Anonymising Pathology Data Using Generative Adversarial Networks

David Morrison and David Harris-Birtill
School of Computer Science, University of St Andrews

INTRODUCTION

Anonymising medical data for use in machine learning is important to preserve patient privacy and, in many circumstances, is a requirement before data can be made available[1]. One approach to anonymising image data is to train a generative model to produce data that is statistically similar to the input data and use the synthetic data in place of the real. In this work, we present a study of the effects of such a process on an exemplar downstream task, histology image classification.

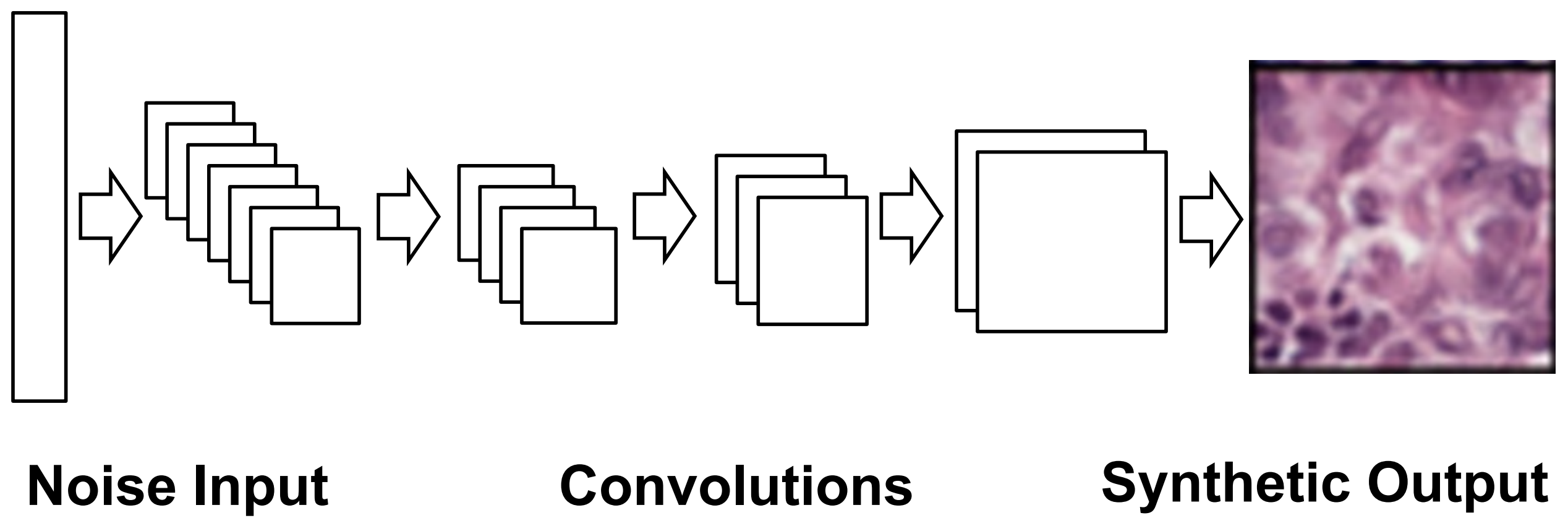

Figure 1 – Deep Convolutional Generative Adversarial Network

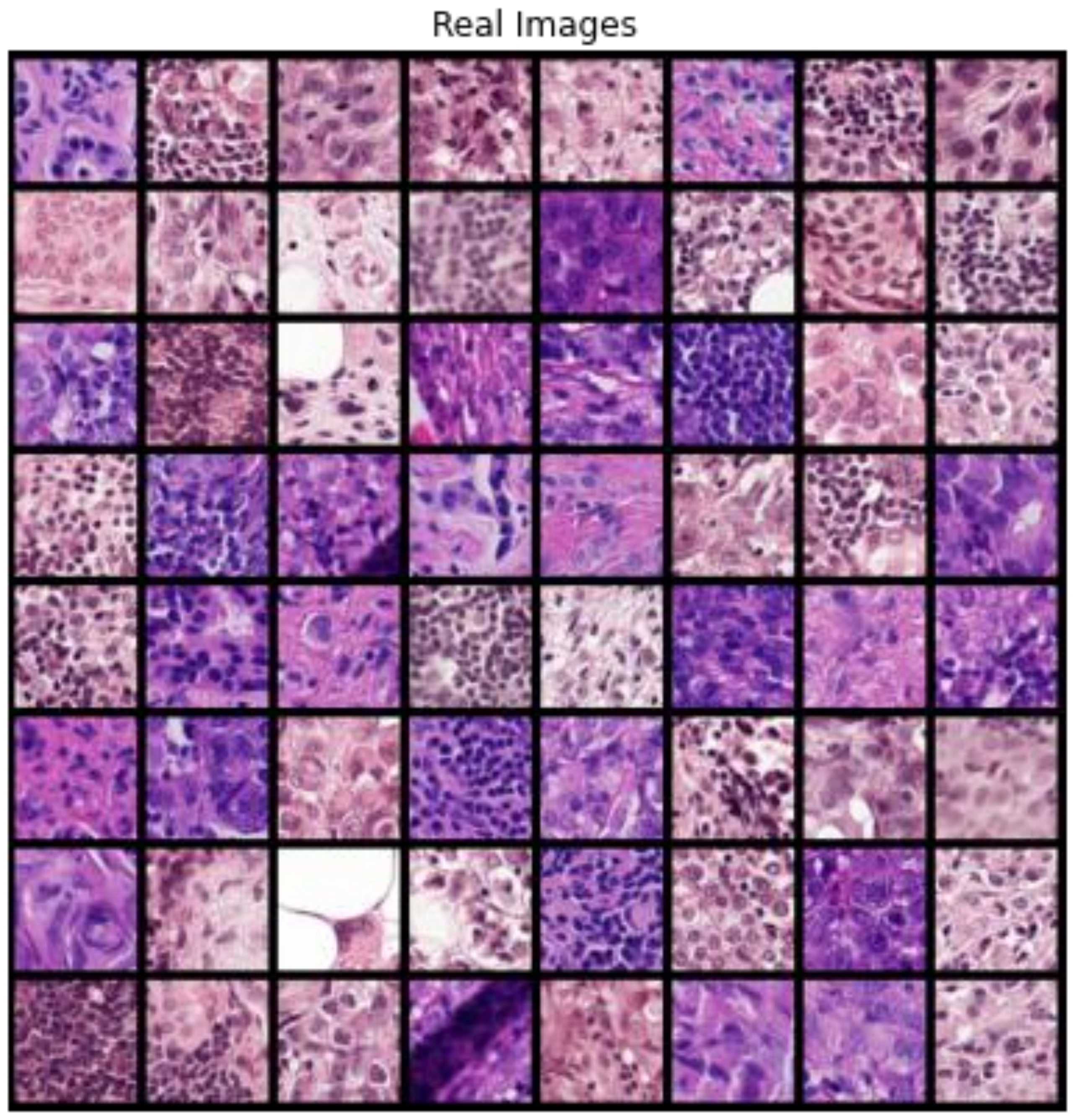

Figure 2 – 64 tumor patches extracted from the Camelyon16 data set.

METHODS

Three identical classification networks were trained. Each was a simple six-layer convolutional neural network. These were labelled:

- **Original** – trained with 70,000 patches extracted from the Camelyon16 dataset.
- **Mixed** – trained with 35,000 patches extracted from the Camelyon16 dataset and 35,000 synthetic patches generated by our DCGAN.
- **Synthetic** – trained on 70,000 synthetic patches.

Independent validation and test sets were used, each consisting of patches extracted from Camelyon16. The synthetic patches were generated using two separate Deep Convolutional Generative Adversarial Network (DCGAN[3]), , one synthesizing tumor patches and the other synthesizing normal patches. Figures 2 and 3 show examples of real tumor patches versus synthetic.

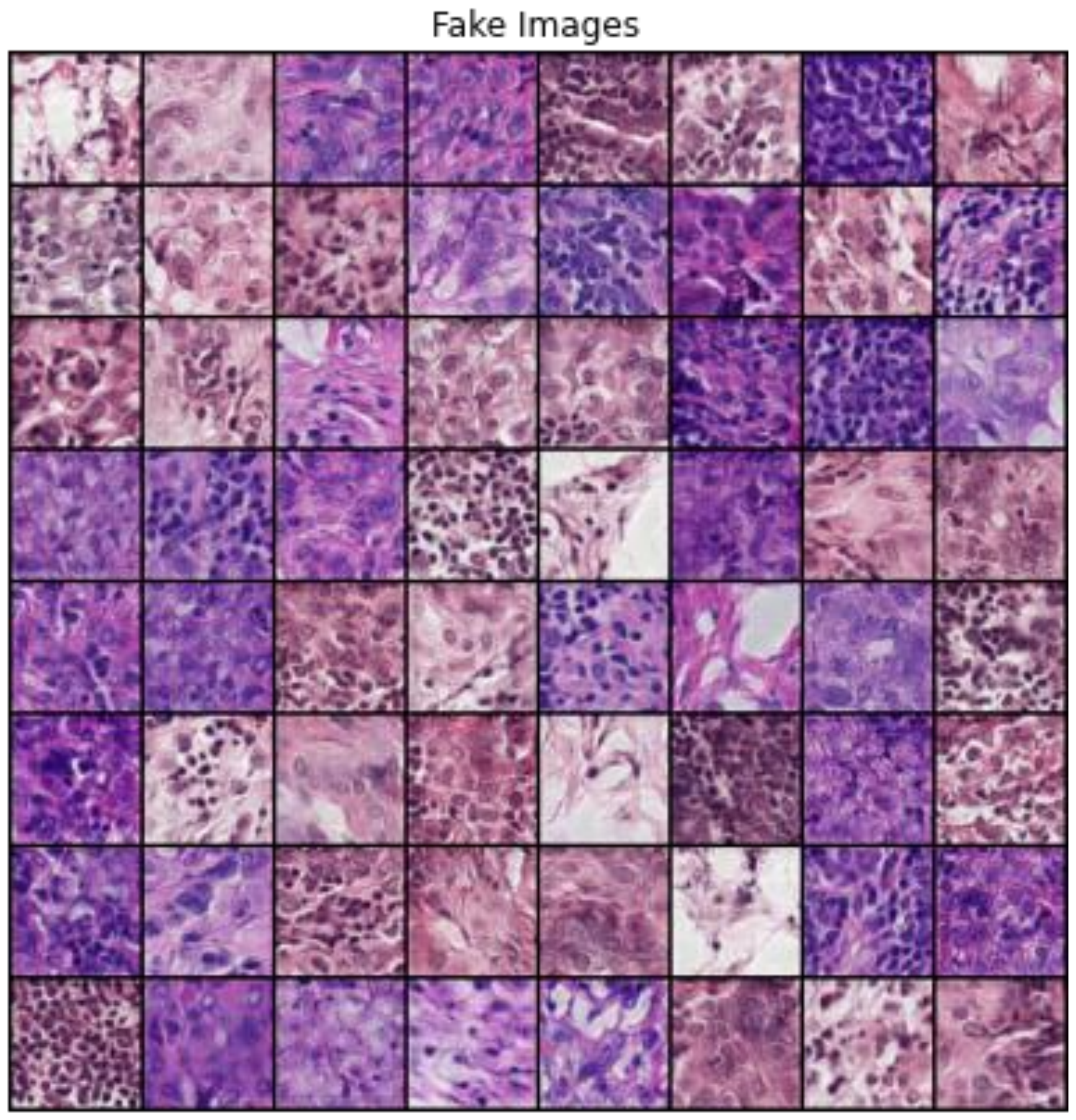

Figure 3 – 64 synthetic tumor patches generated using out DCGAN.

RESULTS

When predicting the class of an image patch as either cancer or normal it's shown that the accuracy reduces from 0.78 for original alone to 0.59 for synthetic alone, and the recall is significantly reduced from 0.70 to 0.44 when training exclusively on the same amount of synthetic data.

Model	Accuracy	Precision	Recall	F1-Score
Original	0.78	0.83	0.70	0.76
Mixed	0.75	0.82	0.65	0.73
Synthetic	0.59	0.64	0.44	0.52

CONCLUSIONS

1. **Classification accuracy was reduced.**
2. **This method is not suitable for anonymization.**
3. **More sophisticated data synthesis is required.**

If retaining a similar accuracy is required for the downstream task, then either the original data must be used or an improved anonymisation strategy must be devised. We conclude that using this DCGAN to anonymise the dataset, degrades the accuracy of the classifier which implies that it has failed to capture the required variation in the original data to generalise and act as a sufficient anonymisation strategy.

Future work will investigate improving the generative model and simulation as a synthesis technique.

REFERENCES

1. **Patel, M., Looney, P., Young, K., and Halling-Brown, M., "Automated collection of medical images for research from heterogeneous systems: trials and tribulations," in [Medical Imaging 2014: PACS and Imaging Informatics: Next Generation and Innovations], 9039, 90390C, International Society for Optics and Photonics (2014).**
2. **Maximov, M., Elezi, I., and Leal-Taix´e, L., "Ciagan: Conditional identity anonymization generative adversarial networks," in [Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition], 5447-5456 (2020).**
3. **Radford, A., Metz, L., and Chintala, S., "Unsupervised representation learning with deep convolutional generative adversarial networks," arXiv preprint arXiv:1511.06434 (2015).**