

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

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.