.g., real H&E and real Masson trichrome stains) to train the network. As discussed in the article, this particular model has been used for stain normalisation and stain transfer in the past. An interesting development by Levy et al[41], demonstrated the use of CycleGAN to create virtual

Masson trichrome stains from real H&E liver sections – validating their algorithm in a real-world clinical setting. More recently, Naglah et al[42], have produced a similar technique for virtual Masson trichrome generation, this time using paired data to train a conditional GANs rather than a CycleGAN, and suggest that their methodology produces more optimal results.

While the development of GANs by Goodfellow et al[15] in 2014 was undeniably revolutionary in the field of AI and machine learning, GANs initially suffered from important limitations. Briefly, preliminary GANs models were limited to low-resolution, small images and required very large datasets (from several hundred thousand to several million images) to train, could only generate a limited variety of synthetic images and required enormous computing resources to train[21,22,56]. A further challenge with GANs is that the generator can start to produce only a limited variety of believable examples – termed *mode collapse*[22]. These constraints understandably impeded the utility of this technology in a medical setting.

For the objective of this research to be met, we required our GANs model to:

i. Generate high-resolution images.
ii. Train on a modest dataset.
iii. Train on reasonable computing resources.
iv. Replicate the morphological diversity inherent in urine cytology.

Fortuitously, research into addressing these limitations of early GANs has been ongoing since the initial GANs publication in 2014 and enormous progress has been made in this regard.

One attempt to increase the resolution of synthetic images is BigGAN[57], which is capable of generating images up to 512x512px with sufficient variability for medical image generation. This architecture is, however, severely limiting in a medical setting in that it requires computing resources far in excess of the reach of most pathology departments worldwide, requiring specialised Tensor Processing Unit Pods to train due to very large batch sizes.

The StyleGAN family of algorithms provides alternative solutions to these challenges in GANs training with several additional benefits for medical/pathology image generation. This series of GANs have been created by the GPU manufacturer, NVIDIA, and are capable of generating very large resolution images up to 1024x1024px. The first architecture in this series is the so-called ProGAN/Progressive GAN/Progressively Growing GAN (PGGAN)[56]. This architecture has been successfully used to generate synthetic medical images[50,51] and stabilises the GANs training process by progressively growing the network from 4x4px up to 1024x1024px images. It also introduced a method to improve the variability of the synthetic

images created – called *minibatch standard deviation*. The first successor of ProGAN is StyleGAN[58], which carried much of the developments of ProGAN forward, together with introducing a mapping network into the generator architecture. This allows not only for the generation of high-quality images but also allows for a degree of control over the "styles" generated (in the initial article, this is illustrated with variations in features in facial portraits). The first iteration of StyleGAN2[59] corrected several artefacts seen in the synthetic images generated by StyleGAN. As mentioned in **Article #3**, a development in the StyleGAN series called Adaptive Discriminator Augmentation (ADA) – first introduced in the second iteration of StyleGAN2[60] – allows for the generation of synthetic images with orders of magnitude less training data than previously required for GANs. This is of particular importance in medical image generation where access to large training sets is a distinct challenge. The iteration of StyleGAN utilised in this work is the latest in the series, StyleGAN3[61], which further improves the results of StyleGAN2-ADA – specifically in more realistic texturing (especially important in animation and video generation).

Another potential benefit of the StyleGAN family of algorithms in medical image generation that was unexplored in the present work is style mixing. Style mixing allows for the "styles" of two images to be mixed during the generation process – e.g., mixing facial features and the presence or absence of spectacles in human face generation as illustrated in the original StyleGAN publication[58]. However, as the creators of the StyleGAN family state, mixing regularisation can be detrimental to complex datasets, is highly dataset dependent and when regularisation was disabled during training, fewer unrealistic images were generated[61]. As creating realistic images was our primary goal, we used a very low regularisation weight during our training, limiting our model's ability to learn image attributes to allow for reliable style mixing. Furthermore, in its current form, StyleGAN requires that the GANs be trained in an unconditional manner to allow for style mixing. As we needed control over the classes generated by our GANs (i.e., benign, atypical or malignant examples), we trained our GANs in a conditional fashion. Further research into the use of style mixing, and the broader domain of style transfer, in generating pathology images seems an obvious next step.

As briefly alluded to, GANs can be trained in a conditional manner, allowing for control over the type of images generated. Traditionally, GANs have been regarded as an unsupervised machine learning technique, however, shortly after the development of GANs in 2014, the idea of conditional GANs (cGANs)[62] was proposed. This technique trains both the generator and the discriminator on class labels, allowing for target data generation by GANs – i.e.,

allowing a user to direct the generator to create an image of a specific class (benign, atypical or malignant in this case). Unlike ProGAN (in its original form), StyleGAN3 allows for this class-specific image generation, making it well-suited for medical image generation.

It is for the reasons set out above that StyleGAN3 was selected for this project and **Article #3** and **Article #4** present the development of our work using this algorithm.

The results of the StyleGAN3 output at different stages in the training process are illustrated in Figures 8 to 11. These figures are presented at the end of this discussion.

The potential use cases for GANs in pathology extend well past what has currently been achieved. Generating entirely synthetic images and transforming images (image-to-image translation) as done in stain transfer and stain normalisation are good initial pathology-specific uses for this technology but generative models offer uses beyond these applications. Tschuchnig et al[63], suggest extension of image-to-image translation to morphological translation (e.g., frozen section to formalin-fixed paraffin-embedded tissue sections or text-to-image translation – providing a GANs with a text input and obtaining a synthetic output image). An interesting extension of this would be the transformation of histology into cytology and vice versa, for teaching and training purposes. Another use case could be image-to-text translation – a histological/cytological microscopic description could be generated by a GANs given a visual input (e.g., a WSI) – this could be used to decrease reporting times for pathologists as pathology reports will be able to be generated prior to the case being examined and the generated report could be edited before sigh out.

More recently a new class of generative models have emerged in the text-to-image arena that warrants brief inclusion in this discussion. Diffusion models are a new class of deep learning models that generate synthetic images[64]. These generative models adopt a different approach to image synthesis: in a forward diffusion process, an input image is perturbed by the step-wise addition of gaussian noise, while in the reserve diffusion process, the model learns to recover the original image by reversing the diffusion process in a step-wise manner[64]. In this way, the model learns to generate images from random noise (the end result of the forward diffusion process). Diffusion models have been shown to produce high-quality, diverse synthetic images and the application of this technology to pathology images promises to be a very interesting future avenue for research. Two very successful examples of this technology in non-pathology text-to-image synthesis include DALL·E 2[65] and Stable Diffusion[66]. Importantly, however, is that these models currently require absolutely enormous datasets and computing resources to train – euphemistically termed

"computation burden"[64] – and GANs remain faster at image synthesis. Training the Stable Diffusion model, for example, is estimated to have cost $600,000[67] – clearly not within reach of most pathology departments worldwide.

As discussed in **Article #3**, using generative models to create fully synthetic patient records for teaching has enormous potential for medical education – with synthetic pathology data (e.g., WSIs) forming only one part of these records. Goncalves et al[68] and Walonoski et al[69], both illustrated the ability of generative models to create synthetic healthcare data – epidemiological data in the case of the former and in creating synthetic health records that simulate the lifespan of virtual patients in the case of the latter. More recently, Shi et al[70], made use of GANs technology to create a large synthetic dataset for predicting the effect of different hypertensive medications on blood pressure. Including synthetic pathology data (clinical pathology results, histology and cytology data) and synthetic radiology seems a logical progression of this. Ideally, creating fully synthetic pathology teaching cases would include patient demographic details, cytology specimens, histology specimens and associated ancillary investigations such as immunocytochemistry/immunohistochemistry and molecular tests. Clearly, this is an ideal that still requires an enormous amount of future research to achieve.

Finally, an interesting avenue for future research could be the evaluation of which models generate the best synthetic pathology images. A method for assessing synthetically generated images from various GANs has been developed by Zhou et al[71]. This is termed HYPE (Human eYe perceptual Evaluation). The proposed technique allows for the tracking of improvements between models. Specifically, HYPE∞, a simplified, faster and cheaper technique is particularly well suited to a pathology context, where pathologists with diagnostic time pressures will need to be involved in the evaluation process.

*Clustering*

Clustering, considered a part of unsupervised machine learning, refers to a group of algorithms that attempt to identify patterns in unlabelled data. We made use of clustering (specifically, K-means clustering with cosine similarity) extensively in our work – primarily as a form of diversity sampling[14]. In **Article #2**, we made use of clustering to initially remove a large proportion of unsuitable images (e.g., images with obscuring inflammation or blood, artefacts such as cracked coverslips or images containing only crystals) prior to labelling. In both **Articles #3 and #4**, we used clustering to balance the training dataset on a

morphological basis in an attempt to increase the diversity of synthetic images created by our GANs. Additionally, we used clustering to select a morphologically diverse set of both real and synthetic images for assessment by pathology personnel in **Article #4**.

Clustering has been proposed as a means of decreasing the amount of data that needs to be labelled manually by domain experts. Peikari et al[72], presented a "cluster-then-label" semi-supervised approach to dataset annotation. In this proposal, the labels from a limited number of data points are propagated to unlabelled data points. In our work, cluster purity (a measure indicating to what degree clusters contain a single class) was 0.799 when assessed using our initially labelled dataset described in **Article #2** (unpublished data). The image features extracted by the pretrained Resnet50 model (not trained on cytology images), were rather used to group visually (and by extrapolation, morphologically) similar images together for manual assessment by a pathologist, rather than as a form of semi-supervised learning where labels are propagated to similar images.

### Article #2 – The Dynamics of Pathology Dataset Creation Using Urine Cytology as an Example[12]

**Article #2** presents an analysis of the creation of a modest urine cytology dataset. Dataset creation is a time-consuming and laborious process but is of fundamental importance for supervised machine learning in pathology. The importance of high-quality, accurate and morphologically diverse datasets cannot be understated in the development of machine learning in pathology. Such datasets should contain sufficient diversity to allow algorithms to learn variability in illumination, staining, focus and morphology, to enable algorithms to generalise well. Our article presents summarised guidelines relating to pathology dataset creation combined from a variety of sources (Table 1 from **Article #2**). Briefly, pathology datasets should:

i. Account for variability encountered in the real world
ii. Be of high quality
iii. QA should be considered in the annotation process
iv. Algorithms should be evaluated appropriately
v. Details about data acquisition and the annotation process should be published with the dataset.

This article's main goal was to highlight the challenges relating to pathology dataset creation and discuss some dynamics relating to this process, along with providing a synthesis of

dataset creation guidelines. Relating to the dynamics of dataset creation, we quantified the time taken to weakly annotate urine cytology images and assess whether this differs by diagnostic class. Additionally, we assessed intra-observer variability and whether the duration of labelling sessions affects the quality of annotation. To achieve these objectives, we labelled a modest urine cytology dataset, by consensus via video conference using custom-created software that allowed us to record the time taken to annotate each image and to randomly introduce QA images into the labelling process. Two experienced pathologists labelled images during 7 sessions over a 14-day period. This was conducted via video conference, rather than in person, due to the COVID-19 pandemic.

The median time taken to label each urine cytology image was 2.9 seconds but ranged from 1 to 42 seconds. We demonstrated a statistically significant difference in the time taken to label images by class/diagnostic category – assigning an image to the atypical urothelial category took the longest at 7.2 seconds, followed by the malignant urothelial category at 3.8 seconds and 2.9 seconds for the benign urothelial category. Figure 4 presents a bar chart of labelling time by diagnostic category (from **Article #2**).

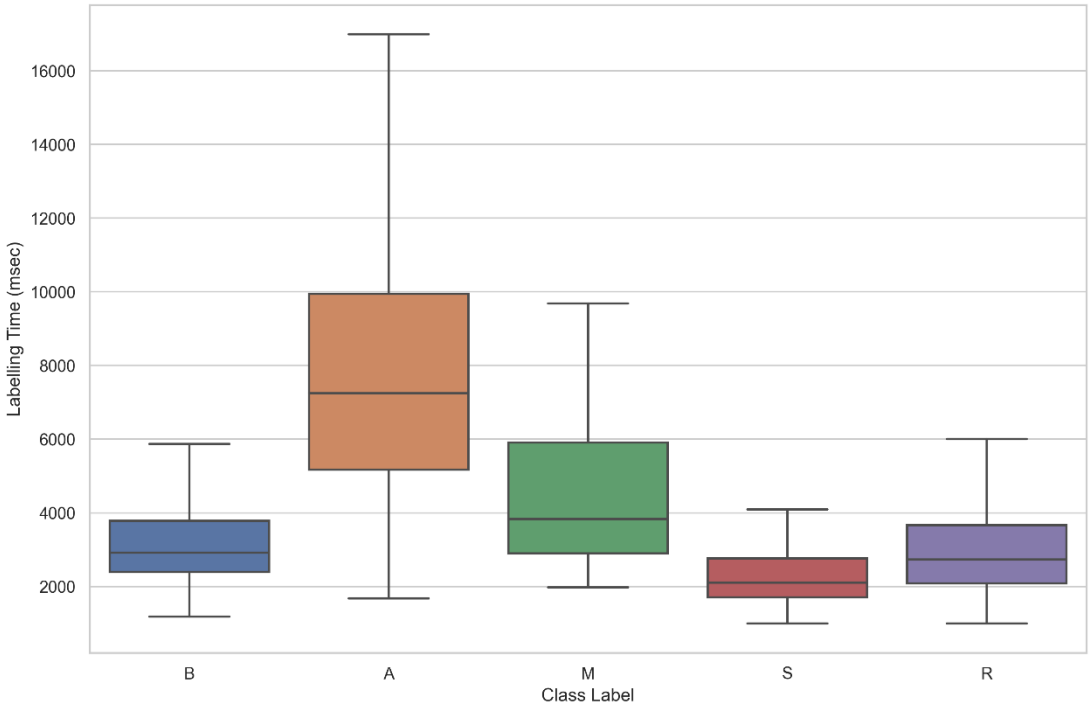



Figure 4: Image labelling time by diagnostic category. B, benign; A, atypical; M, malignant; S, squamous; R, reject. From McAlpine et al[12], used with permission.

Our QA agreement also varied among urothelial cell categories: benign and malignant categories showed almost perfect agreement while atypical diagnoses showed moderate agreement. A summary of this QA set is shown in Table 1 (from **Article #2**). Level of agreement is interpreted from Cuff et al[73].

Table 1: Summary of the QA dataset and associated level of agreement. From McAlpine et al[12], used with permission.

| **Urothelial cell category** | **Frequency (%)** | **Kappa score** | **Level of agreement** |
|---|---|---|---|
| **Benign** | 188 (77.05) | 0.82 | Almost perfect agreement |
| **Atypical** | 29 (11.89) | 0.49 | Moderate agreement |
| **Malignant** | 27 (11.07) | 0.81 | Almost perfect agreement |
| **Overall** | | 0.73 | Substantial agreement |

To evaluate pathologist performance over the length of the labelling sessions we assessed image labelling time and QA agreement in the first and second half of the sessions (based on median relative labelling time – additional explanation of this is presented in **Article #2**). Our data showed no evidence of a decrease in labelling time or diagnostic accuracy in the second half of the labelling sessions. On the contrary, our data suggest that image labelling time decreases in the second half of the labelling session, without diagnostic accuracy, as measured by kappa scores, showing deterioration. This point is illustrated in Figure 5 and Table 2 – both from **Article #2**.

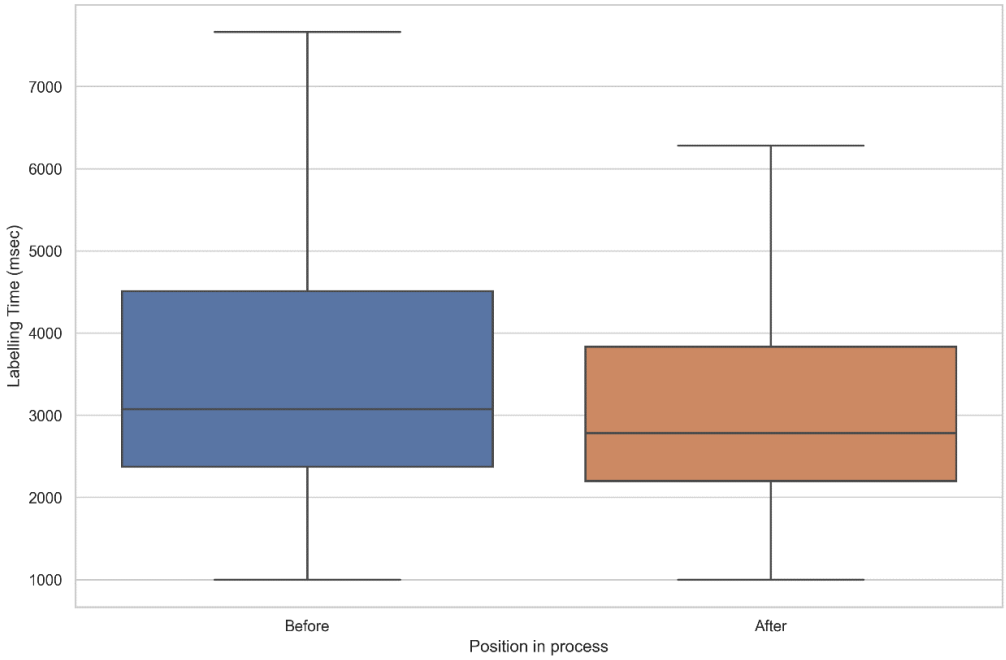



Figure 5: Image labelling time for images before and after the median relative label time. From McAlpine et al[12], used with permission.

Table 2: Summary of the level of agreement between urothelial categories before and after the median relative label time. From McAlpine et al[12], used with permission.

| **Category** | **Both images labelled before median *relative label time*** | **One image labelled before median *relative label time*** | **Both images labelled after median *relative label time*** |
|---|---|---|---|
| **Number of images** | 61 | 118 | 65 |
| **Kappa score** | 0.56 | 0.81 | 0.78 |
| **Level of agreement** | Moderate agreement | Almost perfect agreement | Substantial agreement |

Our data support the literature regarding the difficulties in the reproducibility of TPS specifically in the AUC category. As stated above, TPS was developed to specifically detect HGUC. The literature indicates that TPS has largely met this goal and has led to improvements in the specificity and positive predictive value in urine cytology[27,28].

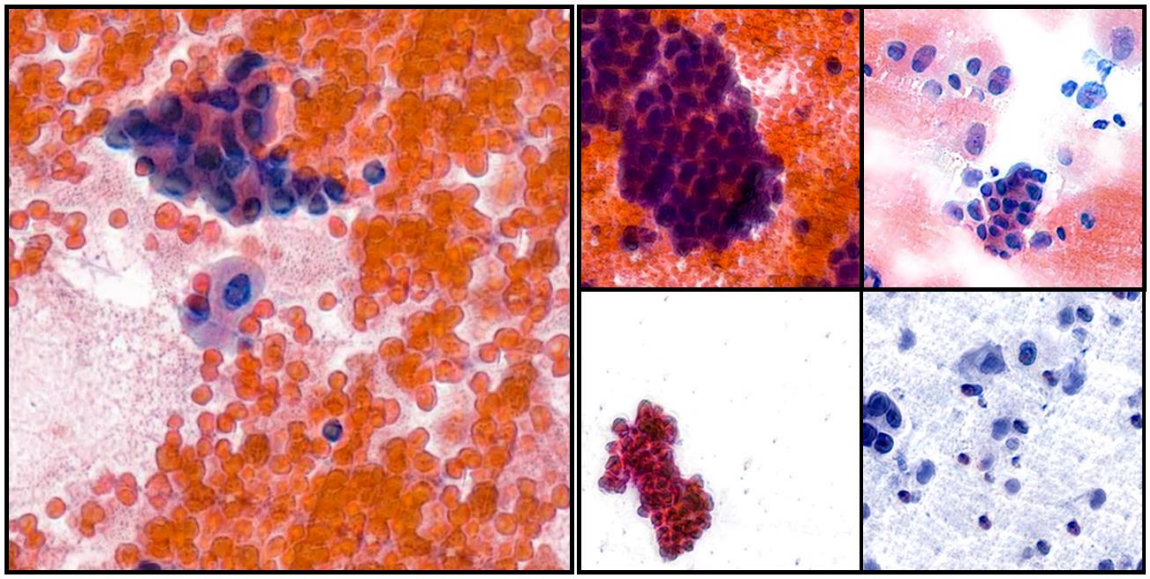
As indicated above, and in **Article #4**, the dataset ultimately used to train our StyleGAN3 model contained 7,460 images. This set of images was derived from a dataset including the 1,515 urothelial images labelled in **Article #2** and a second set of images 710 images annotated by two pathologists in tandem. An additional 11,521 images were labelled by the candidate to produce a dataset of 13,746 that underwent K-means clustering with cosine similarity to balance morphological diversity. QA was not performed at the initial annotation after the initial labelling experiment presented in **Article #2** as it was decided that QA would be performed on the generated synthetic dataset instead. As indicated in **Article #4**, both real and synthetic images utilised in the final experiment were visually assessed and accepted by two pathologists. K-means clustering with cosine similarity was used to improve morphological diversity in the final training set.

**Article #3: Is it real or not? Towards AI-based realistic synthetic cytology image generation to augment teaching and quality assurance in pathology**[10]

**Article #3** is a proof-of-concept article that illustrates the process of generating synthetic malignant urothelial images using StyleGAN3 using a small dataset and on modest hardware. This article presents the motivation for this project – the relative lack of malignant urine cytology specimens in our public sector setting. We demonstrate how synthetic images can be created with a small dataset using StyleGAN3 due to Adaptive Discriminator Augmentation (ADA). We also demonstrate that StyleGAN3 is able to generate a wide range of morphological patterns.

On the hardware available to us at the time (a 12Gb NVIDIA 3060 GPU) we were only able to generate 256x256px images. Between **Article #3** and **Article #4**, we obtained an NVIDIA Academic Hardware Grant and received a 24Gb RTX6000 GPU that allowed us to produce 512x512px images – overcoming one of the main limitations of this work.

Figure 6 illustrates a comparison between 256x256px and 512x512px images.



**Full resolution 512x512px image generated by the final StyleGAN3 model (Article #4)**

**Lower resolution 256x256px image generated by the initial StyleGAN3 model (Article #3)**

Figure 6: An illustration of the difference in image size/resolution between the images generated in Article #3 and Article #4.

Another key contribution of this work is a discussion on the use of synthetic data in healthcare and associated legal and ethical issues of this technology in medicine. Literature relating to the use of synthetic data in healthcare has been presented above under the discussion on GANs. Synthetic health data has been proposed as a solution to the risk of patient re-identification when traditional anonymisation approaches are used in health data research[74]. Legislation and ethical guidelines relating to the safe use of synthetic health data are lagging behind technological advancements in generative models[74,75]. While synthetic data has the potential to address the paucity of large, annotated datasets in medicine, it can be used maliciously. Chen et al[75], make the point that this technology has been used to generate what are known as “deepfakes” – fake impersonation videos that make use of face swapping technology. Ironically, the authors also point out that this technology could be used to protect patient anonymity by replacing real faces with synthetically generated ones in videos and images. Both Bellovin et al[74] and Chen et al[75], point out that unless specific, sophisticated modelling techniques are employed during training, there remains a risk of “data leakage” where some real data is leaked into the synthetic data – compromising patient anonymity. It is this problem that limits the use of such algorithms currently[75] and what lawmakers and ethical review boards should consider when addressing the use of synthetic data in healthcare[74,75]. Finally, Chen et al[75], make the point that the tedious and

time-consuming process of annotating data for training supervised algorithms may just be offset by the equally time-consuming and tedious task of curating synthetically generated data – shifting the problem to a different place in the machine learning cycle.

We used a dataset of 1,000 morphologically balanced malignant urothelial images to train a StyleGAN3 model to generate synthetic urine cytology images. This training process took 9-days and 21-hours to complete on an NVIDIA 3060 GPU. We present the results of this process by illustrating both highly realistic and less realistic synthetic images generated by our model. Unrealistic examples included images with unrealistic nuclear details (e.g., "wavy" artefacts), unrealistic staining patterns, acellular images, and unrealistic repetitive patterns in both the image background and cellular elements. Furthermore, we present morphologically diverse examples of these images. Variability in staining patterns, slide/image background and variations in cellular elements (e.g., cell clusters, individual cells, non-diagnostic benign cellular elements) were replicated by the GANs model.

To our knowledge, this is the first publication demonstrating the ability of GANs to generate highly realistic and morphologically diverse synthetic cytology images in an unconditional way (i.e., not relying on image-to-image transfer) – creating entirely novel synthetic cytology images.

**Article #4: Are synthetic cytology images ready for prime time? A comparative assessment of real and synthetic urine cytology images**[13]

The final article in this body of work presents two main contributions:

i. We extend our previous work of generating 256x256px malignant urine cytology images to 512x512px images of all three urine cytology classes – benign, atypical and malignant.
ii. We present the results of testing these synthetic images on pathology personnel.

Furthermore, we summarise the existing literature relating to the generation of synthetic pathology images and the assessment of these images by pathologists.

A dataset of 7,460 benign, atypical and malignant urine cytology images was used to train a conditional StyleGAN3 model. Once trained, the model was used to generate a synthetic dataset of 3,000 images per class (i.e., 9,000 in total). This synthetic dataset was clustered and a representative subset of 30 images was extracted for testing by pathology personnel. A morphologically balanced set of 60 images (30 real and 30 synthetic) was presented to

volunteers as an image survey. Participants were asked to place each image in a diagnostic category, provide a subjective quality score per image and state whether they would include the image in a pathology teaching set. Responses were assessed on an individual observer basis – i.e., responses for the real and synthetic subsets were compared per individual observer. The important overall outcome of this work is the finding that there was no statistically significant difference between real and synthetic urine cytology images when assessed for diagnostic error rate, subjective quality score or inclusion in a teaching set when assessed on an individual observer basis.

As the primary goal of this research project was to assess the ability of GANs to generate synthetic urine cytology images to supplement real-world examples, the above findings are of particular importance. Although the synthetic images required curation prior to their inclusion in the image survey, our trained StyleGAN3 model was able to generate realistic examples of all three urine cytology classes and at a high enough resolution for these images to be used for teaching purposes.

While Turing Tests are commonly used to determine whether human observers are able to distinguish real from synthetic, GANs-generated images, we opted to keep our participants blinded to the fact that our image survey contained synthetic images. The Turing Test approach to assessing the success of GANs in pathology has been adopted by various previous researchers[41,46,48,76]. Interestingly, a number of these research groups opted to include additional measures in addition to Turing Tests. Levy et al[41], focused on validating their synthetically generated Masson trichrome stains in liver biopsies in a real-world, clinical setting. Quiros et al[48], assessed stain quality subjectively in synthetically generated breast carcinoma images. In our work, we combined a subjective quality score, similar to a proposal by Platiša et al[77] for assessing the quality of digital pathology images, together with a more objective diagnostic error rate to compare real and synthetic images. As the primary goal of this research is to use synthetic images in teaching and training, we also assessed whether there is a difference in whether observers would include each type of image in a teaching set.

Our approach of assessing responses for real and synthetic images on an individual observer basis was influenced by approaches adopted in previous literature[41,46,48]. This approach was necessary for two reasons:

i. Respondents with varying pathology/cytology experience were included in the study; and

ii. Responses relating to image quality and inclusion in a teaching set were subjective in nature.

Comparing the individual observer's responses between real and synthetic images allowed us to test for a difference in the images of these two groups. As stated, there was no statistically significant difference between real and synthetic images in our study. The heatmap of p-values of individual observers presented in Figure 7, originally from **Article #4**, illustrates this.




Figure 7: A heatmap comparing the p-values for the 12 individual observers (rows 0 through 11), for error rates, inclusion in a teaching set and subjective image quality score. From McAlpine et al[13], used with permission.

An interesting pattern that has emerged from the literature is that the size of the images presented to pathologists plays an important role in their ability to successfully detect synthetic images – the larger the images, the easier it is for pathologists to identify the synthetic images. Levy et al[41], showed that zero out of four pathologists could detect synthetic images at 256x256px, at 512x512px one of the four pathologists could distinguish real from synthetic and at 1024x1024px, two of the four pathologists were able to make the distinction. We were able to generate realistically looking images at 512x512px (the original

resolution they were cropped at from WSIs). The study of synthetic pathology images to date has predominantly been on low-resolution images[44,46,48,78]. The question of optimal image resolution for synthetic images generated for teaching and training purposes will require further research but it is likely that 512x512px will be a minimum standard at this stage.

**Article #4** also presents an assessment of this technology in a larger cohort of image observers than in previous research. Both Levy et al[41] and Wei et al[46], used four pathologists, Senaras et al[76], three pathologists and Quiros et al[48], two pathologists. We recruited 12 participants, ranging from pathology trainees/registrars to pathologists with a wide range of cytology experience.

## DIGITAL PATHOLOGY AND SYNTHETIC DATA IN PATHOLOGY EDUCATION AND QUALITY ASSURANCE

Hassell et al[79], state that pathology training has traditionally followed a type of apprenticeship model where trainees are taught pathology by using real-world diagnostic cases in collaboration with an experienced pathologist. This teaching model makes use of glass slides, either derived from current diagnostic cases or recuts of archived material, to create sets of slides that a trainee would examine and diagnose in consultation with a pathologist. The digitisation of glass slides to create WSIs (i.e., digital pathology/DP) allows for digital replicas of these glass slides to be used for both diagnostic and educational purposes, however, the adoption of this technology in pathology education has been slow[79].

Similarly, proficiency testing as part of internal and external QA has traditionally made use of glass slides. Proficiency testing involves the practice of constructing a set of slides, circulating the set amongst pathologists and cytotechnologists, and comparing the answers of individuals against a set of targets[80]. Proficiency testing is an essential component of QA programmes and assists in ensuring that laboratories produce reliable test results and the continuous improvement of diagnostic pathology laboratores[80]. In recent years, the use of WSIs has become the standard for circulating QA slide sets for this purpose and such WSIs have been shown to be equivalent to glass slides for this purpose[81].

The use of DP and WSIs in QA and pathology education offers a number of potential advantages over the more traditional use of glass slides. DP improves the equity and standardisation of learning resources for medical students and pathology trainees and attempts to make teaching material available to trainees in remote and under-resourced

locations[81]. Moreover, DP allows for better standardisation of testing and assessment in pathology[82]. The use of WSIs is desirable in cases of low-volume material – such as cytology and small tissue biopsies – where it is difficult to replicate and standardise material for teaching and QA purposes[81]. Additionally, the use of such technology allows for improved statistical assessment because all participants are presented with identical diagnostic material[81].

The adoption of DP for diagnostics and educational purposes was rapidly accelerated by the COVID-19 pandemic which saw pathology laboratories in the United States adopting digital tools for routine diagnostics and to train residents when physical contact sessions were difficult to told[79]. Interestingly, this shift to distance learning with DP was viewed as inferior in terms of both quality and effectiveness by pathologists and trainees alike[79]. In contrast, evidence suggests that the use of WSIs for QA purposes is acceptable[83]. The perception of DP as an inferior teaching tool in Hassell et al[79] may, however, be due to the social distancing requirements of COVID-19, rather than the use of digital teaching material itself. The assessment of the acceptability of digital teaching material rather than glass slides in physical contact sessions offers further research opportunities in the developing field of DP in pathology education and training.

The purpose of this body of work was to attempt a proof of concept illustrating the next logical step in the process of acquiring and collating teaching and QA material in pathology. The traditional paradigm of the use of physical glass slides has been followed by the use of DP and WSIs and now, the use of AI to produce synthetic teaching and QA material. This is especially useful in the setting illustrated in this work when real-world teaching material is difficult to obtain due to scarcity. This approach also allows for the use of realistic synthetic data in place of real-world patient samples. As discussed in **Article #3** and in the discussion above, the use of synthetic health data is preferred in teaching in an attempt to better preserve patient anonymity.

While our work makes a significant contribution to the realisation of this goal, a number of limitations remain. Principally, the next step would be to create entire synthetic WSIs rather than the static images produced in this work. Further control over the visual appearance of such synthetic data would also be desirable – namely, control over the cellularity of specimens, the appearance of the lesional material (e.g., individual cells or cell clusters), and the appearance of the cytological background (e.g., clean, inflamed or bloody). Directions for future research to allow for the incorporation of generative AI in pathology education are expanded on below.

## FUTURE DIRECTION OF GENERATIVE ARTIFICIAL INTELLIGENCE IN PATHOLOGY EDUCATION AND QUALITY ASSURANCE

As indicated above, while this body of work presents a novel use for GANs technology and synthetic data in pathology – namely, illustrating its potential use in teaching and QA as opposed to improving supervised machine learning algorithms as previous literature has done, our work has several limitations that offer avenues for future research.

The first is that we generated individual static images rather than entirely synthetic WSIs. To take full advantage of synthetic data in pathology training and QA, the creation of synthetic WSIs is optimal. The creation of synthetic WSIs has been performed in the past, specifically by Levy et al[41], however, these authors created a virtual special stain using image-to-image translation (CycleGAN[84]) rather than entirely synthetic, novel examples of WSIs. The creation of such synthetic WSIs would be a major step forward in the incorporation of synthetic data in pathology but would require significant computing resources to achieve. Of note, is the fact that creating synthetic cytology WSIs would also require the creation of simulated Z-staking, to allow viewers of such synthetic WSIs to focus up and down the focal plane as is done in real-world cytology, creating further challenges for researchers and additional storage and computing resources. Following on from this, the generation of additional stains (e.g., a Giemsa stain) and immunocytochemical stains in addition to the initial synthetic Papanicolaou stain would allow for the creation of an entirely synthetic teaching case that could be used in education and QA.

The use of image-to-image translation to create histology images from cytology images, and vice versa, using technology such as CycleGAN[84] offers a potential additional educational tool to assist in pathology training by assisting trainees in understanding cytology/histology correlations.

In **Article #3** and **Article #4**, we discuss that manual curation of synthetic images generated by our models remains essential – especially if this technology is utilised for education and QA purposes. A further avenue for future research would be to develop a system that diminishes the role of a pathologist in this curation process by automating it. This would allow for the real-time generation of synthetic data without the intervention of a pathologist prior to its incorporation in a teaching and QA set.

As alluded to above, more control over the morphological features of the synthetic data is desirable. This could include control over cytological background (e.g., clean, bloody,

mucinous or inflamed), additional cellular elements (e.g., benign, non-lesional cells) and the nature of the lesional cells (e.g., clusters or individual cells and morphological variation). This may be better achieved using diffusion models (as discussed above) – specifically, with text-to-image generation[64–66].

Achieving these additional goals of more completely incorporating synthetic data in pathology education and QA, faces two major challenges: the creation and annotation of the required pathology-specific datasets and the computing resources required to train the models. Overcoming these hurdles will, however, allow for a far richer training and assessment experience for pathologists and trainees alike, by incorporating generative AI in the training and QA workflows of pathology laboratories in the future. Further challenges face low- and middle-income countries when it comes to AI research: the lack of a routine DP workflow which creates a challenge in digitising pathology material for annotation and access to the computing resources required (including stable, uninterrupted electricity). Working around such challenges may require partnerships with private sector and international laboratories, as well as the use of cloud computing services and the use of academic grants to fund such research.

The work presented in this thesis takes one step forward in the incorporation of generative AI in pathology by shifting the paradigm from using generative AI to improve supervised machine learning algorithms to using realistic synthetic data to supplement real pathology data when such real-world material is difficult to obtain or is associated with ethical issues such as patient permission to allow their biopsy material to be used for teaching and learning. A number of avenues for future research are also presented, the achievement of which will allow for a richer and more complete educational experience using synthetic pathology data.

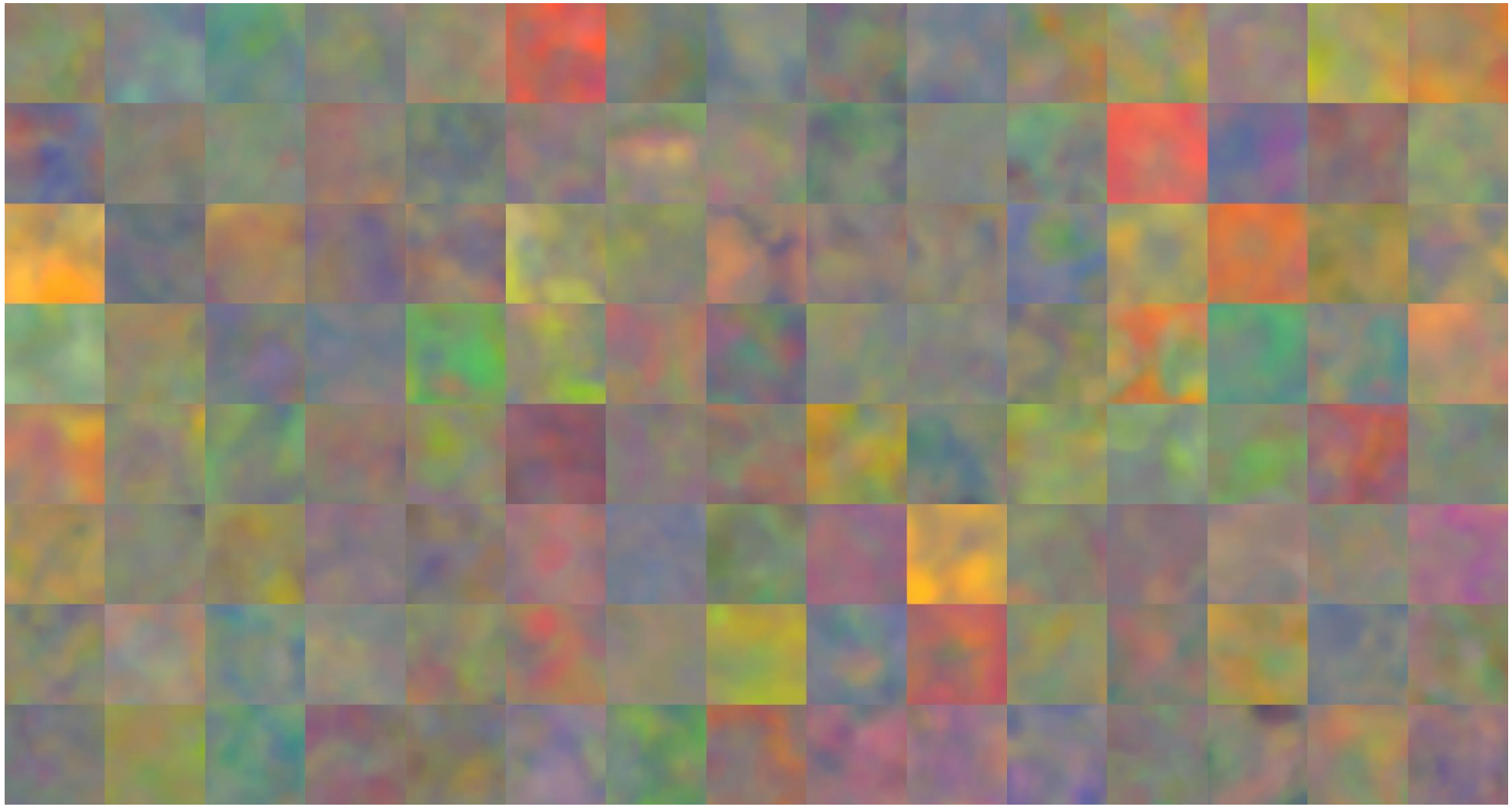


Figure 8:StyleGAN3 initiation results.

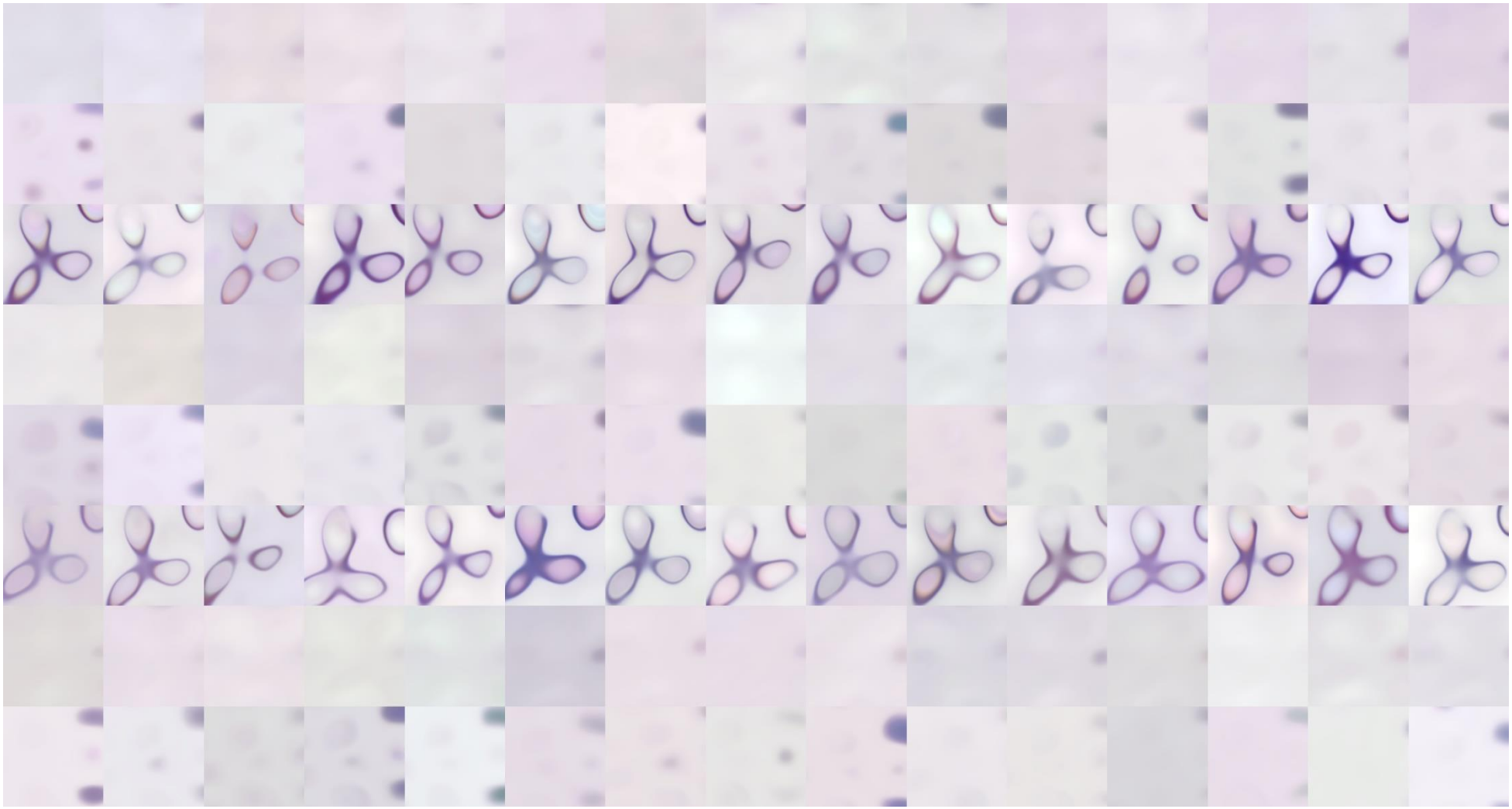


Figure 9: StyleGAN3 results at 20kimgs.

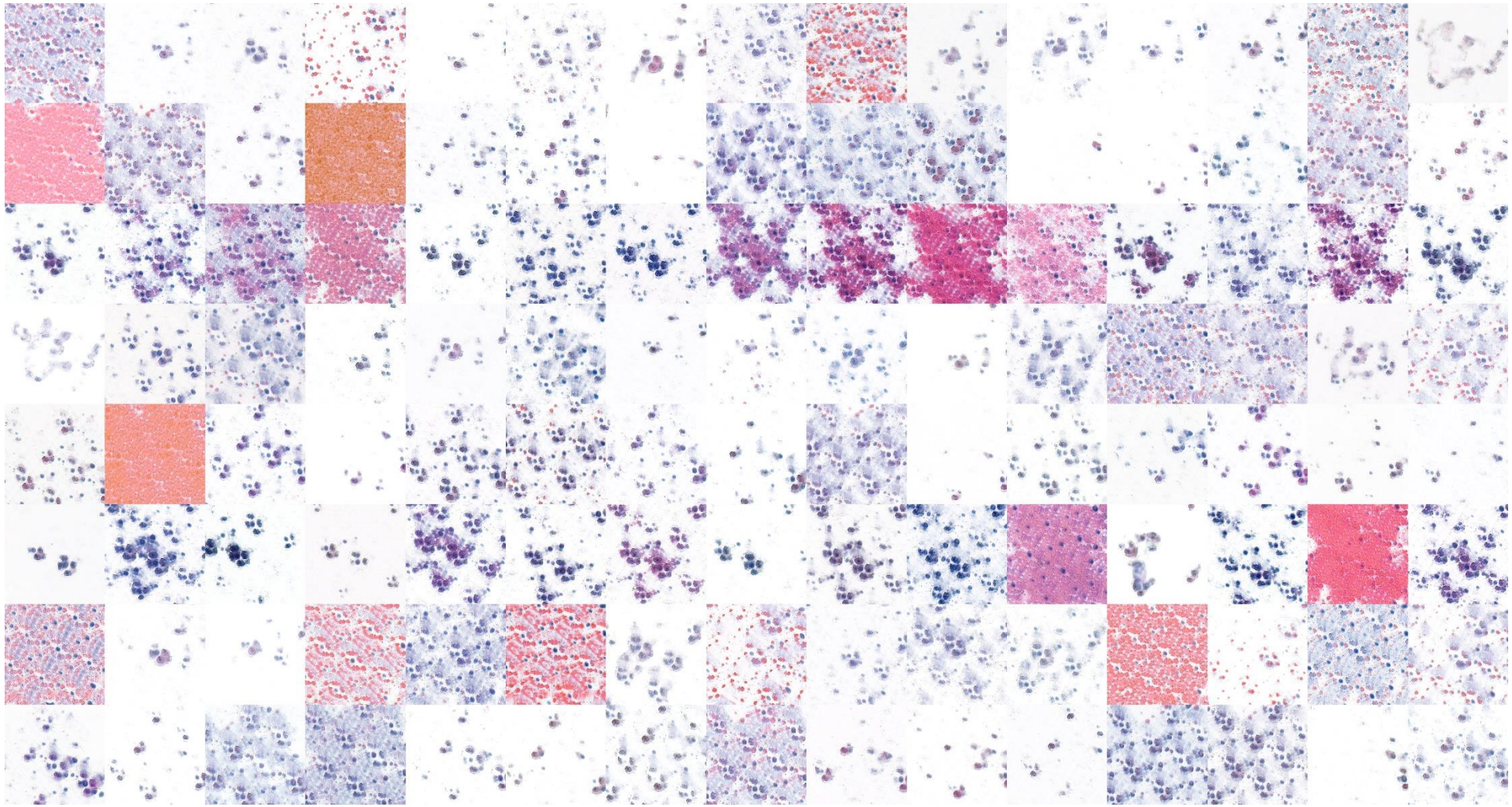


Figure 10: StyleGAN3 results at 100kimgs

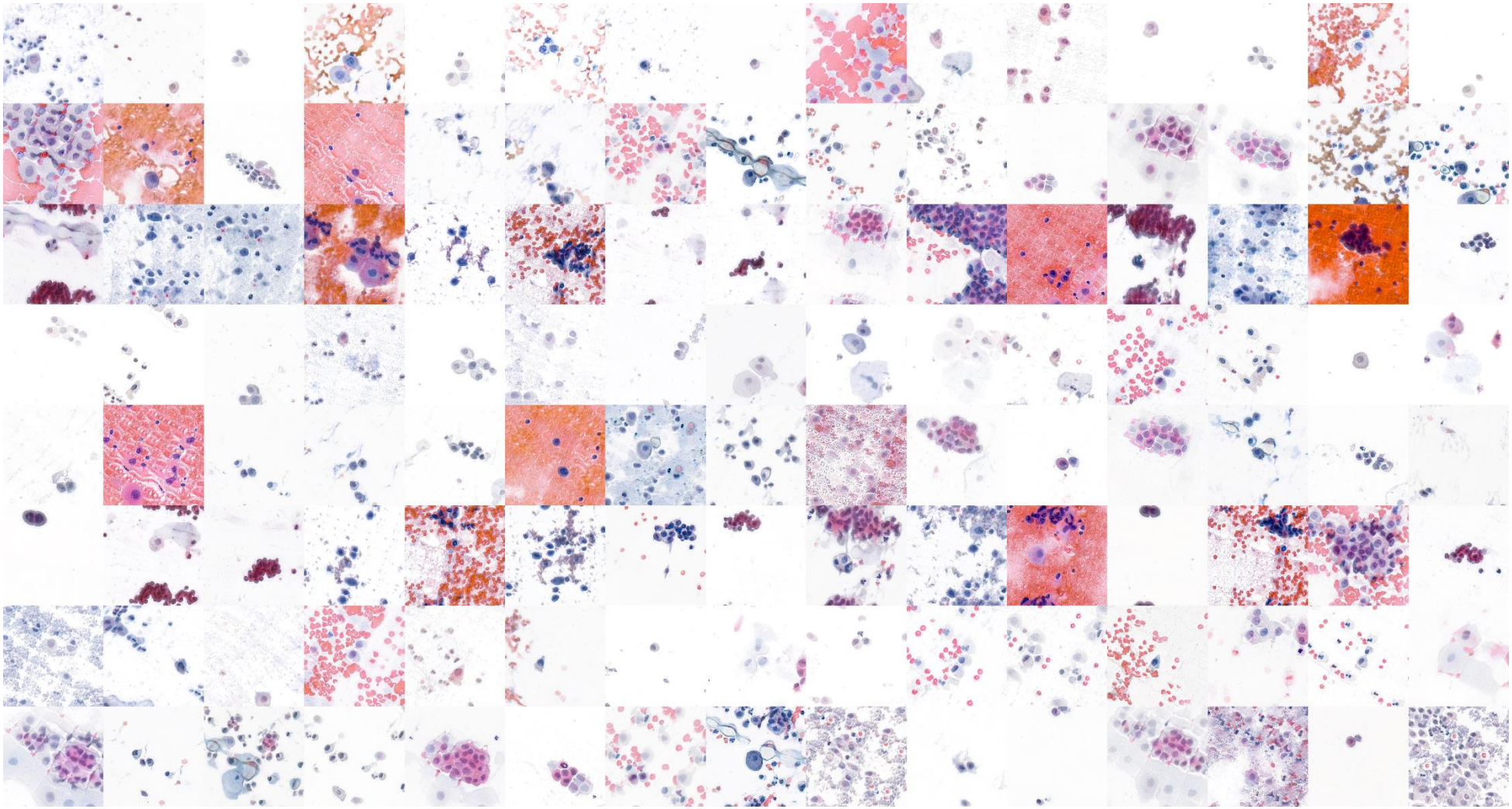


Figure 11: StyleGAN3 results at 3,410kimgs

# CHAPTER 5: REFERENCES


1 Salto-Tellez M, Maxwell P, Hamilton P. Artificial intelligence-the third revolution in pathology. *Histopathology* 2019; **74**: 372–376.

2 McAlpine ED, Pantanowitz L, Michelow PM. Challenges Developing Deep Learning Algorithms in Cytology. Acta Cytol. 2021; **65**: 301–309.

3 Chollet F. *Deep Learning with Python*. 1st ed. Manning Publications Co.: Greenwich, CT, USA, 2017.

4 Cui M, Zhang DY. Artificial intelligence and computational pathology. Laboratory Investigation. 2021; **101**: 412–422.

5 Hanna MG, Hanna MH. Current applications and challenges of artificial intelligence in pathology. *Human Pathology Reports* 2022; **27**. doi:10.1016/j.hpr.2022.300596.

6 Mudenda V, Malyangu E, Sayed S, Fleming K. Addressing the shortage of pathologists in Africa: Creation of a MMed Programme in Pathology in Zambia. *Afr J Lab Med* 2020; **9**: 2225–2002.

7 Metter DM, Colgan TJ, Leung ST, Timmons CF, Park JY. Trends in the US and Canadian Pathologist Workforces From 2007 to 2017. *JAMA Netw Open* 2019; **2**: e194337.

8 Royal College of Pathologists of Australia. Australian Pathologist Workforce Study 2018. 2018https://www.rcpa.edu.au/getattachment/4a38b4f9-5f6a-45eb-8947-dfa072797685/APW.aspx.

9 McAlpine ED, Michelow P. The cytopathologist's role in developing and evaluating artificial intelligence in cytopathology practice. *Cytopathology* 2020; **31**: 385–392.

10 McAlpine E, Michelow P, Liebenberg E, Celik T. Is it real or not? Towards AI-based realistic synthetic cytology image generation to augment teaching and quality assurance in pathology. *J Am Soc Cytopathol* 2022. doi:https://doi.org/10.1016/j.jasc.2022.02.001.

11 McAlpine ED, Michelow P, Celik T. The Utility of Unsupervised Machine Learning in Anatomic Pathology. *Am J Clin Pathol* 2022; **157**: 5–14.

12 McAlpine ED, Michelow P, Celik T. The Dynamics of Pathology Dataset Creation Using Urine Cytology as an Example. *Acta Cytol* 2022; **66**: 46–54.

13 McAlpine E, Michelow P, Liebenberg E, Celik T. Are synthetic cytology images ready for prime time? A comparative assessment of real and synthetic urine cytology images. *J Am Soc Cytopathol* 2023; **12**: 126–135.

14 Monarch R. *Human-in-the-Loop Machine Learning*. Manning Publications Co.: Shelter Island, 2021.

15 Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S *et al.* Generative Adversarial Networks. *ArXiv* 2014. doi:10.1001/jamainternmed.2016.8245.

16 National Health Insurance briefing by Department of Health | PMG. https://pmg.org.za/committee-meeting/13268/ (accessed 1 Dec2022).

17 Detecto. https://detecto.readthedocs.io/en/latest/index.html (accessed 2 May2020).

18 McAlpine E, Michelow P. Implementing Deep Learning Algorithms in Anatomic Pathology Using Open-source Deep Learning Libraries. *Adv Anat Pathol* 2020; **27**: 260–268.

19 Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans Pattern Anal Mach Intell* 2017; **39**: 1137–1149.

20 Tzutalin. LabelImg. 2015.https://github.com/tzutalin/labelImg.

21 Brownlee J. *Generative Adversarial Networks with Python*. v1.4. 2019https://machinelearningmastery.com/.

22 Langr J, Bok V. *GANs in Action*. Manning Publications Co.: Shelter Istand, 2019.

23 Bladder cancer statistics | World Cancer Research Fund International. https://www.wcrf.org/cancer-trends/bladder-cancer-statistics/ (accessed 4 Dec2022).

24 NICD, National Cancer Registry 2014. http://www.nicd.ac.za/wp-content/uploads/2017/03/2014-NCR-tables-1.pdf (accessed 2 Mar2020).

25 Barkan GA, Wojcik EM, Nayar R, Savic-Prince S, Quek ML, Kurtycz DFI *et al.* The Paris System for Reporting Urinary Cytology: The Quest to Develop a Standardized Terminology. *Acta Cytol* 2016; **60**: 185–197.

26 Wojcik EM, Kurtycz DFI, Rosenthal DL. We'll always have Paris The Paris System for Reporting Urinary Cytology 2022. *J Am Soc Cytopathol* 2022; **11**: 62–66.

27 Stanzione N, Ahmed T, Fung PC, Cai D, Lu DY, Sumida LC *et al.* The continual impact of the Paris System on urine cytology, a 3-year experience. *Cytopathology* 2020; **31**: 35–40.

28 Wang YH, Hang JF, Wen CH, Liao KC, Lee WY, Lai CR. Diagnostic agreement for high-grade urothelial cell carcinoma in atypical urine cytology: A nationwide survey reveals a tendency for overestimation in specimens with an N/C ratio approaching 0.5. *Cancers (Basel)* 2020; **12**. doi:10.3390/cancers12020272.

29 Long T, Layfield L, Esebua M, Frazier S, Giorgadze D, Schmidt R. Interobserver reproducibility of the Paris system for reporting urinary cytology. *Cytojournal* 2017; **14**. doi:10.4103/cytojournal.cytojournal_12_17.

30 Kurtycz DFI, Barkan GA, Pavelec DM, Rosenthal DL, Wojcik EM, VandenBussche CJ *et al.* Paris Interobserver Reproducibility Study (PIRST). *J Am Soc Cytopathol* 2018; **7**: 174–184.

31 Allison DB, Vandenbussche CJ. A Review of Urine Ancillary Tests in the Era of the Paris System. Acta Cytol. 2020; **64**: 182–192.

32 Sanghvi AB, Allen EZ, Callenberg KM, Pantanowitz L. Performance of an artificial intelligence algorithm for reporting urine cytopathology. *Cancer Cytopathol* 2019; **127**: 658–666.

33 Nojima S, Terayama K, Shimoura S, Hijiki S, Nonomura N, Morii E *et al.* A deep learning system to diagnose the malignant potential of urothelial carcinoma cells in cytology specimens. *Cancer Cytopathol* 2021; **129**: 984–995.

34 Kaneko M, Tsuji K, Masuda K, Ueno K, Henmi K, Nakagawa S *et al.* Urine cell image recognition using a deep-learning model for an automated slide evaluation system. *BJU Int* 2022; **130**: 235–243.

35 Ou YC, Tsao TY, Chang MC, Lin YS, Yang WL, Hang JF *et al.* Evaluation of an artificial intelligence algorithm for assisting the Paris System in reporting urinary cytology: A pilot study. *Cancer Cytopathol* 2022. doi:10.1002/CNCY.22615.

36 Lebret T, Pignot G, Colombel M, Guy L, Rebillard X, Savareux L *et al.* Artificial intelligence to improve cytology performances in bladder carcinoma detection: results of the VisioCyt test. *BJU Int* 2022; **129**: 356–363.

37 de Bel T, Hermsen M, Jesper Kers R, van der Laak J, Litjens G. Stain-Transforming Cycle-Consistent Generative Adversarial Networks for Improved Segmentation of Renal Histopathology. *Proc Mach Learn Res* 2019; **102**: 151–163.

38 Shaban MT, Baur C, Navab N, Albarqouni S. StainGAN: Stain style transfer for digital histological images. *Proceedings - International Symposium on Biomedical Imaging* 2019; **2019-April**: 953–956.

39 BenTaieb A, Hamarneh G. Adversarial Stain Transfer for Histopathology Image Analysis. *IEEE Trans Med Imaging* 2018; **37**: 792–802.

40 Ren J, Hacihaliloglu I, Singer EA, Foran DJ, Qi X. Adversarial domain adaptation for classification of prostate histopathology whole-slide images. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 2018; **11071 LNCS**: 201–209.

41 Levy JJ, Azizgolshani N, Andersen MJ, Suriawinata A, Liu X, Lisovsky M *et al.* A large-scale internal validation study of unsupervised virtual trichrome staining technologies on nonalcoholic steatohepatitis liver biopsies. *Modern Pathology* 2020; **34**: 808–822.

42 Naglah A, Khalifa F, El-Baz A, Gondim D. Conditional GANs based system for fibrosis detection and quantification in Hematoxylin and Eosin whole slide images. *Med Image Anal* 2022; **81**. doi:10.1016/j.media.2022.102537.

43 Gadermayr M, Appel V, Klinkhammer BM, Boor P, Merhof D. Which way round? A study on the performance of stain-translation for segmenting arbitrarily dyed histological images. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 2018; **11071 LNCS**: 165–173.

44 Hou L, Agarwal A, Samaras Di, Kurc TM, Gupta RR, Saltz JH. Robust histopathology image analysis: To label or to synthesize? *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 2019; **2019-June**: 8525–8534.

45 Mahmood F, Borders D, Chen RJ, Mckay GN, Salimian KJ, Baras A *et al.* Deep Adversarial Training for Multi-Organ Nuclei Segmentation in Histopathology Images. *IEEE Trans Med Imaging* 2020; **39**: 3257–3267.

46 Wei J, Suriawinata A, Vaickus L, Ren B, Liu X, Wei J *et al.* Generative image translation for data augmentation in colorectal histopathology images. *Proc Mach Learn Res* 2019; **116**: 10–24.

47 Bug D, Gräbel P, Feuerhake F, Oswald E, Schüler J, Merhof D. Supervised and Unsupervised Cell-Nuclei Detection in Immunohistology. *MICCAI 2019 Computational Pathology Workshop* 2019.

48 Quiros AC, Murray-Smith R, Yuan K. Pathology GAN: Learning deep representations of cancer tissue. In: *Proceedings of the Third Conference on Medical Imaging with Deep Learning*. 2020, pp 669–695.

49 Hou L, Agarwal A, Samaras D, Kurc TM, Gupta RR, Saltz JH. Unsupervised Histopathology Image Synthesis. 2017.http://arxiv.org/abs/1712.05021.

50 Bellemo V, Burlina P, Yong L, Wong TY, Ting DSW. Generative Adversarial Networks (GANs) for Retinal Fundus Image Synthesis. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 2019; **11367 LNCS**: 289–302.

51 Segal B, Rubin DM, Rubin G, Pantanowitz A. Evaluating the Clinical Realism of Synthetic Chest X-Rays Generated Using Progressively Growing GANs. *SN Comput Sci* 2021; **2**: 1–17.

52 Beers A, Brown J, Chang K, Campbell JP, Ostmo S, Chiang MF *et al.* High-resolution medical image synthesis using progressively grown generative adversarial networks. 2018.http://arxiv.org/abs/1805.03144.

53 Fanous MJ, Popescu G. GANscan: continuous scanning microscopy using deep learning deblurring. *Light Sci Appl* 2022; **11**. doi:10.1038/s41377-022-00952-z.

54 Kang H, Luo D, Feng W, Zeng S, Quan T, Hu J *et al.* StainNet: A Fast and Robust Stain Normalization Network. *Front Med (Lausanne)* 2021; **8**. doi:10.3389/fmed.2021.746307.

55 Razavi S, Khameneh FD, Nouri H, Androutsos D, Done SJ, Khademi A. MiNuGAN: Dual Segmentation of Mitoses and Nuclei Using Conditional GANs on Multi-center Breast H&E Images. *J Pathol Inform* 2022; **13**: 100002.

56 Karras T, Aila T, Laine S, Lehtinen J. Progressive Growing of GANs for Improved Quality, Stability, and Variation. *CoRR* 2017; **abs/1710.1**.http://arxiv.org/abs/1710.10196.

57 Brock A, Donahue J, Simonyan K. Large Scale GAN Training for High Fidelity Natural Image Synthesis. 2018.http://arxiv.org/abs/1809.11096.

58 Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 2019; **2019-June**: 4396–4405.

59 Karras T, Laine S, Aittala M, Hellsten J, Lehtinen J, Aila T. Analyzing and improving the image quality of stylegan. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2020, pp 8107–8116.

60 Karras T, Aittala M, Hellsten J, Laine S, Lehtinen J, Aila T. Training generative adversarial networks with limited data. *Adv Neural Inf Process Syst* 2020; **2020-Decem**.

61 Karras T, Aittala M, Laine S, Harkonen E, Hellsten J, Lehtinen J *et al.* Alias-Free Generative Adversarial Networks. 2021.https://arxiv.org/abs/2106.12423 (accessed 13 Dec2021).

62 Mirza M, Osindero S. Conditional Generative Adversarial Nets. 2014.http://arxiv.org/abs/1411.1784 (accessed 6 Sep2022).

63 Tschuchnig ME, Oostingh GJ, Gadermayr M. Generative Adversarial Networks in Digital Pathology: A Survey on Trends and Future Potential. Patterns. 2020; **1**. doi:10.1016/j.patter.2020.100089.

64 Croitoru F-A, Hondru V, Ionescu RT, Shah M. Diffusion Models in Vision: A Survey. 2022.http://arxiv.org/abs/2209.04747 (accessed 25 Oct2022).

65 Ramesh A, Dhariwal P, Nichol A, Chu C, Chen M. Hierarchical Text-Conditional Image Generation with CLIP Latents. 2022.http://arxiv.org/abs/2204.06125 (accessed 25 Oct2022).

66 Rombach R, Blattmann A, Lorenz D, Esser P, Ommer B. High-Resolution Image Synthesis with Latent Diffusion Models. 2021.http://arxiv.org/abs/2112.10752 (accessed 25 Oct2022).

67 Stable Diffusion - Wikipedia. https://en.wikipedia.org/wiki/Stable_Diffusion#cite_note-stable-diffusion-model-card-1-4-25 (accessed 25 Oct2022).

68 Goncalves A, Ray P, Soper B, Stevens J, Coyle L, Sales AP. Generation and evaluation of synthetic patient data. *BMC Med Res Methodol* 2020; **20**: 1–40.

69 Walonoski J, Kramer M, Nichols J, Quina A, Moesel C, Hall D *et al.* Synthea: An approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record. *Journal of the American Medical Informatics Association* 2018; **25**: 230–238.

70 Shi J, Wang D, Tesei G, Norgeot B. Generating high-fidelity privacy-conscious synthetic patient data for causal effect estimation with multiple treatments. *Front Artif Intell* 2022. doi:10.3389/frai.2022.918813.

71 Zhou S, Gordon ML, Krishna R, Narcomey A, Fei-Fei L, Bernstein MS. HYPE: A Benchmark for Human eYe Perceptual Evaluation of Generative Models. 2019.http://arxiv.org/abs/1904.01121 (accessed 4 Sep2022).

72 Peikari M, Salama S, Nofech-Mozes S, Martel AL. A Cluster-then-label Semi-supervised Learning Approach for Pathology Image Classification. *Sci Rep* 2018; **8**: 1–13.

73 Cuff J, Higgins JPT. Statistical analysis of surgical pathology data using the R program. Adv Anat Pathol. 2012; **19**: 131–139.

74 Bellovin SM, Dutta PK, Reitinger N. Privacy and Synthetic Datasets. *Stanford Technology Law Review* 2019; **22**: 1–52.

75 Chen RJ, Lu MY, Chen TY, Williamson DFK, Mahmood F. Synthetic data in machine learning for medicine and healthcare. Nat Biomed Eng. 2021; **5**: 493–497.

76 Senaras C, Muhammad KKN, Sahiner B, Pennell MP, Tozbikian G, Lozanski G *et al.* Optimized generation of high-resolution phantom images using cGAN: Application to quantification of Ki67 breast cancer images. *PLoS One* 2018; **13**: 1–12.

77 Platiša L, Van Brantegem L, Haeghen Y Vander, Marchessoux C, Vansteenkiste E, Philips W. Psycho-visual evaluation of image quality attributes in digital pathology slides viewed on a medical color LCD display. In: *Proc.SPIE*. 2013 doi:10.1117/12.2006991.

78 Xu Z, Moro CF, Bozóky B, Zhang Q. GAN-based virtual re-staining: A promising solution for whole slide image analysis. *ArXiv* 2019.https://arxiv.org/abs/1901.04059 (accessed 20 Dec2020).

79 Hassell LA, Peterson JE, Pantanowitz L. Pushed Across the Digital Divide: COVID-19 Accelerated Pathology Training onto a New Digital Learning Curve. *Acad Pathol* 2021; **8**. doi:10.1177/2374289521994240.

80 WHO manual for organizing a national external quality assessment programme for health laboratories and other testing sites. 2016.http://apps.who.int/iris/bitstream/handle/10665/250117/9789241549677-eng.pdf;jsessionid=2ECC5E007F715197826E29E44938B9C5?sequence=1 (accessed 17 May2023).

81 Van Es SL. Digital pathology: semper ad meliora. *Pathology* 2019; **51**: 1–10.

82 Capitanio A, Dina RE, Treanor D. Digital cytology: A short review of technical and methodological approaches and applications. Cytopathology. 2018; **29**: 317–325.

83 Ross J, Greaves J, Earls P, Shulruf B, Van Es SL. Digital vs traditional: Are diagnostic accuracy rates similar for glass slides vs whole slide images in a non-gynaecological external quality assurance setting? *Cytopathology* 2018; **29**: 326–334.

84 Zhu JY, Park T, Isola P, Efros AA. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. *Proceedings of the IEEE International Conference on Computer Vision* 2017; **2017-Octob**: 2242–2251.

# CHAPTER 6: PUBLISHED RESEARCH

## 6.1: THE DYNAMICS OF PATHOLOGY DATASET CREATION USING URINE CYTOLOGY AS AN EXAMPLE

**McAlpine ED**, Michelow P, Celik T. The dynamics of pathology dataset creation using urine cytology as an example. *Acta Cytol*. 2022 Jan. 66:46-54. doi: 10.1159/000519273.

This article has been publised in Acta Cytologica (Karger).

Due to copyright limitations this article can be located at the following website address:

https://karger.com/acy/article/66/1/46/821127/The-Dynamics-of-Pathology-Dataset-Creation-Using

## 6.2: IS IT REAL OR NOT? TOWARDS AI-BASED REALISTIC SYNTHETIC CYTOLOGY IMAGE GENERATION TO AUGMENT TEACHING AND QUALITY ASSURANCE IN PATHOLOGY.

**McAlpine ED**, Michelow P, Liebenberg E, Celik T, Is it real or not? Towards AI-based realistic synthetic cytology image generation to augment teaching and quality assurance in pathology. *Journal of the American Society of Cytopathology*. 2022 May – June. 11:123-132. doi: 10.1016/j.jasc.2022.02.001.

An interview relating to this publication can be found on CytoPathPod – the American Society of Cytopathology Podcast, published 5th of August 2022.

https://cytopathpod.podbean.com/

This article has been published in Journal of the American Society of Cytopathology (Elsevier). Authors retain the right to use and share the contents of the published article for scholarly purposes – including in a thesis or dissertation.

Journal of the American Society of Cytopathology (2022) **11**, 123–132

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.jascyto.org/

# Is it real or not? Toward artificial intelligence-based realistic synthetic cytology image generation to augment teaching and quality assurance in pathology

**Ewen McAlpine, FCPath (SA)[a,b,*], Pamela Michelow, MSc[a,b], Eric Liebenberg, BSc[b], Turgay Celik, PhD[c]**

[a] *Department of Anatomical Pathology, National Health Laboratory Service, Johannesburg, South Africa*
[b] *Division of Anatomical Pathology, School of Pathology, University of the Witwatersrand, Johannesburg, South Africa*
[c] *School of Electrical and Information Engineering and Wits Institute of Data Science, University of the Witwatersrand, Johannesburg, South Africa*

Received 26 December 2021; received in revised form 20 January 2022; accepted 3 February 2022

**KEYWORDS**
Artificial intelligence;
Urine cytology;
Synthetic data;
Digital pathology;
Generative adversarial networks

**Introduction** Urine cytology offers a rapid and relatively inexpensive method to diagnose urothelial neoplasia. In our setting of a public sector laboratory in South Africa, urothelial neoplasia is rare, compromising pathology training in this specific aspect of cytology. Artificial intelligence-based synthetic image generation—specifically the use of generative adversarial networks (GANs)—offers a solution to this problem.

**Materials and methods** A limited, but morphologically diverse, dataset of 1000 malignant urothelial cytology images was used to train a StyleGAN3 model to create completely novel, synthetic examples of malignant urine cytology using computer resources within reach of most pathology departments worldwide.

**Results** We have presented the results of our trained GAN model, which was able to generate realistic, morphologically diverse examples of malignant urine cytology images when trained using a modest dataset. Although the trained model is capable of generating realistic images, we have also presented examples for which unrealistic and artifactual images were generated—illustrating the need for manual curation when using this technology in a training context.

*Corresponding author: Ewen McAlpine, FCPath (SA); Division of Anatomical Pathology, School of Pathology, University of the Witwatersrand Medical School, 7 York Road, Johannesburg 2193, South Africa; Tel.: (011) 489-8527; Fax: (011) 489 8512.
*E-mail address:* ewen.mcalpine@wits.ac.za (E. McAlpine).

2213-2945/$36 © 2022 American Society of Cytopathology. Published by Elsevier Inc. All rights reserved.
https://doi.org/10.1016/j.jasc.2022.02.001

**Conclusions** We have presented a proof-of-concept illustration of creating synthetic malignant urine cytology images using machine learning technology to augment cytology training when real-world examples are sparse. We have shown that despite significant morphologic diversity in terms of staining variations, slide background, variations in the diagnostic malignant cellular elements, the presence of other nondiagnostic cellular elements, and artifacts, visually acceptable and varied results are achievable using limited data and computing resources.
© 2022 American Society of Cytopathology. Published by Elsevier Inc. All rights reserved.

## Introduction

Urine cytology is a commonly used screening modality for the detection and follow-up of bladder cancer. In South Africa, bladder cancer (urothelial neoplasia) is the 9th most common cause of cancer in men, with an incidence of 3.56 cases per 100,000, and the 17th most common cancer in women, with an incidence of 1.36 cases per 100,000.[1] The use of urine cytology offers a rapid and relatively inexpensive method to diagnose urothelial neoplasia. However, the lack of optimal sensitivity, diagnostic accuracy, and high interobserver variability have remained challenges in this field. The Paris System (TPS) for reporting urinary cytology was introduced in 2016 to standardize the diagnosis and reporting of urine cytology.[2]

In the setting of the South African public health sector, metastatic squamous cell carcinoma from cervical primary cancer accounts for most malignant urine cytology diagnoses, with malignant urothelial samples accounting for <20% of malignant cases. This is understandable because cervical carcinoma is far more common than bladder cancer locally, with an incidence of 23.11 per 100,000 persons.[1] This has several ramifications for teaching, training, and quality assurance in our academic laboratory setting.

1. Trainees (residents, registrars, and cytology practitioner students) might not be exposed to an adequate number or sufficient variety of urine cytology specimens during their training.
2. A limited number of cases are available from which to draw for quality assurance purposes to reduce diagnostic error and interobserver variability (eg, with the use of calibration and/or reference slide sets to inform pathologists and technologists to standardized diagnostic criteria and terminology).

Furthermore, unlike histopathology, for which a relatively large amount of tissue will be present in paraffin blocks, cytology slides cannot be duplicated or recut for teaching or quality assurance purposes. To a certain extent, the advent of whole slide images (WSIs) has allowed for the creation of teaching and quality assurance slide sets. However, this process remains very labor intensive and onerous and requires that each slide undergo manual validation by multiple pathologists or cytology practitioners before its inclusion in a reference slide set.[3]

Machine learning and artificial intelligence offers a potential solution to the problem of limited teaching and training material. Generative adversarial networks (GANs), first introduced in 2014 by Goodfellow et al,[4] broadly speaking are a family of machine learning models that are able to generate novel, synthetic examples of the data used to train them. GANs comprise 2 deep neural networks that compete with one another during training—a generator network that creates synthetic data points (eg, cytology images, such as in the present study) and a discriminator network that learns to distinguish these synthetically generated data points from real-world examples. The general concept of GANs and how these paired networks are trained are shown in Fig. 1. Since their introduction 2014, great strides have been made in improving the quality and resolution of images that these models produce.[5] Limitations remain in that these models are extremely resource intensive, requiring multiple high-end, expensive graphics processing units (GPUs) to train and, until recently, requiring enormous image datasets for training (ie, several hundred thousand to several million images). In addition to the computer resources required for training, the principal limitation to adapting this technology to a discipline such as cytopathology has been the lack of large, high-quality datasets. Dataset creation is a time-consuming and laborious process, and creating datasets of several million cytology images is largely unfeasible.[6] Despite these limitations, the prospect of using GANs to generate unlimited and diverse synthetic examples of pathology images is clearly appealing and has very definite potential for teaching and training undergraduate and postgraduate pathology students and for ongoing quality assurance of trained personnel.

A major advancement with particular promise for pathology is the development of a GAN architecture called StyleGAN2-adaptive discriminator augmentation (ADA), described by Karras et al[7] in 2020 in an aptly titled report "Training generative adversarial networks with limited data." The main contribution of their work was to introduce the ADA technique, allowing the GANs to achieve realistic results with orders of magnitude fewer data than previously required. Data augmentation, a commonly used technique in supervised machine learning when training using limited datasets, is particularly undesirable in the setting of GANs because the generator learns to generate the augmented data distribution (eg, rotation, translation,

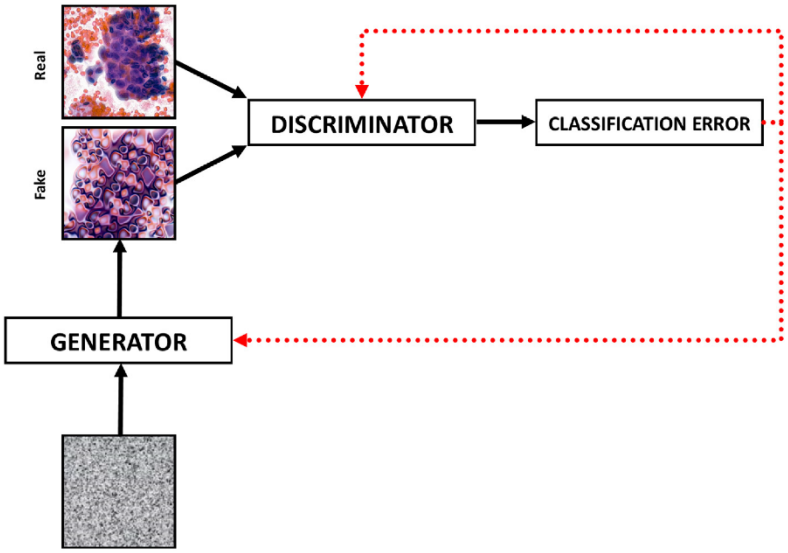



**Figure 1** An illustration of a generative adversarial network model, comprising a generator that learns to generate a synthetic image from random noise and a discriminator that learns to distinguish these synthetic images from real-world examples. The classification error of the discriminator is back propagated to both the discriminator (aiming to minimize the error) and the generator (aiming to maximize the discriminator's error).[8]

color transformation, added image noise). StyleGAN2-ADA introduced a "nonleaky" augmentation technique that allowed the discriminator to be augmented without these image augmentations being replicated by the generator. This ADA technique has persisted in the latest iteration of the StyleGAN family of models, StyleGAN3, introduced in 2021. In view of the ability of these models to learn from limited datasets, their potential for pathology image generation is clear.

Another unique challenge to synthetic image generation in cytology is the morphologic diversity that exists in the discipline—even a single diagnostic category (eg, malignant urine cytology images) can present with a wide array of morphologic patterns with which pathology trainees must become acquainted. In cytopathology, this morphologic diversity stems from factors such as staining variation, background factors (eg, blood, inflammation), variations in the diagnostic cellular elements (eg, position, rotation, individual cells, cellular clusters), the presence of other nondiagnostic cellular elements (ie, benign, nonlesional cells), and the presence of noncellular elements (ie, contamination, crystals), and artifacts. To be useful in teaching and training settings, synthetically generated images must represent this morphologic diversity as far as possible.

Traditional GANs models have a number of pitfalls that have been the subject of extensive research since 2014.[8] One of these limitations is of particular importance when generating synthetic pathology images for teaching and quality assurance purposes and is termed "mode collapse." When generating synthetic images for the purposes illustrated in the present study, one will necessarily need generated images that will reliably represent the diversity of real-world pathology images. In mode collapse, certain modes (ie, classes, examples) will not be well represented in the generated images. Karras et al[9] introduced a technique to incentivize the generation of a variety of images. This technique has been carried forward into the StyleGAN family of algorithms.

Given the combination of morphologic diversity and limited training data, synthetic cytology image data can be challenging to generate. In the present study, we have provided an example of how the use of StyleGAN3 can ameliorate these difficulties using malignant urine cytology as an example.

Additionally, given the morphologic diversity present in cytology, we have introduced a preprocessing step in an attempt to ensure that a wide array of morphologic patterns are represented in the training data. Our method makes use of unsupervised K-means clustering to group images with similar visual features together and then sample images equally from these clusters. Unsupervised machine learning attempts to identify patterns in unannotated data and is very well suited to grouping similar data points (eg, images) together without user input.[10] Specifically, K-means clustering involves iteratively assigning data points to a cluster based on their distance (or similarity) from a randomly selected cluster center (called a "centroid"). This process is repeated until the algorithm converges.

Finally, we have briefly discussed the concept of synthetic data in the context of medical education and some of the legal and ethical concerns regarding the use of synthetic data.

## Materials and methods

The present study involved the use of anonymized, archived cytology slides. The Human Research Ethics committee (medical) of the University of the Witwatersrand granted ethical clearance for the present study (certificate no. M190604). After ethical approval, a set of 26 malignant urine cytology slides, derived from 2 different cytology laboratories—one serving the public sector and one a private laboratory, were digitized using a Pannoramic 250 digital scanner (3DHISTECH Ltd, Budapest, Hungary) at ×400 magnification. The WSIs were tiled into 512 × 512 pixels (px) still images. A total of 3000 images representing malignant urothelial cytology samples were manually selected by 2 cytopathologists (either in tandem or individually).[11] Most of the dataset was annotated by a single expert pathologist, with a subset of the dataset (~10%) annotated by two expert pathologists in unison as a quality assurance measure. The criteria stipulated in TPS 2016 for reporting urine cytology was used to select malignant examples.[2] This initial set of images was unbalanced because of morphologic diversity (some slides were more cellular than others, creating an overrepresentation of certain morphologic and staining patterns in the dataset), and an additional preprocessing step to counter this was used. The image set was grouped into 20 clusters using K-means clustering with cosine similarity, and the 50 images closest to each cluster centroid were automatically included in the 1000-image training set. More details of the method used in this process have been presented in our previous work.[11] This preprocessing step was used to prevent the trained model from learning to synthesize only the more common morphologic examples in the dataset. Also, because of the limited computing resources available for our study, the images were resized from 512 × 512 px to 256 × 256 px before model training.

The present limited, but morphologically diverse and representative, dataset was used to train a StyleGAN3 model to generate 256 × 256 px synthetic urine cytology images. The official NVIDIA PyTorch (NVIDIA, Santa Clara, CA) implementation of StyleGAN3, made available for noncommercial or research purposes under the NVIDIA source code license, was used without modification.[12,13] The network was trained using 2,820,000 images shown to the discriminator network on an NVIDIA RTX 3060 GPU during a 9-day, 21-hour period.

Following manual visual inspection of the generated results, the saved checkpoint (at 2,44 million images shown to the discriminator) was selected as the checkpoint that generated the best images, as determined subjectively by an expert pathologist. A synthetic dataset of 1000 images was then generated by the trained GANs. To visually assess the morphologic diversity of both the real and synthetic datasets, each image set was again grouped into 20 clusters using K-means clustering with cosine similarity.

## Results

The trained StyleGAN3 model generated realistic, morphologically diverse examples of malignant urine cytology images, even when trained using a modest dataset of 1000 real cytology images and the limited computing resources of a single 12-Gb NVIDIA GPU. Our results have illustrated that generating synthetic cytology images is achievable for most pathology institutions worldwide. In the present study, we have demonstrated that although this technology is capable of producing realistic synthetic images that can be used for teaching, training, or quality assurance purposes, limitations exist, and synthetically generated image sets require manual curation. This manual curation or image selection is commonly referred to as "cherry-picking" and is a known limitation of GANs technology. We have presented visual examples illustrating the morphologic diversity created by the GANs (by comparing it with the training set of real images), successful examples of synthetic images created by the GANs, and examples of unrealistic synthetic urine cytology images. Furthermore, although only malignant examples of urine cytology were synthesized in our proof-of-concept study, a broader array of urine cytology examples covering the spectrum of TPS for classifying urine cytology could potentially be generated with additional data and computing time.[2] Although synthetic image quality lies on a spectrum, after visual inspection by pathologists, ~15% of images were considered unrealistic and/or nonrepresentative (contained no diagnostic cellular elements).

Two commonly used methods to determine the success of a trained GANs are visual inspection and, more objectively, measurement of the Fréchet inception distance (FID) between the real and synthetic images.[5,8] The feature vectors of these 2 groups were extracted using a pretrained Inception v3 model trained on the ImageNet dataset. The difference between these 2 image sets was calculated mathematically. Although the FID has been shown to correlate with image quality,[14] 2 important aspects should be noted. First, the pretrained feature extraction model has not been trained in cytology-specific datasets. Second, the network appeared to focus more on textures than on shapes.[15] The FID between the real and synthetic images throughout the training process in the present work is shown in Fig. 2. An appreciable exponential decrease in the FID was seen at the beginning of the training process; however, the FID plateaued at ~30. Although an FID of 0 represents a perfect score, given that the network used to extract the image features was not trained in a cytology-specific dataset

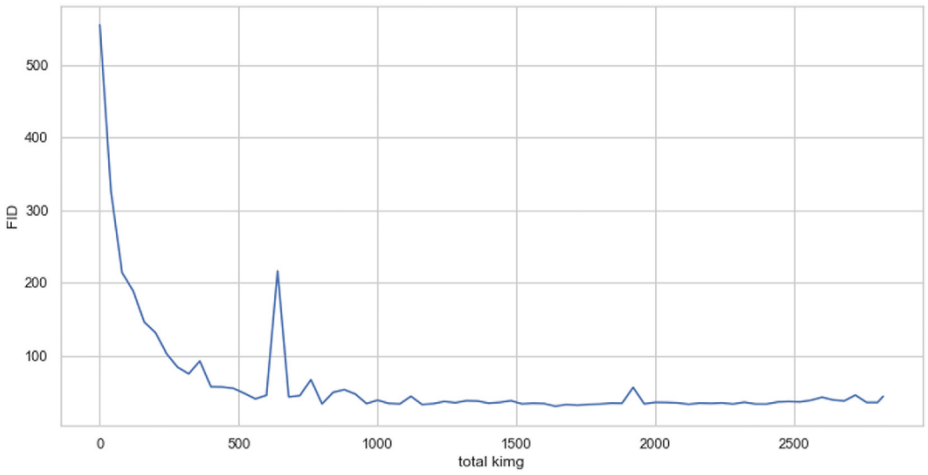



**Figure 2** Fréchet inception distance (FID) between real and synthetic images during the generative adversarial network training process.

and a small dataset was used in the present work, achieving a lower FID seemed unlikely. An important advantage to the use of the FID in this context was its inherent sensitivity to interclass mode collapse, making it well suited to assessing the diversity of images created by the GANs.[14] Mode collapse is particularly undesirable in the setting of cytology images when wide diversity is present.

Using the more subjective approach of visual inspection, examples of images deemed realistic and diagnostic of malignant urine cytology by 2 cytopathologists are presented in Fig. 3. Apart from being highly realistic, the images represented a significant spectrum of morphologic diversity present in urine cytology specimens.

Examples of real and synthetic images, derived from each of the 20 clusters and presented side-by-side for ease of comparison, are shown in Fig. 4. The morphologic diversity of the real dataset in terms of staining variations, slide background (eg, blood, acute inflammation), variation in the diagnostic malignant cellular elements (ie, position, rotation, individual cells, cellular clusters), the presence of other nondiagnostic cellular elements (eg, benign urothelial cells), and artifacts were replicated in the synthetic image dataset.

The synthetic image sets created by GANs require manual curation because unrealistic images can be created by these models. Examples of unrealistic images are presented in Fig. 5. This subset of images includes examples in which unrealistic nuclear detail was created, including merging the nuclei of cellular clusters into a "wavy" pattern, inconsistencies in cytoplasmic staining, nondiagnostic images containing artifact only, and repetitive unrealistic patterns, in both the background and cellular elements. Similar background artifacts, such as a mosaic-like pattern in the background whitespace, was noted by Wei et al[16] in their synthetically generated images of colonic polyps.

Furthermore, to illustrate the unrealistic staining patterns generated by our StyleGAN3 model, side-by-side comparisons between the synthetic images demonstrating more believable, natural staining that would be seen in a real-life setting and staining that would be less believable and more artifactual are shown in Fig. 6.

## Discussion

GANs have been used in pathology in a variety of settings to date. The use of this technology has predominantly been to improve supervised machine learning algorithms (eg, classification algorithms for diagnostics) by normalizing the straining of slides from different laboratories, creating synthetic versions of special histochemical and immunohistochemical stains from a real stain (termed "stain transfer"), and augmenting real datasets with synthetically generated data.[10]

Generating realistic synthetic images for medical education is a clear progression of such a process. Senaras et al[17] explored the creation of synthetic ("phantom images") Ki67 immunohistochemical images in breast carcinoma. They noted the potential of GANs to theoretically produce an unlimited amount of teaching material for pathology training. In their study, they demonstrated an example of image-to-image translation using GANs.[17] A GANs was trained to create a Ki67 image given an input nuclear segmentation mask. In work more akin to that in the present study, Beers et al[18] produced synthetic fundoscopy images using an older iteration of GANs technology termed a "progressively growing GAN" (ProGAN) developed in 2017.[9] The fundoscopy images generated by this architecture were realistic; however, the investigators reported that some unrealistic images were also produced, which were

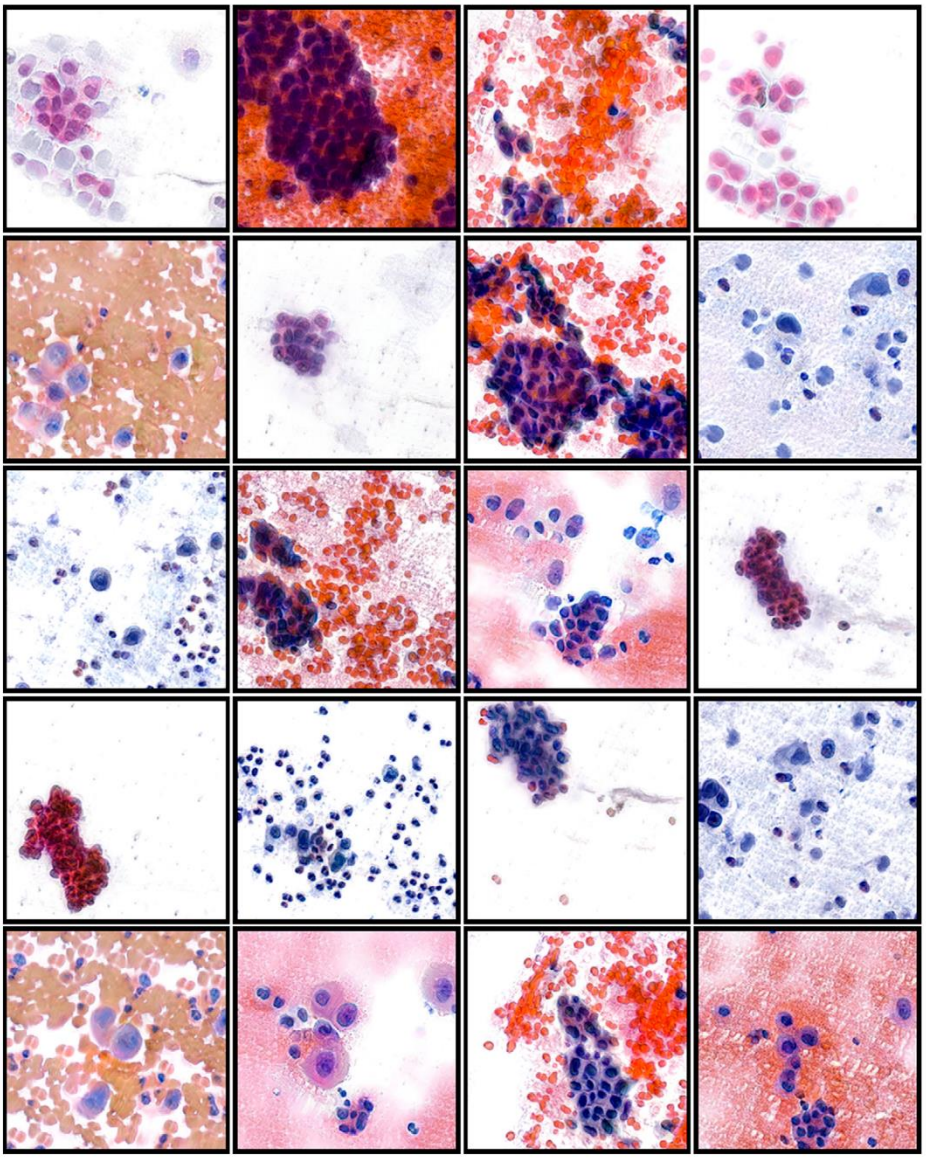


**Figure 3** Realistic, diagnostic synthetic malignant urine cytology images. Note the morphologic diversity present in the synthetic image set. Note also that these images show largely appropriate background features (including cleaner images, blood, and inflammation), cellular orientation (cluster and individual cells), cell size, and nuclear features (irregularity and hyperchromasia).

also seen in the present work. In addition, the dataset used to train this ProGAN was >5 times the size of the dataset used to train our StyleGAN and the morphologic diversity in cytology is subjectively greater than that seen in fundoscopy images.

Wei et al[16] used a CycleGAN to transform histologic images of normal colonic mucosa into examples of adenomatous polyps and sessile serrated lesions to balance a dataset and not to supplement pathology teaching material, as we have performed in the present study.[19] CycleGAN offers enormous potential in pathology owing to its ability to transform images of one type (eg, a stain) into a different type in an unpaired manner.[10] The techniques described in the present study do not rely on conditional input and use a GANs architecture that is able to generate entirely novel synthetic examples of malignant urine cytology.

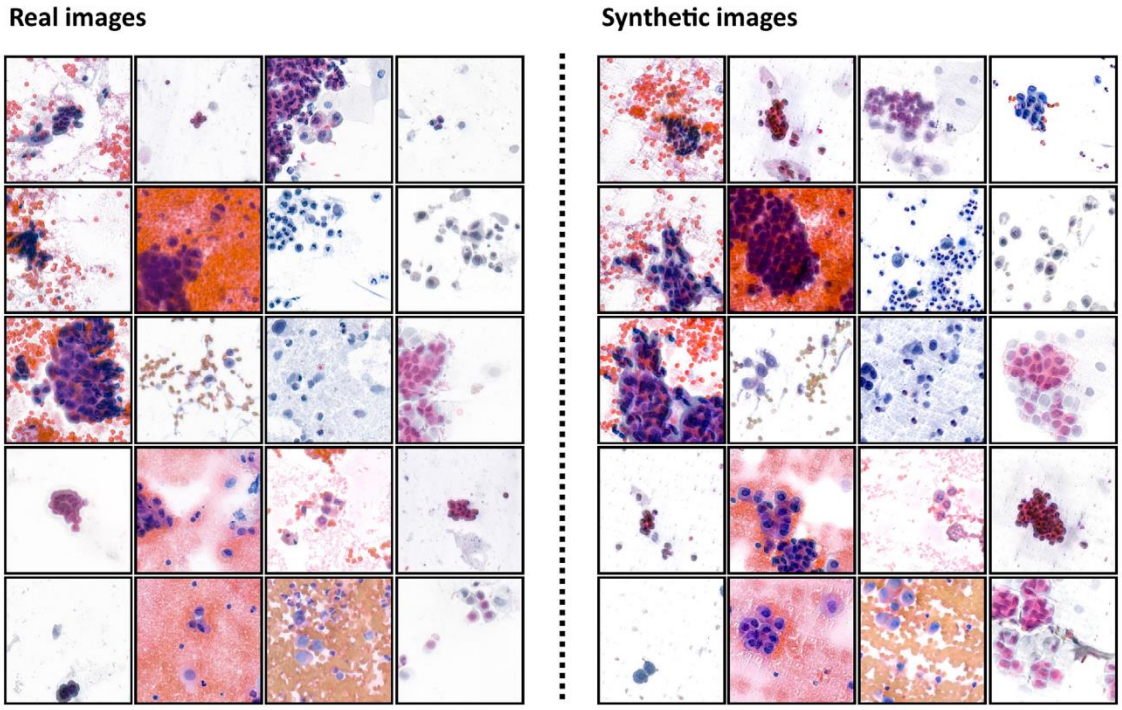



**Figure 4** A side-by-side comparison illustrating that the significant morphologic diversity seen in the real dataset (training images) is replicated by the generative adversarial network in the synthetic (fake) images.

We restricted our study to the generation of relatively small (256 × 256 px), single, still cytology images. The reasons for this are twofold: first, the enormous computing resources and time required for training the GANs to higher resolutions; and, second, the need for larger training datasets to achieve believable results at these higher resolutions. The

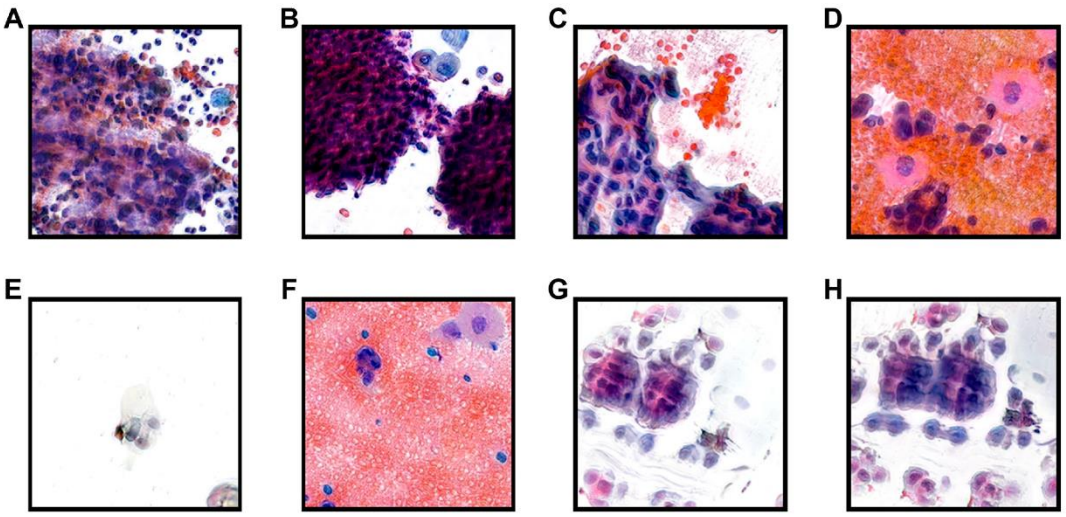



**Figure 5** Examples of less realistic synthetic urine cytology images generated by the trained generative adversarial network model. A-C, Unrealistic nuclear detail, including a "wavy" artifact (c). D, Unrealistic cytoplasmic staining of benign urothelial cells. E, Nondiagnostic image containing artifact only. F, Unnatural repeating pattern in background blood. G and H, Unrealistic repetitive pattern in diagnostic cellular elements.

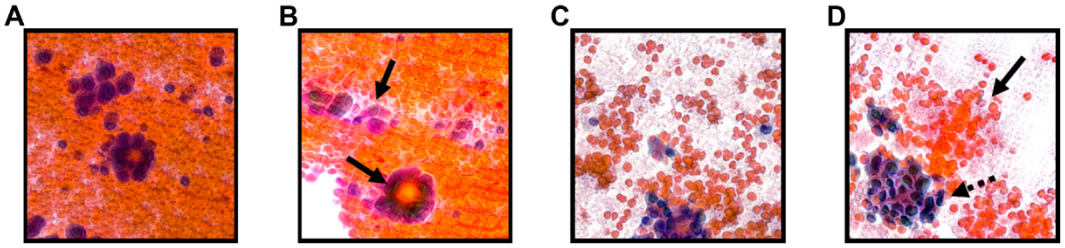



**Figure 6** A side-by-side comparison between synthetic images demonstrating more believable staining and staining that is less believable and artifactual. Images A and B and images C and D represent synthetic images from similar morphologic groups, with images A and C showing more natural and realistic staining.

computer resources and appropriately trained personnel required for this are not readily available in our low middle-income setting. Nevertheless, we have achieved visually acceptable results with a limited dataset and limited computing resources. StyleGAN3 is capable of generating high-resolution images ($\leq$1024 $\times$ 1028 px). However, training this model would require a full week using 8 high-end GPUs at a cost unachievable for most pathology departments and several months with the GPU used in the present study.[12] Another prospect for future research would be to create entirely synthetic WSIs for cytology teaching. Levy et al[20] demonstrated the use of creating virtual WSIs of Masson trichrome stains from a hematoxylin and eosin-stained liver sections. Again, although GANs technology is capable of being adapted in this manner, the computing resources required are substantial, although they should become more obtainable in the future.

The concept of using synthetic data in healthcare has been explored by several groups in recent years. Such data can be used for both research and education purposes; however, just as with most advancements in medicine, the use of synthetic data has introduced novel legal and ethical dilemmas.[21] Large datasets of patient data are highly prized by both health researchers and commercial organizations alike but are not easily accessible because of privacy concerns.[22] Additionally, access to large datasets of patient data are an invaluable resource for teaching and training purposes.[22,23] The traditional approach to address privacy and ethical concerns has been anonymization and the de-identification of datasets. However, a risk of re-identification remains as additional and/or auxiliary information becomes available in the future.[21,22] At present, anonymized patient data are routinely sold by governments, corporate entities, and clinical groups.[24] Synthetic health data have been posited as a solution to the inherent problems of anonymized data. Specifically, to address the risk of re-identification and the tradeoff between dataset utility and privacy, all personally identifiable information is removed from the anonymized data. However, these data will remain useful for data mining and research purposes.[21,22] One of the main problems with synthetic data is that, without specifically instituted privacy safeguards, a theoretical risk exists of "data leakage" such that some real data will be leaked into the synthetic dataset, compromising patient privacy.[21]

Both Goncalves et al[22] and Walonoski et al[24] have demonstrated the use of generative algorithms to create synthetic patient datasets, although neither of these groups included the generation of synthetic cytology data in their methods. Goncalves et al[22] explored different generative models (including GANs) to generate synthetic data from an epidemiologic dataset, while Walonoski et al[24] developed open-source software that creates synthetic electronic health records by simulating the lifespan of virtual patients. This technology has significant potential for healthcare education and will allow for the generation of synthetic patients for teaching and training and maximize patient privacy. The potential of including synthetically generated visual pathology and radiology data is a logical next step in this process. The prospect of creating synthetic images from synthetically generated pathology or radiology reports using GANs technology is becoming at least theoretically possible, with research into text-to-image generation,[25,26] although research into this specific use of the technology in pathology has yet to be undertaken.

However, in terms of ethical and legal concerns and because the use of synthetic data is so new in healthcare, legislation has not addressed the use of synthetic data directly,[21] nor is it likely that ethics review boards have sufficient knowledge and experience with synthetic data and the theoretical risk of "data leakage" to make informed decisions regarding the use of such data at present. Although the risk of data leakage (and compromised patient privacy) when creating single synthetic cytology images, as has been performed in the present study, is negligible, when incorporating such images in the synthetic electronic health records of virtual patients, this risk increases. Bellovin et al,[21] in their recent commentary on the topic in the Stanford Technology Law Review, made the point that the current legislation in the United States does not specifically deal with synthetic data and, as such, when current legislation is applied to the subject, the risk exists of both an overly conservative and overly lenient interpretation of privacy. Bellovin et al[21] also made 2 recommendations because of the advent of generative models and synthetic

data. First, they recommended that legislation be amended to encapsulate the nuances of synthetic data, the risk of data leakage, and the steps taken to mitigate against this risk. Second, they recommended that the use of synthetic data should remain preferable for medical research when available.[21] Moreover, until such legislative changes occur, it would seem prudent to adopt a cautious ethical approach and ensure that ethics review boards retain oversight of the use of synthetic healthcare data.

## Conclusions

In the present report, we have presented a proof-of-concept illustration for creating synthetic malignant urine cytology images using machine learning technology. The motivation behind this adaptation of GANs to a cytology setting is the potential use of synthetic data to augment teaching and training when real-world material is sparse, instead of the traditional use of GANs to improve supervised machine learning algorithms. Furthermore, we have illustrated that despite the significant morphologic diversity inherent in cytology, visually acceptable and varied results are achievable with limited data and computing resources. Finally, we have provided a brief discussion of the legality and privacy concerns arising with the use of synthetic data in the context of health research and education.

## Funding sources

The present study was partially funded by grants from the University of the Witwatersrand, Faculty of Health Sciences Research Equipment Grants.

## Conflict of interest disclosures

The authors made no disclosures.

## Statement of ethics

This study involved the use of anonymized, archived cytology slides. Ethical clearance for this study was granted by Human Research Ethics Committee (Medical) of the University of the Witwatersrand (Certificate number: M190604).

## Author contributions

All 4 authors contributed to the conceptualization of this study, the analysis of the results, and the writing of the manuscript.

## Acknowledgments

The authors acknowledge the National Health Laboratory Service, Lancet Laboratories, and 3F Scientific for their assistance with providing research material and digitizing the slides for the present research, respectively.

## References

1. National Institute for Communicable Diseases, Cancer Registry. Cancer in South Africa 2019 Full Report National Cancer Registry. Available at: https://www.nicd.ac.za/centres/national-cancer-registry/. Accessed March 2, 2020.
2. Barkan GA, Wojcik EM, Nayar R, et al. The Paris System for reporting urinary cytology: the quest to develop a standardized terminology. *Acta Cytol*. 2016;60:185–197.
3. Khalbuss W, Pantanowitz L, Parwani AV. Digital imaging in pathology. *Pathol Res Int*. 2011;2011:1–10.
4. Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks. Available at: https://arxiv.org/abs/1406.2661; 2014. Accessed July 31, 2021.
5. Brownlee J. Generative Adversarial Networks with Python. v1.4. Available at: https://machinelearningmastery.com/; 2019. Accessed July 12, 2020.
6. Abels E, Pantanowitz L, Aeffner F, et al. Computational pathology definitions, best practices, and recommendations for regulatory guidance: a white paper from the digital pathology association. *J Pathol*. 2019;249:286–294.
7. Karras T, Aittala M, Hellsten J, et al. Training generative adversarial networks with limited data. *Adv Neural Inf Process Syst*. 2020;33:12104–12114.
8. Langr J, Bok V. *GANs in Action*. Shelter Island, NY: Manning Publications Co.; 2019.
9. Karras T, Aila T, Laine S, et al. Progressive growing of GANs for improved quality, stability, and variation. 1-26. Available at: https://arxiv.org/abs/1710.10196; 2017. Accessed December 6, 2018.
10. McAlpine ED, Michelow P, Celik T. The utility of unsupervised machine learning in anatomic pathology. *Am J Clin Pathol*. 2022;157:5–14.
11. McAlpine ED, Michelow P, Celik T. The dynamics of pathology dataset creation using urine cytology as an example. *Acta Cytol*. 2022;66:46–54.
12. StyleGAN3. Available at: https://github.com/NVlabs/stylegan3. Accessed December 13, 2021.
13. Karras T, Aittala M, Laine S, et al. Alias-Free Generative Adversarial Networks. Available at: https://arxiv.org/abs/2106.12423; 2021. Accessed December 13, 2021.
14. Borji A. Pros and cons of GAN evaluation measures: new developments. Available at: https://arxiv.org/abs/2103.09396; 2021. Accessed December 17, 2021.
15. Karras T, Laine S, Aittala M, et al. *Analyzing and improving the image quality of styleGAN*. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020:8107–8116.
16. Wei J, Suriawinata A, Vaickus L, et al. Generative image translation for data augmentation in colorectal histopathology images. Available at: https://arxiv.org/abs/1910.05827. Accessed December 10, 2020.
17. Senaras C, Muhammad KKN, Sahiner B, et al. Optimized generation of high-resolution phantom images using cGAN: application to quantification of Ki67 breast cancer images. *PLoS One*. 2018;13:1–12.
18. Beers A, Brown J, Chang K, et al. High-resolution medical image synthesis using progressively grown generative adversarial networks. Available at: https://arxiv.org/abs/1805.03144. Accessed December 6, 2018.

19. Zhu JY, Park T, Isola P, et al. *Unpaired image-to-image translation using cycle-consistent adversarial networks*. 2017 IEEE International Conference on Computer Vision (ICCV); 2017:2242–2251.
20. Levy JJ, Azizgolshani N, Andersen MJ, et al. A large-scale internal validation study of unsupervised virtual trichrome staining technologies on nonalcoholic steatohepatitis liver biopsies. *Mod Pathol*. 2020; 34:808–822.
21. Bellovin SM, Dutta PK, Reitinger N. Privacy and synthetic datasets. *Stan Tech L Rev*. 2019;22:1–52.
22. Goncalves A, Ray P, Soper B, et al. Generation and evaluation of synthetic patient data. *BMC Med Res Methodol*. 2020;20:1–40.
23. Dube K, Gallagher T. International Symposium on Foundations of Health Informatics Engineering and Systems. Foundations of Health Information Engineering and Systems. *Approach and method for generating realistic synthetic electronic healthcare records for secondary use*. 2014;8315:69–86.
24. Walonoski J, Kramer M, Nichols J, et al. Synthea: an approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record. *J Am Med Inform Assoc*. 2018;25:230–238.
25. Tschuchnig ME, Oostingh GJ, Gadermayr M. Generative adversarial networks in digital pathology: a survey on trends and future potential. *Patterns*. 2020;1:100089.
26. Hu K, Liao W, Yang MY, et al. Text to image generation with semantic-spatial aware GAN. Available at: https://arxiv.org/abs/2104.00567; 2021. Accessed December 21, 2021.

## 6.3: ARE SYNTHETIC CYTOLOGY IMAGES READY FOR PRIME TIME? A COMPARATIVE ASSESSMENT OF REAL AND SYNTHETIC URINE CYTOLOGY IMAGES.

**McAlpine ED**, Michelow P, Liebenberg E, Celik T. Are synthetic cytology images ready for prime time? A comparative assessment of real and synthetic urine cytology images. *Journal of the American Society of Cytopathology. 2023.* March – April 12:126-135. doi: 10.1016/j.jasc.2022.10.001

This article has been published in Journal of the American Society of Cytopathology (Elsevier). Authors retain the right to use and share the contents of the published article for scholarly purposes – including in a thesis or dissertation.

Journal of the American Society of Cytopathology (2023) **12**, 126–135

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.jascyto.org/

# Are synthetic cytology images ready for prime time? A comparative assessment of real and synthetic urine cytology images

**Ewen McAlpine, FCPath (SA)[a,b,*], Pamela Michelow, MSc[a,c], Eric Liebenberg, BSc[a], Turgay Celik, PhD[d]**

[a] *Division of Anatomical Pathology, School of Pathology, University of the Witwatersrand, Johannesburg, South Africa*
[b] *Ampath National Laboratories, Johannesburg, South Africa*
[c] *National Health Laboratory Services, Johannesburg, South Africa*
[d] *School of Electrical and Information Engineering and Wits Institute of Data Science, University of the Witwatersrand, Johannesburg, South Africa*

Received 20 August 2022; received in revised form 17 September 2022; accepted 1 October 2022

**KEYWORDS**
Artificial intelligence;
Urine cytology;
Synthetic data;
Digital pathology;
Generative adversarial networks;
Cytology teaching and learning

**Introduction** The use of synthetic data in pathology has, to date, predominantly been augmenting existing pathology data to improve supervised machine learning algorithms. We present an alternative use case—using synthetic images to augment cytology training when the availability of real-world examples is limited. Moreover, we compare the assessment of real and synthetic urine cytology images by pathology personnel to explore the usefulness of this technology in a real-world setting.

**Materials and methods** Synthetic urine cytology images were generated using a custom-trained conditional StyleGAN3 model. A morphologically balanced 60-image data set of real and synthetic urine cytology images was created for an online image survey system to allow for the assessment of the differences in visual perception between real and synthetic urine cytology images by pathology personnel.

**Results** A total of 12 participants were recruited to answer the 60-image survey. The study population had a median age of 36.5 years and a median of 5 years of pathology experience. There was no significant difference in diagnostic error rates between real and synthetic images, nor was there a significant difference between subjective image quality scores between real and synthetic images when assessed on an individual observer basis.

**Conclusions** The ability of Generative Adversarial Networks technology to generate highly realistic urine cytology images was demonstrated. Furthermore, there was no difference in how pathology personnel perceived the subjective quality of synthetic images, nor was there a difference in diagnostic error rates

*Corresponding author: Ewen McAlpine, University of the Witwatersrand Medical School, 7 York Road, Parktown, Johannesburg, South Africa 2193; Tel.: (711) 489-8492; Fax: (711) 489-8512.
*E-mail address:* ewen.mcalpine@wits.ac.za (E. McAlpine).

2213-2945/$36 © 2022 American Society of Cytopathology. Published by Elsevier Inc. All rights reserved.
https://doi.org/10.1016/j.jasc.2022.10.001

Downloaded for Anonymous User (n/a) at University of the Witwatersrand Johannesburg from ClinicalKey.com by Elsevier on May 12, 2023. For personal use only. No other uses without permission. Copyright ©2023. Elsevier Inc. All rights reserved.

between real and synthetic urine cytology images. This has important implications for the application of Generative Adversarial Networks technology to cytology teaching and learning.
© 2022 American Society of Cytopathology. Published by Elsevier Inc. All rights reserved.

## Introduction

The use of synthetic data in pathology has, to date, predominantly been augmenting existing pathology data to improve supervised machine learning algorithms.[1] This synthetic data augmentation includes normalizing slide staining, creating synthetic versions of special histochemical/immunohistochemical stains from a real stain (stain transfer), and augmenting real data sets with synthetically generated data. In our previous work,[2] we illustrated a novel use for synthetic data in cytology motivated by a unique challenge faced when teaching cytology trainees in our setting of a public sector laboratory in South Africa. Briefly, the majority of our malignant urine cytology specimens represent metastatic squamous cell carcinomas from cervical primaries, resulting in a paucity of examples of urothelial pathology for teaching and training purposes.

A novel machine learning technology, termed Generative Adversarial Networks (GANs), offers a potential solution to this problem. GANs have proven effective in generating novel, synthetic examples of real-world data. Nonmedical examples of the use of this technology include human faces,[3-6] artwork,[6] and other everyday images such as birds, dogs, and cats.[7] This technology has also been used for medical image generation—for example, for creating synthetic fundoscopy images[8] and chest radiographs.[9] In our previously published research,[2] we demonstrated the ability of GANs to generate realistic 256 × 256 px examples of malignant urothelial cytology images from a limited data set of 1000 images. In the present work, we extend this adaption of GANs technology to cytology by generating images double the previous resolution (now 512 × 512 px) and including benign and atypical examples together with malignant images.

Additionally, we investigate how these synthetic cytology images are visually perceived by an audience of pathology personnel. In previous research, a so-called Turing test has been adopted to assess the ability of a human observer to distinguish between real and synthetic images.[1] In this work, we make use of a modified single stimulus method to compare a subjective image quality score and diagnostic accuracy between real and synthetic urine cytology images in place of a Turing test.

The subjective assessment of medical image quality has previously been studied by assessing subjective quality of experience,[10] and it seems a logical step to assess synthetic medical image data similarly. Current methodologies for assessment fall into 2 groups: single-stimulus and multistimulus. Single-stimulus methods present the observer with a single image (or "stimulus") at a time and they are asked to rate the image using a quantitative scale. Multistimulus methods present the observer with multiple images as reference, with 1 image being scored against the others. When assessing the quality of medical images, Platiša et al.[11] suggest that pathologists and nonmedical image experts place importance on different image attributes when assessing image quality and propose that pathologists should determine the image quality of digital pathology images when developing domain-specific image algorithms or imaging systems.

Given this background, this work presents a comparison between real and synthetic benign, atypical and malignant urine cytology images using 3 measures: (i) the relative diagnostic error rate between real and synthetic images; (ii) assessing whether pathology personnel would include synthetic images in a teaching set; and (iii) comparing the subjective image quality scores between real and synthetic images.

## Methods

### Synthetic image generation

Following ethical approval, a set of 214 voided urine cytology slides, derived from 2 different cytology laboratories—one a public sector laboratory and the other a private laboratory—were digitized using a Pannoramic 250 digital scanner (3DHISTECH Ltd., Budapest, Hungary) at ×400 magnification.

Of the 214 urine cytology slides, 142 (66.4%) were retrieved from a private pathology laboratory and the remaining 72 slides (33.6%) were obtained from a public sector laboratory. Initially, cases were retrieved retrospectively from 2020, with most cases being derived from 2017 and 2018; however, following an initial review of the research material, more recent cases from 2019 and 2020 were included in the data set, as were slides from the department's teaching collection dating back to 1988. This decision was made specifically to increase the morphological diversity of the training data set.

The whole slide images (WSIs) were tiled into 512 × 512 px still images. A data set of 13,746 images was manually annotated by 2 cytopathologists (either in tandem or individually[12]). This set comprised 7320 benign urothelial images, 3367 atypical urothelial images, and 3059 malignant urothelial images. On analysis of this data set of 13,746 images, it was found that benign images were derived from 174 of the 214 slides, atypical images from

Downloaded for Anonymous User (n/a) at University of the Witwatersrand Johannesburg from ClinicalKey.com by Elsevier on May 12, 2023. For personal use only. No other uses without permission. Copyright ©2023. Elsevier Inc. All rights reserved.

120 slides, and malignant images from 84 slides. The reason behind this is that malignant cases include a spectrum of benign and atypical cells in addition to malignant ones and, similarly, atypical cases include both atypical and benign urothelial cells.

As in our previous work,[2] clustering with cosine similarity was used to balance the data set in terms of morphological diversity. This preprocessing step resulted in a final training data set of 3240 benign urothelial images, 2400 atypical urothelial images, and 1820 malignant urothelial images.

This multiclass, representative and morphologically balanced data set was used to train a conditional StyleGAN3 model to generate 512 × 512 px synthetic benign, atypical, and malignant urine cytology images. The official NVIDIA PyTorch implementation of StyleGAN3[6,13] was used without modification. The network was trained for 3410 kimg (ie, 3,410,000 images shown to the discriminator network) on a single 24Gb NVIDIA RTX 6000 graphics processing unit over a 9-day-12-hour period.

A synthetic data set of 3000 synthetic images per class was created using the trained StyleGAN3 model. The entire training (real) image set and the synthetic data set were clustered into 10 clusters per class using K-means clustering with cosine similarity. Following visual inspection by a pathologist, a single image was selected from each cluster to create a set of 60 images for assessment by pathology personnel. The diagnostic category and visual suitability of each image (both real and synthetic) were confirmed independently by a second pathologist. Images were selected specifically to represent a diversity of real-world examples of urine cytology and to cover the morphological diversity of the data set as far as possible. The rationale behind the clustering was to include as much morphological diversity in the 60-image set as possible.

## Image assessment by pathology personnel

Real and synthetic urine cytology images were assessed by pathology personnel using a single stimulus method. Pathologists and pathology trainees (registrars/residents) were voluntarily recruited from the Division of Anatomical Pathology, University of the Witwatersrand, Johannesburg. These participants were invited to an optional, short didactic lecture presented by EM to refamiliarize participants with The Paris System for Reporting Urine Cytology.[14] Participation in this study was open to all pathologists and residents who had completed a 6-month cytology rotation. Informed consent was obtained from participants and, following this, participants used a custom-designed web application to complete an online, self-conducted questionnaire. Participants were blinded to the fact that synthetic images were included in the study. Images were randomized and assigned a filename from 0001-0060. Participants were presented with the images either in ascending or descending numerical order, alternating between participants. Participants were able to stop the evaluation process at any time and recommence the assessment process at a later point. A total of 60 images were individually presented to participants—comprising 10 real and synthetic images from the benign, atypical, and malignant classes. For each image, participants were required to (i) subjectively determine image quality (using a 5-point Likert scale—presented in Table 1); (ii) state whether they would include the image in a teaching set (yes or no); and (iii) provide a diagnosis (benign, atypical, or malignant). Participants were provided with a graphical summary of The Paris System criteria, including a visual reference to assist them with estimating nuclear to cytoplasmic ratios (Fig. 1). While the current Paris System[14] separates Suspicious for high-grade urothelial carcinoma and high-grade urothelial carcinoma on a quantitative basis, for the present study, these 2 categories were combined into a single malignant class as quantification of cells is difficult on static images. Study participants were made aware of this, and this was further illustrated in the visual reference provided to them (Fig. 1).

Although participants were instructed to assess image quality subjectively, they were asked to take factors such as image sharpness/focus, acceptability of staining, presence or absence of artifacts, image brightness, and image noise into consideration when assessing the images. There was no time for limit viewing each image and participants were not able to skip images. Participants were able to conduct image assessments at a time that was convenient to them.

## Statistical methods

Statistical analysis was conducted in the statistical analysis software R: A language and environment for statistical computing, R Core Team, Vienna, Austria (http://www.R-project.org/).[15] The heatmap of $P$ values was created using Python 3 (https://www.python.org/).[16]

The individual differences between real and synthetic images among participants were assessed across 3 categories: error rate, subjective inclusion in a teaching set, and subjective image quality score. Fisher exact tests were employed in the case of error rates and inclusion in a teaching set and a 2-sample Kolmogorov–Smirnov test was used when assessing quality scores. The advantage of the Kolmogorov–Smirnov test in this setting is that it allows for testing the difference in the distributions of quality scores rather than simply the median of the 2 distributions.[17]

**Table 1** The 5-point Likert scale used to subjectively assess image quality.

| Quality score | Category |
|---|---|
| 1 | Very low quality |
| 2 | Low quality |
| 3 | Average quality |
| 4 | Good quality |
| 5 | Very good quality |

Downloaded for Anonymous User (n/a) at University of the Witwatersrand Johannesburg from ClinicalKey.com by Elsevier on May 12, 2023. For personal use only. No other uses without permission. Copyright ©2023. Elsevier Inc. All rights reserved.

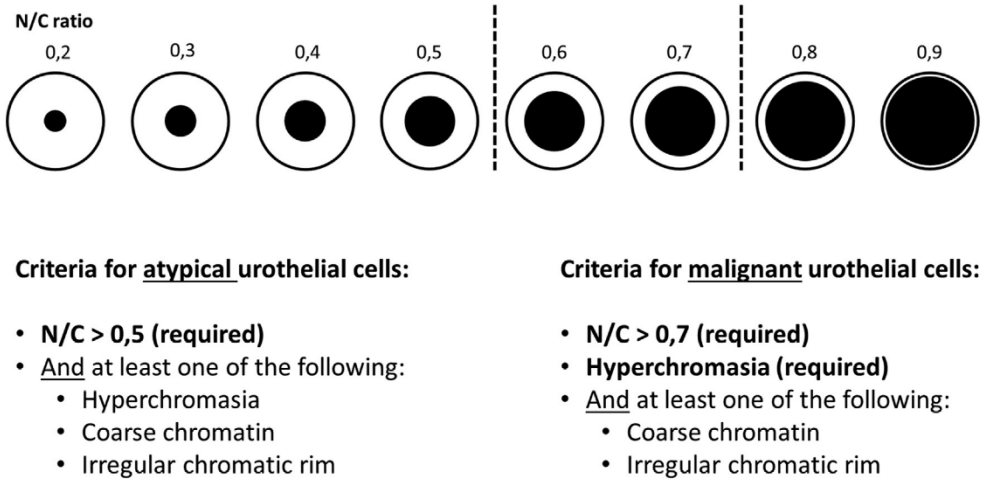



**Figure 1** Graphical summary of the morphological criteria for diagnosing urine cytology, based on TPS,[14] available to study participants as a visual aid. Abbreviation: TPS, The Paris System.

## Results

In our previous work on this topic,[2] our trained StyleGAN3 model was able to produce highly realistic 256 × 256 px images of malignant urine cytology. In the current work, we extend this to include benign, atypical, and malignant urine cytology and double the resolution of the synthetic images to 512 × 512 px. Figure 2 illustrates the Fréchet Inception Distance (FID) between real and synthetic images over the GANs training process. The FID is an objective measure of how similar the synthetic images are to real examples and has been shown to correlate with image quality.[18,19] The FID of our model at the end of the training process was 14.1 (approximately half that of our previous model[2]). As stated in our previous work,[2] manual curation of synthetically created data sets remains common practice because GANs can generate unrealistic images. Our trained model generated a minority of images regarded as unacceptable by a pathologist (between 5% and 10%) of images, although it must be emphasized that the synthetic images lie on a morphological spectrum, and determining an exact cutoff for what is acceptable is difficult. Causes for exclusion were similar to our previous findings: unrealistic artifacts and staining patterns, acellular images, and telltale repetitive patterns.[2]

A subset of images utilized in the image survey component of this project is presented in Figure 3. All 3 classes of urine cytology images are presented side by side.

### Image observers

A total of 12 participants completed the 60-image survey. This group comprised 7 pathologists (58.3%) and 5 residents (41.7%). The ages of the participants ranged from 31 to 57 years, with a median age of 36.5 years (interquartile

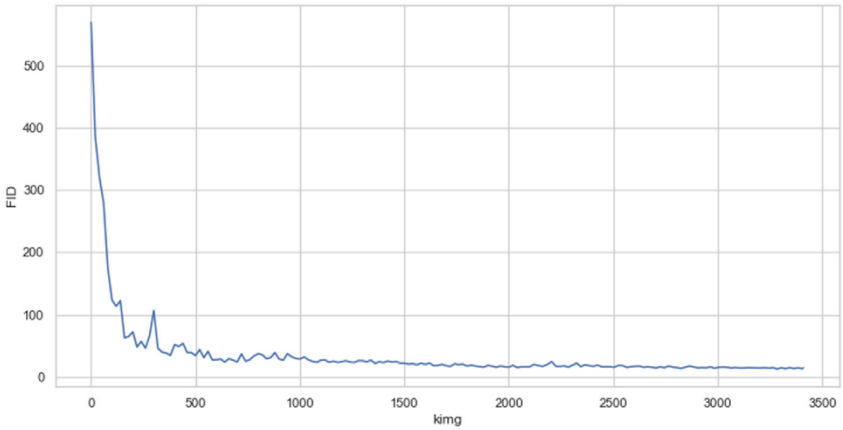



**Figure 2** Fréchet Inception Distance (FID) between real and synthetic images over the StyleGAN3 training process.

Downloaded for Anonymous User (n/a) at University of the Witwatersrand Johannesburg from ClinicalKey.com by Elsevier on May 12, 2023. For personal use only. No other uses without permission. Copyright ©2023. Elsevier Inc. All rights reserved.

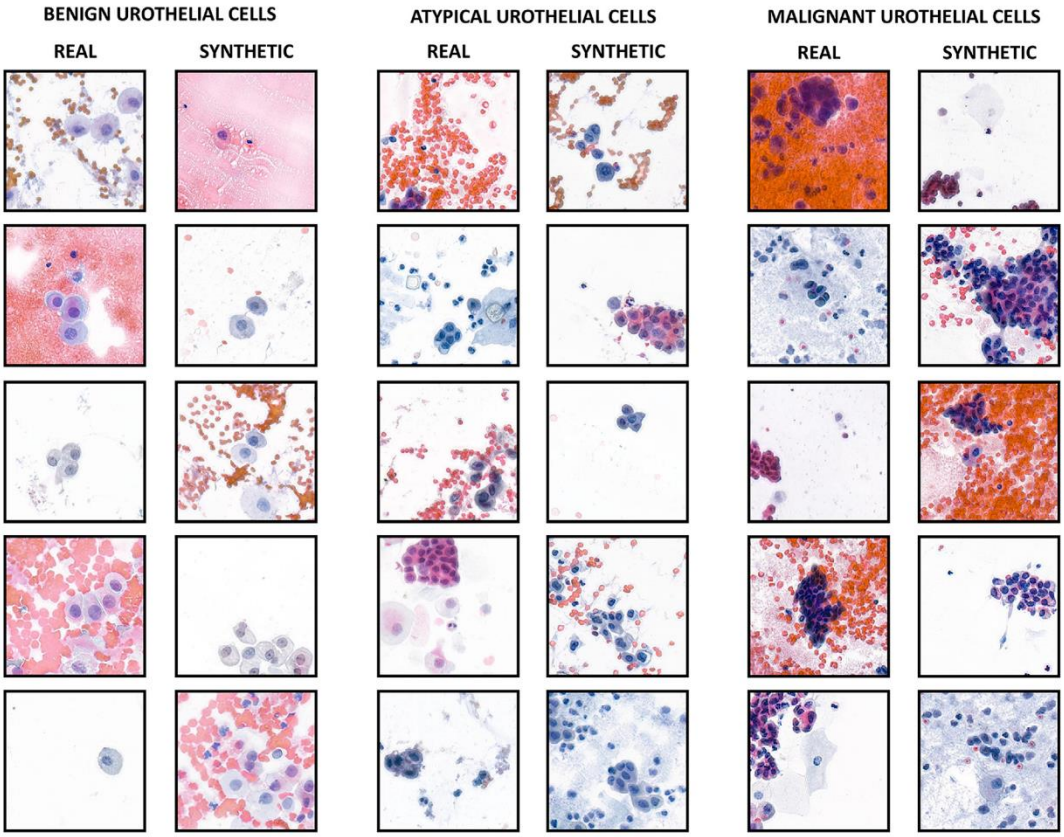



**Figure 3** A side-by-side comparison of a subset of real and synthetic benign, atypical, and malignant urothelial images.

range [IQR] = 5.5). The length of pathology experience of these observers ranged from 1 to 31 years, with a median of 5 years (IQR = 4).

## Diagnostic errors

The overall diagnostic error rate across all 12 observers was 202 errors over 720 observations (28.1%). A slightly higher rate of diagnostic errors was present in the synthetic image set, with 105 errors over 360 observations (29.2%), compared to the real image set where 97 errors were made over 360 observations (26.9%). The majority of errors was made among malignant images (55.5%), followed by atypical images (39.1%), and then benign images (5.4%). Errors between malignant and atypical (ie, atypical cases called malignant or vice versa) accounted for the great majority of errors (64.4%), errors between benign and atypical the second most common source of errors (29.2%), and the major diagnostic error of confusion between benign and malignant images were the least common (6.4%).

Errors per observer ranged from 8 to 29 images of the total 60 images, with a median of 16 (IQR = 9.25). Errors per observer across the 30 real image subset ranged from 2 to 15 with a median of 8 (IQR = 5.25), while errors per observer across the 30 synthetic image subset ranged from 4 to 14, with a median of 7.00 (IQR = 5.25).

## Inclusion in a teaching set

Overall, across all 720 observations, 421 images observations (58.5%) were selected for inclusion in a teaching set. More synthetic than real images were selected for inclusion — 225 of 360 synthetic images (62.5%) versus 196 of 360 real images (54.4%).

On an individual observer basis, images selected for inclusion in a teaching set ranged from 10 to 59 of the 60 images, with a median of 38 (IQR = 19). For real images, between 2 and 29 images were selected for inclusion, with a median of 18 (IQR = 7.5). For synthetic images, between 8 and 30 images were selected for inclusion, with a median of 18 (IQR = 8.25).

## Image quality score

Regarding the 720 image observations as a whole, quality scores ranged from 1 to 5, with a median of 3.0 (IQR = 1).

Downloaded for Anonymous User (n/a) at University of the Witwatersrand Johannesburg from ClinicalKey.com by Elsevier on May 12, 2023. For personal use only. No other uses without permission. Copyright ©2023. Elsevier Inc. All rights reserved.

Quality scores for real images ranged from 1 to 5, with a median of 3.0 (IQR = 1), while quality scores for synthetic images ranged from 1 to 5, with a median of 4 (IQR = 1).

### Real versus synthetic images on an individual observer basis

Analysis of the differences between real and synthetic images on an individual observer basis is presented as a heatmap of *P* values in Figure 4.

Testing the differences in error rates between real and synthetic images on an individual observer basis using a Fisher exact test confirmed that there was no statistically significant difference among any of the observers, with *P* values ranging from 0.25 to 1.

Similarly, the differences in selection for inclusion in a teaching set between real and synthetic images were not statistically significant, with *P* values ranging from 0.07 to 1, although 2 observers approached significance (*P* values of 0.07 and 0.08, respectively).

When tested on an individual observer basis, there was no statistically significant difference in the quality scores between real and synthetic images, with *P* values ranging from 0.07 to 1. Again, a single observer approached significance (*P* value of 0.07).

Interestingly, observers who approached a significant *P* value in 1 of the categories tested (error rate, inclusion in a teaching set, or subjective image quality score) did not show a pattern of approaching significance in the 2 remaining categories. As can be seen in Figure 3, observers 1, 4, and 8 approached significance in a single of the subjective categories only. None of the observers approached significance in the objective error rate category.

These results indicate that the *alternative hypothesis* (that there is a difference in perception between real and synthetic urine cytology images by observers) must be rejected across all 3 categories (error rate, inclusion in a teaching set, and subjective image quality) and for all 12 observers.

## Discussion

The first contribution of this work is the extension of our previous research[2] to include benign and atypical urothelial images in addition to malignant examples and to double the resolution of the images from 256 × 256 px to 512 × 512 px. In addition to this, we achieved an FID approximately half that of our previous work. This improvement in the realism of our images was achieved as a result of having access to a larger and faster graphics processing unit (a 24Gb NVIDIA RTX 6000 versus a 12Gb NVIDIA RTX 3060) while, at the same time, leveraging the larger data set resulting from including benign and atypical examples alongside the malignant images (a total data set size of 7460 images in this work compared to 1000 images in our previous work).

The second contribution of this work is to illustrate an evaluation of these synthetic urine cytology images by a

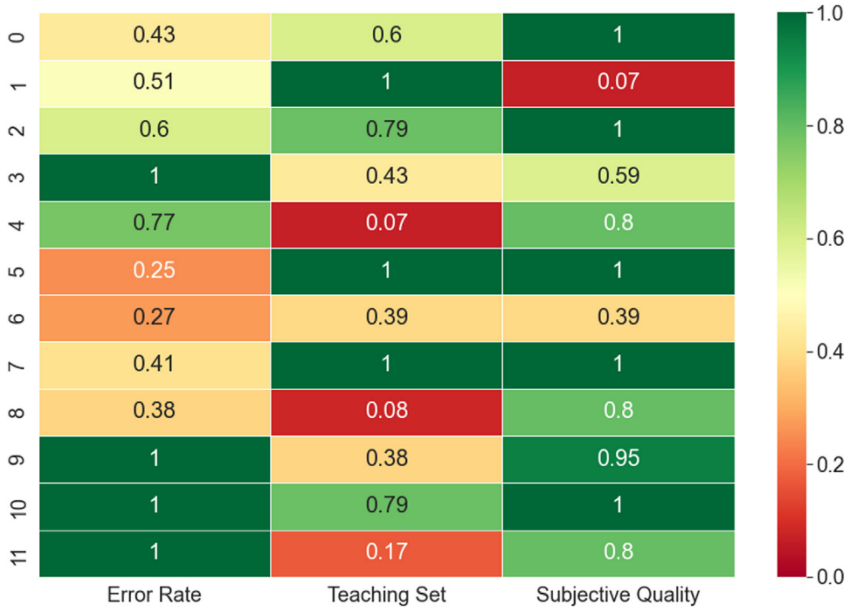



**Figure 4** A heatmap comparing the *P* values for the 12 individual observers (rows 0 through 11), for error rates (Fisher exact test), inclusion in a teaching set (Fisher exact test) and subjective image quality score (Kolmogorov–Smirnov test). Observer order has been randomized to protect anonymity.

Downloaded for Anonymous User (n/a) at University of the Witwatersrand Johannesburg from ClinicalKey.com by Elsevier on May 12, 2023. For personal use only. No other uses without permission. Copyright ©2023. Elsevier Inc. All rights reserved.

cohort of pathology personnel and to confirm that there was no difference in how such image observers perceive real and synthetic cytology images.

While testing the ability of pathologists to distinguish between real and synthetic images is not unreported in the literature, this work presents an alternative approach to the traditional Turing test where observers are asked to state whether an image is real or synthetic. In this study, we blinded image observers to the fact that the survey set contained both real and synthetic images, and rather than asking them to detect the synthetic data, the observers were asked to rate the overall quality of the image subjectively, state whether they would include the image in a teaching set and to provide a cytological diagnosis for each image. The main aim of this methodology was to assess whether there was an objective difference in the diagnostic error rate between real and synthetic images and how observers subjectively perceived the image quality of these images without knowing that half of the survey set was generated synthetically using GANs technology.

The overarching result of this investigation is that, in our study, synthetic urine cytology images are not inferior to real examples when viewed blindly by trained pathology personnel when assessed objectively by error rate and subjective by quality score. While the diagnostic error rate was marginally higher in the synthetic image subset compared to the real images (29.2% versus 26.9%), this was not statistically significant for any of the image observers.

Similarly, the overall subjective image quality score for synthetic images was marginally higher than the real examples — however, this was not statistically significant for any of the image observers. Interestingly, synthetic images were more likely than real images to be selected for inclusion in a teaching set. It must be stated, however, that observers were not restricted in terms of the number of images they were allowed to include in the teaching set and such a restriction may have altered this finding. Importantly, when comparing observers on an individual basis (ie, whether they individually rated real and synthetic images differently), there was no statistically significant difference in the selection of real and synthetic images for the teaching set. Similarly, on an individual basis, there were no significant differences between real and synthetic images in terms of image quality scores and diagnostic errors. This comparison of real and synthetic images on an individual basis is important as it allows for individual biases and preferences to be controlled for to a degree.

Little literature exists regarding the subjective quality assessment of pathology images. Platiša et al[11] investigated the subjective perception of image quality in veterinary pathology images by 3 groups of observers — pathology experts, pathology students, and image experts. Importantly, this study compared the perceived image quality of images that had been manipulated (eg, blurring, addition of noise, and change in color) rather than real versus synthetic images. The authors of this study also used a single stimulus method with a 6-point Likert scale. One of the major findings of this study was that domain experts and image experts emphasize different image parameters when assessing image quality — pathology experts were more sensitive to changes in image color and contrast than image experts, who were most sensitive to image noise. This may be explained by the fact that pathologists know when staining patterns are most realistic given past professional experience. This is imperative to consider when generating synthetic data — realistic staining patterns are an important parameter to monitor. This study[11] also makes the important point that the assessment of pathology image quality should be done by pathology personnel, as undertaken in the present research. Image quality assessment is likely to become important as digital pathology makes inroads in cytology — both for diagnostics and for teaching.

Lévêque et al[10] reviewed the topic of the subjective assessment of medical image quality and the authors make mention of the fact that recruiting expert participants for studies involving image quality assessment is difficult due to existing diagnostic workloads that such observers experience. They state that the use of single stimulus methods, as was used in the present study, may be preferred as they are quicker to perform. The authors also indicate that several factors can influence the subjective assessment of medical images (termed *Influence Factors*). Such factors include environmental factors (eg, variability in viewing environments) and human factors (eg, physical and emotional status). They do, however, point out that medical imaging studies have been conducted in different environments. The present study did not control for the environment in which the participants undertook the image survey, nor did we restrict participants to undertaking the survey at a specific time. As indicated by Lévêque et al,[10] recruiting participants is challenging and, the authors of the present work decided that it is important not to place stringent restrictions on participants but rather to allow them to complete the survey in a realistic setting that mimics routine pathology practice or a teaching setting. Having said this, a few important measures were instituted in this study in an attempt to limit this potential variability: (i) only pathologists or residents who had completed a 6-month cytology rotation were eligible for recruitment; (ii) the order of the image set was randomized at the start of the experiment and participants were presented with the images in either ascending or descending order to control for score drift that may occur over single stimulus sessions[10]; (iii) real and synthetic images were compared on an individual observer basis as well as over the population as a whole.

In previously published pathology literature, the ability of pathology personnel to distinguish between real and synthetic pathology images has mainly centered around Turing tests — where observers are asked to state whether an image is real or synthetic — and predominantly in the setting of histopathology images rather than cytopathology images. Furthermore, a more commonly used technique for the

Downloaded for Anonymous User (n/a) at University of the Witwatersrand Johannesburg from ClinicalKey.com by Elsevier on May 12, 2023. For personal use only. No other uses without permission. Copyright ©2023. Elsevier Inc. All rights reserved.

generation of synthetic pathology images is to use an image-to-image translation method (such as CycleGAN[1]) rather than generating fully synthetic images as has been illustrated in this and our previous work.[2] Using these methods, synthetic images are usually created from existing pathology sections (eg, the generation of a special stain from a haematoxylin and eosin–stained section).

Levy et al[20] demonstrated such a method in a large-scale study based in a real-world setting. In this study, the authors created virtual Masson trichrome stains from haematoxylin and eosin–stained liver biopsies for the assessment of liver fibrosis. The authors make the point that they selected a real-world setting as it is envisioned that this is where this technology will be deployed. Similarly, we decided to use a similar philosophy when undertaking this study. Levy et al[20] demonstrated that similar variability in diagnostics existed between the real and synthetic stains – a similar finding to our diagnostic error rate which did not vary significantly between real and synthetic images. Additionally, Levy et al[20] presented a Turing test where 4 pathologists were asked to identify real and synthetic images. The pathologists were generally not able to do this reliably. Of note, however, is the fact that the authors extend this to study the effect of image resolution on the ability of pathologists to identify synthetic images. At lower resolutions (256 × 256 px), none of the pathologists were able to distinguish between real and synthetic images. At 512 × 512 px, a single pathologist was able to make this distinction and at 1024 × 1024 px, 2 pathologists were able to identify the synthetic images. In our study we generated 512 × 512 px images. This was the size that the initial image patches were cropped at from the WSIs. As was illustrated in our initial work on this topic,[2] the generation of higher resolution images requires larger training sets and larger and faster graphics processing units. Importantly, cytology images at 512 × 512 px are likely to be diagnostically useful as cytology slides are typically examined at a higher magnification than histopathology slides where tissue architecture is a more useful diagnostic feature.

Other studies have attempted to test the ability of human observers to distinguish between real and synthetic images. Senaras et al[21] made use of a Turing test when testing the ability of 3 pathologists and 3 image analysts to detect synthetic Ki67 immunohistochemical stains in breast carcinoma. These observers could not reliably distinguish between real and synthetic images. Wei et al,[22] created synthetic colorectal polyp images using image-to-image translation and tested the ability of 4 pathologists to distinguish real from synthetic images. The ability of pathologists to identify the synthetic images varied by the type of polyps generated – 1 out of 4 for tubular adenomas and 3 out of 4 for sessile serrated lesions.

Alternative approaches to the traditional Turing test have been reported in a pathology-specific context. Quiros et al[23] generated 224 × 224 px breast carcinoma histology images and recruited 2 pathologists to view a set of real and synthetic images (50 images in total) and provided a staining rating from 1 to 5 (5 being the most realistic staining). The authors showed no significant difference between the ratings of real and synthetic images. Additionally, in a supplementary experiment, the pathologists were unable to distinguish synthetic images from real examples. This technique of subjectively rating the staining of the image rather than specifically asking observers to identify the synthetic images is similar to the approach we adopted in this work, except for the fact that our participants were blinded to the presence of synthetic data whereas, in the former study, the pathologists had initially been asked to identify synthetic images in a Turing test prior to the staining quality assessment.

An interesting remark by Quiros et al,[23] is that the pathologists specifically commented on the small size of the images they were being asked to assess. In fact, the low resolution of synthetic pathology images studied to date[2,22-25] is of concern given the findings by Levy et al,[20] suggesting that image resolution plays an important part in the ability of pathologists to distinguish between real and synthetic images. As stated above, the resolution of the images in this study (512 × 512 px) seems an appropriate resolution for cytology at this stage of GANs technology development.

The technique presented in this and, our previous, work[2] may prove useful for training pathologists and cytotechnologists in laboratories with insufficient examples of certain types of cytology specimens or disease types. Although literature addressing this particular point is lacking, it seems likely that inadequate exposure to appropriate teaching material will compromise training and may affect success in examinations and accurate diagnostics in a real-world setting.

Ethical and legal questions around the use of synthetic data are bound to arise going forward, however, as this field is relatively novel in a healthcare context, there is a lack of appropriate legislation and guidelines currently.[26] As it stands, patient and/or institutional permission is required to use pathology specimens to generate data sets.[27] The question as to whether such permission is required once a generative model has been trained remains unanswered. Given the lack of current legislation and guidelines, a cautious ethical approach, ensuring that ethics review boards retain oversight of the use of synthetic medical data, seems appropriate. We suggest that, if the generated synthetic data and its associated trained model are not open source, permission from the institution/company that owns the model would still be necessary.[28]

One issue that may arise from the use of synthetic images is that of trust. Further research needs to be undertaken to investigate whether pathology and cytology teachers will trust synthetic images for training purposes. In this study, where participants were unaware that some images were synthetic, both real and synthetic urine cytology images were chosen for teaching and learning purposes. Whether real or synthetic images would be preferred for training if the teacher knows which images are real or synthetic remains unanswered.

Downloaded for Anonymous User (n/a) at University of the Witwatersrand Johannesburg from ClinicalKey.com by Elsevier on May 12, 2023. For personal use only. No other uses without permission. Copyright ©2023. Elsevier Inc. All rights reserved.

Our study had several limitations that further research could address. Firstly, we had 12 participants — a larger number than most other studies investigating the perception of real versus synthetic pathology images but a larger cohort of respondents may yield different results. Secondly, some of our image observers had limited cytology experience (6 months). We included these respondents to test this technology in a group of people who may use the technology in the future. Furthermore, to control for this, we did not compare observers to one another, instead, we tested real and synthetic images on an individual observer basis (similar to other studies in the pathology literature[20,22,23]). Lastly, our study was limited to urine cytology alone. To test the generalizability of this technology to cytology, further research will be required. Having said this, the extension of this technology to subdomains of cytopathology and histopathology, other than urine cytology, seems a logical next set. Looking forward, the potential of this technology specifically in the setting of teaching/training and quality assurance seems clear. Levy et al[20] demonstrated the ability of GANs to generate entire synthetic WSIs — albeit using image-to-image translation — and the ability to create synthetic WSIs alongside immunocytochemistry/immunohistochemistry and special stains would allow for the creation of large amounts of teaching material. Similarly, GANs technology may allow educators to transform cytology to histology and vice versa to illustrate cytology/histology correlation to learners or to transform benign examples into atypical/dysplastic and malignant examples to illustrate disease progression. While the use of GANs technology in pathology has focused on improving supervised algorithms for diagnostics to date,[1] the unexplored potential for teaching/training and quality assurance testing is beginning to become clear.

## Conclusion

This and our previous work,[2] demonstrate the ability of GANs technology to generate highly realistic urine cytology images. Furthermore, this work, demonstrates that there was no difference in how pathology personnel perceived the subjective quality of these synthetic images nor was there a difference in diagnostic error rates between real and synthetic urine cytology images.

The possible use cases for GANs technology/synthetic data in pathology extend beyond improving supervised classification/segmentation algorithms and the generation of highly realistic synthetic data has enormous potential use in pathology teaching/training and quality assurance testing.

## Author contributions

All 4 authors contributed to the conceptualization of this study, the analysis of the results, and the writing of the manuscript.

## Conflict of interest disclosures

The authors declare no conflicts of interest.

## Statement of ethics

This study involved the use of anonymized, archived cytology slides and human participants. Ethical clearance for this study was granted by Human Research Ethics Committee (Medical) of the University of the Witwatersrand (Certificate number: M190604). Informed consent was obtained from human observers (image observers) prior to their enrolment in the study.

## Acknowledgments

The authors acknowledge the National Health Laboratory Service, Lancet Laboratories, and 3F Scientific for their assistance with providing research material and digitizing the slides in this research, respectively. The authors would also like to extend thanks to Mr. P. Diessel for his assistance in creating the online survey system used in this project. This work has been partially funded by the University of the Witwatersrand, Faculty of Health Sciences Research Equipment grants, and an NVIDIA Academic Hardware Grant.

## References

1. McAlpine ED, Michelow P, Celik T. The utility of unsupervised machine learning in anatomic pathology. *Am J Clin Pathol*. 2022;157:5–14.
2. McAlpine E, Michelow P, Liebenberg E, Celik T. Is it real or not? Towards AI-based realistic synthetic cytology image generation to augment teaching and quality assurance in pathology. *J Am Soc Cytopathol*. 2022;11:123–132.
3. Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks. *IEEE Comput Soc Conf Comput Vis Pattern Recogn*; 2019:4396–4405.
4. Karras T, Aila T, Laine S, Lehtinen J. Progressive growing of GANs for improved quality, stability, and variation. CoRR. Available at: http://arxiv.org/abs/1710.10196; 2017. Accessed December 6, 2018.
5. Karras T, Laine S, Aittala M, Hellsten J, Lehtinen J, Aila T. Analyzing and improving the image quality of stylegan. *IEEE Comput Soc Conf Comput Vis Pattern Recogn*; 2020:8107–8116.
6. Karras T, Aittala M, Laine S, et al. Alias-free generative adversarial networks. *Adv Neural Inf Process Syst*. 2021;34:852–863.
7. Brock A, Donahue J, Simonyan K. Large scale GAN training for high fidelity natural image synthesis. Available at: http://arxiv.org/abs/1809.11096. Accessed September 28, 2018.
8. Bellemo V, Burlina P, Yong L, Wong TY, Ting DSW. Generative adversarial networks (GANs) for retinal fundus image synthesis. In: *Asian Conference on Computer Vision*. Cham: Springer; 2019:289–302.
9. Segal B, Rubin DM, Rubin G, Pantanowitz A. Evaluating the clinical realism of synthetic chest X-rays generated using progressively growing GANs. *SN Comput Sci*. 2021;2:321.
10. Lévêque L, Liu H, Martini M, Outtas M, Zhang L. On the subjective assessment of the perceived quality of medical images and videos.

Downloaded for Anonymous User (n/a) at University of the Witwatersrand Johannesburg from ClinicalKey.com by Elsevier on May 12, 2023. For personal use only. No other uses without permission. Copyright ©2023. Elsevier Inc. All rights reserved.

In: *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE; 2018:1–6.
11. Platiša L, Van Brantegem L, Haeghen YV, Marchessoux C, Vansteenkiste E, Philips W. Psycho-visual evaluation of image quality attributes in digital pathology slides viewed on a medical color LCD display. *Med Imaging*. 2013;8676:184–193.
12. McAlpine ED, Michelow P, Celik T. The dynamics of pathology dataset creation using urine cytology as an example. *Acta Cytol*. 2022;66:46–54.
13. StyleGAN3. Available at: https://github.com/NVlabs/stylegan3. Accessed December 13, 2021.
14. Barkan GA, Wojcik EM, Nayar R, et al. The Paris system for reporting urinary cytology: the quest to develop a standardized terminology. *Acta Cytol*. 2016;60:185–197.
15. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2022. Available at: https://www.R-project.org/.
16. Python. Available at: https://www.python.org/. Accessed February 27, 2020.
17. Kolmogorov-Smirnov Test. *The Concise Encyclopedia of Statistics*. New York: Springer; 2008:283–287.
18. Langr J, Bok V. *GANs in Action*. Shelter Island, NY: Manning Publications Co; 2019.
19. Borji A. Pros and cons of GAN evaluation measures: new developments. *Comput Vis Image Understand*. 2021;215:103329.
20. Levy JJ, Azizgolshani N, Andersen MJ, et al. A large-scale internal validation study of unsupervised virtual trichrome staining technologies on nonalcoholic steatohepatitis liver biopsies. *Mod Pathol*. 2021;34:808–822.
21. Senaras C, Niazi MKK, Sahiner B, et al. Optimized generation of high-resolution phantom images using cGAN: application to quantification of Ki67 breast cancer images. *PLoS One*. 2018;13:e0196846.
22. Wei J, Suriawinata A, Vaickus L, et al. Generative image translation for data augmentation in colorectal histopathology images. *Proc Mach Learn Res*. 2019;116:10–24.
23. Quiros AC, Murray-Smith R, Yuan K. Pathology GAN: learning deep representations of cancer tissue. *Med Phys*. 2021;48:3262–3372.
24. Xu Z, Moro CF, Bozóky B, Zhang Q. GAN-based virtual re-staining: a promising solution for whole slide image analysis. *ArXiv*; 2019:1–16.
25. Hou L, Agarwal A, Samaras D, Kurc TM, Gupta RR, Saltz JH. Robust histopathology image analysis: to label or to synthesize? *IEEE Comput Soc Conf Comput Vis Pattern Recogn*; 2019:8525–8534.
26. Bellovin SM, Dutta PK, Reitinger N. Privacy and Synthetic Datasets. *Stan Tech L Rev*. 2019;22:1–52.
27. McKay F, Williams BJ, Prestwich G, Bansal D, Hallowell N, Treanor D. The ethical challenges of artificial intelligence-driven digital pathology. *J Pathol Clin Res*. 2022;8:209–216.
28. Abels E, Pantanowitz L, Aeffner F, et al. Computational pathology definitions, best practices, and recommendations for regulatory guidance: a white paper from the digital pathology association. *J Pathol*. 2019;249:286–294.

Downloaded for Anonymous User (n/a) at University of the Witwatersrand Johannesburg from ClinicalKey.com by Elsevier on May 12, 2023. For personal use only. No other uses without permission. Copyright ©2023. Elsevier Inc. All rights reserved.

# CHAPTER 7: SUMMARY AND CONCLUSIONS

Our work demonstrates that realistic, morphologically diverse urine cytology images can be generated using existing GANs technology, trained on a relatively small number of images and with limited computing resources and that expert human observers find such synthetic data visually acceptable. Additionally, our data indicate that there is no significant difference in synthetic data in terms of subjective image quality or diagnostic classification as determined by pathology personnel.

This body of work has important implications for pathology education in the future. Synthetic pathology data can be used to supplement real-world training material when such material is limited or when patient privacy concerns prevent the use of real patient records.

Lastly, the introduction of synthetic data in healthcare raises novel ethical and legal issues that will need to be addressed going forward. Given how new the concept of synthetic data is to medicine (and, by extrapolation, pathology) current ethical guidelines and legal frameworks are likely to prove insufficient to address these new challenges. As discussed, although synthetic data offers a potential solution to patient anonymity issues in health research and education, their remains a risk of patient re-identification and responsible use of this technology, governed by appropriate laws and ethical guidelines, is essential to safeguard individual patients, medical practitioners and the profession as a whole.

# APPENDICES

## APPENDIX 1: STUDENT AND CO-AUTHORS DECLARATION

**Declaration: Student's contribution to articles and agreement of co-authors**

I, Ewen David McAlpine, student number 0002220G, declare that this Thesis is my own work and that I contributed adequately towards research findings published in the articles stated below which are included in my Thesis.

**Signature of Student:** E McAlpine.

**Date:** 08/12/2022

**Name of Supervisor: Pamela Michelow**

**Signature of Supervisor:** P Michelow

**Date:** 6/12/2022

**Name of Supervisor: Turgay Celik**

**Signature of Supervisor:** T. Celik

**Date:** 9 December 2022

**Agreement by co-authors:**

By signing this declaration, the co-authors listed below agree to the use of the articles by the student as part of his Thesis.

**Article 1:**

**Title:** The Utility of Unsupervised Machine Learning in Anatomic Pathology
*Am J Clin Pathol*. 2021 Jul 24. https://doi.org/10.1093/ajcp/aqab085.

| Authors | Name | Signature | Date |
|---|---|---|---|
| 1st author | McAlpine, E.D. | E McAlpine. | 08/12/2022 |
| 2nd author | Michelow, P. | P Michelow | 6/12/2022 |
| 3rd author | Celik, T. | T. Celik | 9 December 2022 |

**Comments by supervisor:**

1

**Article 2:**

**Title:** The dynamics of pathology dataset creation using urine cytology as an example
*Acta Cytol.* 2021 Oct 18;1-9. doi: 10.1159/000519273.

| Authors | Name | Signature | Date |
|---|---|---|---|
| 1st author | McAlpine, E.D. | | 08/12/2022 |
| 2nd author | Michelow, P. | | 6/12/2022 |
| 3rd author | Celik, T. | | 9 December 2022 |

**Comments by supervisor:**

**Article 3:**

**Title:** Is it real or not? Towards AI-based realistic synthetic cytology image generation to augment teaching and quality assurance in pathology
*J Am Soc Cytopathol.* 2022 May-Jun;11(3):123-132. doi: 10.1016/j.jasc.2022.02.001.

| Authors | Name | Signature | Date |
|---|---|---|---|
| 1st author | McAlpine, E.D. | | 08/12/2022 |
| 2nd author | Michelow, P. | | 6/12/2022 |
| 3rd author | Liebenberg, E. | | |
| 4th author | Celik, T. | | 9 December 2022 |

**Comments by supervisor:**

**Article 4:**

**Title:** Are synthetic cytology images ready for prime time? A comparative assessment of real and synthetic urine cytology images.
J Am Soc Cytopathol. 2022. Article in press. doi: 10.1016/j.jasc.2022.10.001. Available online: 7 October 2022.

| Authors | Name | Signature | Date |
|---|---|---|---|
| 1st author | McAlpine, E.D. | | 08/12/2022 |
| 2nd author | Michelow, P. | | 6/12/2022 |
| 3rd author | Liebenberg, E. | | |
| 4th author | Celik, T. | | 9 December 2022 |

**Comments by supervisor:**

**Article 2:**

**Title:** The dynamics of pathology dataset creation using urine cytology as an example
*Acta Cytol.* 2021 Oct 18;1-9. doi: 10.1159/000519273.

| Authors | Name | Signature | Date |
|---|---|---|---|
| 1st author | McAlpine, E.D. | | 08/12/2022 |
| 2nd author | Michelow, P. | | 6/12/2022 |
| 3rd author | Celik, T. | | |

**Comments by supervisor:**

**Article 3:**

**Title:** Is it real or not? Towards AI-based realistic synthetic cytology image generation to augment teaching and quality assurance in pathology
*J Am Soc Cytopathol.* 2022 May-Jun;11(3):123-132. doi: 10.1016/j.jasc.2022.02.001.

| Authors | Name | Signature | Date |
|---|---|---|---|
| 1st author | McAlpine, E.D. | | 08/12/2022 |
| 2nd author | Michelow, P. | | 6/12/2022 |
| 3rd author | Liebenberg, E. | | 8/12/2022 |
| 4th author | Celik, T. | | |

**Comments by supervisor:**

**Article 4:**

**Title:** Are synthetic cytology images ready for prime time? A comparative assessment of real and synthetic urine cytology images.
J Am Soc Cytopathol. 2022. Article in press. doi: 10.1016/j.jasc.2022.10.001. Available online: 7 October 2022.

2

| Authors | Name | Signature | Date |
|---|---|---|---|
| 1st author | McAlpine, E.D. | | 08/12/2022 |
| 2nd author | Michelow, P. | | 6/12/2022 |
| 3rd author | Liebenberg, E. | | 8/12/2022. |
| 4th author | Celik, T. | | |

**Comments by supervisor:**

## APPENDIX 2: ETHICS CERTIFICATE

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

R14/49 Dr Ewen David McAlpine

### HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

### CLEARANCE CERTIFICATE NO. M190604

| | |
|---|---|
| **NAME: (Principal Investigator)** | Dr Ewen David McAlpine |
| **DEPARTMENT:** | Division of Anatomical Pathology, School of Pathology |
| **PROJECT TITLE:** | Using deep neural networks to generate synthetic cytology images for training and quality assurance purposes |
| **DATE CONSIDERED:** | 28/06/2019 |
| **DECISION:** | Approved unconditionally |
| **CONDITIONS:** | |
| **SUPERVISOR:** | Prof C. Wright, Prof P. Michelow and Prof T. Celik |
| **APPROVED BY:** | Dr CB Penny, Chairperson, HREC (Medical) |
| **DATE OF APPROVAL:** | 26/07/2019 |

**This clearance certificate is valid for 5 years from date of approval. Extension may be applied for.**

**DECLARATION OF INVESTIGATORS**

To be completed in duplicate and **ONE COPY** returned to the Research Office Secretary on the Third Floor, Faculty of Health Sciences, Phillip Tobias Building, 29 Princess of Wales Terrace, Parktown, 2193, University of the Witwatersrand. I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated, from the research protocol as approved, I/we undertake to resubmit the application to the Committee. **I agree to submit a yearly progress report**. The date for annual re-certification will be one year after the date of convened meeting where the study was initially reviewed. In this case, the study was initially reviewed in **June** and will therefore be due in the month of **June** each year. Unreported changes to the application may invalidate the clearance given by the HREC (Medical).

Principal Investigator Signature Date 06/08/19

**PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES**

## APPENDIX 3: TURNITIN REPORT



ORIGINALITY REPORT

| 15% | 7% | 14% | 2% |
|---|---|---|---|
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

PRIMARY SOURCES

| | | |
|---|---|---|
| 1 | Ewen McAlpine, Pamela Michelow, Eric Liebenberg, Turgay Celik. "Are synthetic cytology images ready for primetime? A comparative assessment of real and synthetic urine cytology images.", Journal of the American Society of Cytopathology, 2022<br>Publication | 4% |
| 2 | Ewen McAlpine, Pamela Michelow, Eric Liebenberg, Turgay Celik. "Is it real or not? Toward artificial intelligence-based realistic synthetic cytology image generation to augment teaching and quality assurance in pathology", Journal of the American Society of Cytopathology, 2022<br>Publication | 3% |
| 3 | www.karger.com<br>Internet Source | 2% |
| 4 | Ewen McAlpine, Pamela Michelow, Eric Liebenberg, Turgay Celik. "Are synthetic cytology images ready for prime time? A comparative assessment of real and synthetic | 1% |

Supervisor : Pamela Michelow

13/1/2023

Turgay Celik

15 January 2023

103

# APPENDIX 4: JOURNAL PERMISSIONS



## KARGER PUBLISHERS LICENSE TERMS AND CONDITIONS

May 22, 2023

This Agreement between Dr. Ewen McAlpine ("You") and Karger Publishers ("Karger Publishers") consists of your license details and the terms and conditions provided by Karger Publishers and Copyright Clearance Center.

| | |
|---|---|
| License Number | 5543461015134 |
| License date | May 07, 2023 |
| Licensed Content Publisher | Karger Publishers |
| Licensed Content Publication | Acta Cytologicaf |
| Licensed Content Title | The Dynamics of Pathology Dataset Creation Using Urine Cytology as an Example |
| Licensed copyright line | Copyright © 2022, Silverchair Publisher |
| Licensed Content Author | McAlpine, Ewen David; Michelow, Pamela M. |
| Licensed Content Date | Jan 6, 2022 |
| Licensed Content Volume | 66 |
| Licensed Content Issue | 1 |
| I would like to... | Website/Intranet |
| Requestor type | editor/author of the requested content |
| Format | Electronic |
| Portion | abstract |
| In the following language(s) | Original language of publication |
| Posting duration | over 12 months |
| Home page URL | https://wiredspace.wits.ac.za/handle/10539/45 |
| Client / Sponsor | Private - Electronic PhD Thesis |
| Expected posting date | Jul 2023 |
| The requesting person/organization | Ewen McAlpine |
| Requestor Location | Dr. Ewen McAlpine<br>7 York Road<br>Parktown<br><br>Johannesburg, 2193<br>South Africa<br>Attn: Dr. Ewen McAlpine |
| Publisher Tax ID | 980406204 |
| Total | **0.00 USD** |
| Terms and Conditions | |

**STANDARD TERMS AND CONDITIONS FOR REPRODUCTION OF MATERIAL**

**Introduction**

The Publisher for this copyrighted material is Karger Publishers. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your CCC account and that are available at any time at http://myaccount.copyright.com.

**Limited License**

KARGER PUBLISHERS LICENSE
TERMS AND CONDITIONS

May 07, 2023

---

This Agreement between Dr. Ewen McAlpine ("You") and Karger Publishers ("Karger Publishers") consists of your license details and the terms and conditions provided by Karger Publishers and Copyright Clearance Center.

| | |
|---|---|
| License Number | 5543461272487 |
| License date | May 07, 2023 |
| Licensed Content Publisher | Karger Publishers |
| Licensed Content Publication | Acta Cytologicaf |
| Licensed Content Title | The Dynamics of Pathology Dataset Creation Using Urine Cytology as an Example |
| Licensed copyright line | Copyright © 2022, Silverchair Publisher |
| Licensed Content Author | McAlpine, Ewen David; Michelow, Pamela M. |
| Licensed Content Date | Jan 6, 2022 |
| Licensed Content Volume | 66 |
| Licensed Content Issue | 1 |
| Type of Use | Website/Intranet |
| Requestor type | editor/author of the requested content |

| | |
|---|---|
| Format | Electronic |
| Portion | figures/tables/illustrations |
| Number of figures/tables /illustrations | 4 |
| Include Image file | no |
| In the following language(s) | Original language of publication |
| Posting duration | over 12 months |
| Home page URL | https://wiredspace.wits.ac.za/handle/10539/45 |
| Client / Sponsor | Private - Electronic PhD Thesis |
| Expected posting date | Jul 2023 |
| Portions | Fig 3, Table 3, Fig 5, Table 4 |
| The requesting person/organization | Ewen McAlpine |
| Requestor Location | Dr. Ewen McAlpine<br>7 York Road<br>Parktown<br><br>Johannesburg, 2193<br>South Africa<br>Attn: Dr. Ewen McAlpine |
| Publisher Tax ID | 980406204 |
| Total | 0.00 USD |

ELSEVIER LICENSE
TERMS AND CONDITIONS

May 07, 2023

---

This Agreement between Dr. Ewen McAlpine ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

| | |
|---|---|
| License Number | 5543461420754 |
| License date | May 07, 2023 |
| Licensed Content Publisher | Elsevier |
| Licensed Content Publication | Journal of the American Society of Cytopathology |
| Licensed Content Title | Is it real or not? Toward artificial intelligence-based realistic synthetic cytology image generation to augment teaching and quality assurance in pathology |
| Licensed Content Author | Ewen McAlpine,Pamela Michelow,Eric Liebenberg,Turgay Celik |
| Licensed Content Date | May–June 2022 |
| Licensed Content Volume | 11 |
| Licensed Content Issue | 3 |

| | |
|---|---|
| Licensed Content Pages | 10 |
| Start Page | 123 |
| End Page | 132 |
| Type of Use | post on a website |
| Requestor type | academic/educational institute |
| Portion | abstract |
| Format | electronic |
| Are you the author of this Elsevier article? | Yes |
| Will you be translating? | No |
| Home page URL | https://wiredspace.wits.ac.za/handle/10539/45 |
| Client / Sponsor | Private - Electronic PhD Thesis |
| Expected posting date | Jul 2023 |
| Requestor Location | Dr. Ewen McAlpine<br>7 York Road<br>Parktown<br><br>Johannesburg, 2193<br>South Africa<br>Attn: Dr. Ewen McAlpine |
| Publisher Tax ID | ZA 4110266048 |

ELSEVIER LICENSE
TERMS AND CONDITIONS

May 07, 2023

---

This Agreement between Dr. Ewen McAlpine ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

| | |
|---|---|
| License Number | 5543470028596 |
| License date | May 07, 2023 |
| Licensed Content Publisher | Elsevier |
| Licensed Content Publication | Journal of the American Society of Cytopathology |
| Licensed Content Title | Is it real or not? Toward artificial intelligence-based realistic synthetic cytology image generation to augment teaching and quality assurance in pathology |
| Licensed Content Author | Ewen McAlpine,Pamela Michelow,Eric Liebenberg,Turgay Celik |
| Licensed Content Date | May–June 2022 |
| Licensed Content Volume | 11 |
| Licensed Content Issue | 3 |
| Licensed Content Pages | 10 |

| | |
|---|---|
| Start Page | 123 |
| End Page | 132 |
| Type of Use | post on a website |
| Requestor type | academic/educational institute |
| Portion | figures/tables/illustrations |
| Number of figures/tables /illustrations | 1 |
| Format | electronic |
| Are you the author of this Elsevier article? | Yes |
| Will you be translating? | No |
| Home page URL | https://wiredspace.wits.ac.za/handle/10539/45 |
| Client / Sponsor | Private - Electronic PhD Thesis |
| Expected posting date | Jul 2023 |
| Portions | Fig 1 |
| Requestor Location | Dr. Ewen McAlpine<br>7 York Road<br>Parktown<br><br>Johannesburg, 2193<br>South Africa<br>Attn: Dr. Ewen McAlpine |

ELSEVIER LICENSE
TERMS AND CONDITIONS

May 07, 2023

---

This Agreement between Dr. Ewen McAlpine ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

| | |
|---|---|
| License Number | 5543640824125 |
| License date | May 07, 2023 |
| Licensed Content Publisher | Elsevier |
| Licensed Content Publication | Journal of the American Society of Cytopathology |
| Licensed Content Title | Are synthetic cytology images ready for prime time? A comparative assessment of real and synthetic urine cytology images |
| Licensed Content Author | Ewen McAlpine,Pamela Michelow,Eric Liebenberg,Turgay Celik |
| Licensed Content Date | March–April 2023 |
| Licensed Content Volume | 12 |
| Licensed Content Issue | 2 |
| Licensed Content Pages | 10 |

| | |
|---|---|
| Start Page | 126 |
| End Page | 135 |
| Type of Use | post on a website |
| Requestor type | academic/educational institute |
| Portion | abstract |
| Format | electronic |
| Are you the author of this Elsevier article? | Yes |
| Will you be translating? | No |
| Home page URL | https://wiredspace.wits.ac.za/handle/10539/45 |
| Client / Sponsor | Private - Electronic PhD Thesis |
| Expected posting date | Jul 2023 |
| Requestor Location | Dr. Ewen McAlpine<br>7 York Road<br>Parktown<br><br>Johannesburg, 2193<br>South Africa<br>Attn: Dr. Ewen McAlpine |
| Publisher Tax ID | ZA 4110266048 |
| Total | 0.00 USD |
| Terms and Conditions | |

ELSEVIER LICENSE
TERMS AND CONDITIONS

May 07, 2023

This Agreement between Dr. Ewen McAlpine ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

| | |
|---|---|
| License Number | 5543640984232 |
| License date | May 07, 2023 |
| Licensed Content Publisher | Elsevier |
| Licensed Content Publication | Journal of the American Society of Cytopathology |
| Licensed Content Title | Are synthetic cytology images ready for prime time? A comparative assessment of real and synthetic urine cytology images |
| Licensed Content Author | Ewen McAlpine,Pamela Michelow,Eric Liebenberg,Turgay Celik |
| Licensed Content Date | March–April 2023 |
| Licensed Content Volume | 12 |
| Licensed Content Issue | 2 |
| Licensed Content Pages | 10 |
| Start Page | 126 |

| | |
|---|---|
| End Page | 135 |
| Type of Use | post on a website |
| Requestor type | academic/educational institute |
| Portion | figures/tables/illustrations |
| Number of figures/tables /illustrations | 2 |
| Format | electronic |
| Are you the author of this Elsevier article? | Yes |
| Will you be translating? | No |
| Home page URL | https://wiredspace.wits.ac.za/handle/10539/45 |
| Client / Sponsor | Private - Electronic PhD Thesis |
| Expected posting date | Jul 2023 |
| Portions | Fig 1 and Fig 4 |
| Requestor Location | Dr. Ewen McAlpine<br>7 York Road<br>Parktown<br><br>Johannesburg, 2193<br>South Africa<br>Attn: Dr. Ewen McAlpine |
| Publisher Tax ID | ZA 4110266048 |

# ELSEVIER LICENSE
# TERMS AND CONDITIONS

May 22, 2023

This Agreement between Dr. Ewen McAlpine ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

| | |
|---|---|
| License Number | 5542060809921 |
| License date | May 04, 2023 |
| Licensed Content Publisher | Elsevier |
| Licensed Content Publication | Journal of the American Society of Cytopathology |
| Licensed Content Title | Is it real or not? Toward artificial intelligence-based realistic synthetic cytology image generation to augment teaching and quality assurance in pathology |
| Licensed Content Author | Ewen McAlpine,Pamela Michelow,Eric Liebenberg,Turgay Celik |
| Licensed Content Date | May–June 2022 |
| Licensed Content Volume | 11 |
| Licensed Content Issue | 3 |
| Licensed Content Pages | 10 |
| Start Page | 123 |
| End Page | 132 |
| Type of Use | post on a website |
| Requestor type | academic/educational institute |
| Portion | full article |
| Format | electronic |
| Are you the author of this Elsevier article? | Yes |
| Will you be translating? | No |
| Home page URL | https://wiredspace.wits.ac.za/handle/10539/45 |
| Client / Sponsor | Private - Electronic PhD Thesis |
| Expected posting date | Jul 2023 |
| Requestor Location | Dr. Ewen McAlpine<br>7 York Road<br>Parktown<br><br>Johannesburg, 2193<br>South Africa<br>Attn: Dr. Ewen McAlpine |
| Publisher Tax ID | ZA 4110266048 |
| Billing Type | Invoice |
| Billing Address | Dr. Ewen McAlpine<br>7 York Road<br>Parktown<br><br>Johannesburg, South Africa 2193<br>Attn: Dr. Ewen McAlpine |
| Total | **0.00 USD** |

# ELSEVIER LICENSE
# TERMS AND CONDITIONS

May 22, 2023

This Agreement between Dr. Ewen McAlpine ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

| | |
|---|---|
| License Number | 5542060989992 |
| License date | May 04, 2023 |
| Licensed Content Publisher | Elsevier |
| Licensed Content Publication | Journal of the American Society of Cytopathology |
| Licensed Content Title | Are synthetic cytology images ready for prime time? A comparative assessment of real and synthetic urine cytology images |
| Licensed Content Author | Ewen McAlpine,Pamela Michelow,Eric Liebenberg,Turgay Celik |
| Licensed Content Date | March–April 2023 |
| Licensed Content Volume | 12 |
| Licensed Content Issue | 2 |
| Licensed Content Pages | 10 |
| Start Page | 126 |
| End Page | 135 |
| Type of Use | post on a website |
| Requestor type | academic/educational institute |
| Portion | full article |
| Format | electronic |
| Are you the author of this Elsevier article? | Yes |
| Will you be translating? | No |
| Home page URL | https://wiredspace.wits.ac.za/handle/10539/45 |
| Client / Sponsor | Private - Electronic PhD Thesis |
| Expected posting date | Jul 2023 |
| Requestor Location | Dr. Ewen McAlpine<br>7 York Road<br>Parktown<br><br>Johannesburg, 2193<br>South Africa<br>Attn: Dr. Ewen McAlpine |
| Publisher Tax ID | ZA 4110266048 |
| Billing Type | Invoice |
| Billing Address | Dr. Ewen McAlpine<br>7 York Road<br>Parktown<br><br>Johannesburg, South Africa 2193<br>Attn: Dr. Ewen McAlpine |
| Total | **0.00 USD** |