Enhancing Medical Image Classification with Diffusion-Based Synthetic Data

Samuel Bohumel*
Supervised by: Ing. Maroš Kollár†

Faculty of Informatics and Information Technologies
Slovak University of Technology
Bratislava / Slovakia

Abstract

Neural networks and deep learning are now widely used approaches for solving tasks in multiple domains, including computer vision. In the field of medical image processing, these approaches can bring efficient and fast diagnosis. However, there is a challenge associated with the lack of annotated training data needed to train the models. The collection and especially the annotation of such data can be time-consuming and expensive. In this work, we explore the use of generative models for medical data synthesis that could effectively complement existing training sets and improve the performance of classification models. The main area of our research is the synthesis of atypical cells, which is the main signal of a tumor. Nuclear atypia is usually manifested by enlarged cells and irregular shapes, which are the features we focus on. We take advantage of diffusion probabilistic models that are used for guided synthesis of samples either from a segmentation mask or an atypia class. This research contributes to the integration of machine learning techniques in healthcare and evaluates the presence of synthetic data in training sets.

1 Introduction

Breast carcinoma is one of the most common breast cancer variants, causing deaths in million-scale across the world every year. The main procedure in grading this cancer is nuclear atypia scoring. In pathology, the term "atypia" describes a deviation from the typical or normal morphology. Atypia describes unusual cell alterations that can impact stromal, epithelial, myoepithelial, and other connective tissue cells, including endothelial cells. An essential component of the histological grading system for both in situ[1] and invasive carcinomas is the classification of epithelial cell atypia as low, intemediate, or high. Score 1 describes nuclei that are similar in size of benign epithelial cell nuclei[5] (less than 1.5 times their size) and show minimal pleomorphism (Pleomorphism is the variation in cell or nuclear shape and size.) that are either invisible or barely noticeable. Score 2 nuclei are moderately larger (1.5–2 times the size of benign nuclei) and exhibit mild to moderate pleomorphism with visible but small and subtle nucleoli. Score 3 nuclei are significantly larger (greater than twice the size of benign nuclei) and display vesicular chromatin, pronounced pleomorphism, and prominent nucleoli. While these characteristics collectively help determine the degree of atypia in invasive breast carcinoma (IBC) and other areas of breast pathology, evaluating individual features remains subjective, and the relative weight of each characteristic is not clearly established. Currently, this is done manually by pathologists, meaning there is susceptibility to errors and disagreement between specialists. Therefore, there is a strong motivation for developing automated methods to solve the problem of manual scoring.

Recent advances in machine learning, particularly generative neural networks, offer the potential to augment histological datasets with synthetic images, reducing reliance on manual annotations and improving consistency in analysis. This is especially important in histology, the study of the microscopic structure of tissues and cells, where detailed examination of histological sections provides critical insights into cellular composition, organization, and spatial relationships [8].

Progress in computational pathology, prognostic evaluations, and computer-aided diagnosis could greatly benefit from automated histological image analysis. However, training neural networks for such tasks requires large, well-annotated datasets [3]. A major challenge in applying deep learning methods in this domain is the scarcity of publicly available histological data. Generative approaches, by synthesizing realistic and annotated histological images, offer a promising solution to this limitation, enabling the development of more robust and reliable machine learning models.

The main goal of this research is to evaluate how well generative neural networks produce synthetic annotated

*xbohumel@stuba.sk
†mt@cescg.org

[1] "In situ" describes a condition where abnormal cells are present but have not spread to nearby tissues. Ductal carcinoma in situ (DCIS) is a type of breast cancer that is contained within the milk ducts and has not spread beyond them.

data that may enhance real histological datasets and enhance model performance. In particular, the research effort intends to use performance indicators to compare the results of neural networks trained on these created datasets with those learned on standard datasets.

2 Related work

Deshpande et al.[4] introduced an interactive framework for generating tissue images based on user-defined gland layouts. The system utilizes generative adversarial networks for tissue image synthesis and latent diffusion models for glandular mask generation. By incorporating user input for gland positions and sizes, the framework produces realistic tissue structures, considering cancer type, and employs a VQ-VAE decoder for refining glandular masks.

Cong et al.[2] proposed a Color Adaptive Generative Adversarial Network (CAGAN) for stain normalization in histopathology image analysis. Their method leverages semi-supervised learning with consistency regularization and co-training, enabling the model to learn from unlabeled data. The results demonstrate that CAGAN improves color consistency with the target domain, enhancing classification performance.

The study Synthesis of diagnostic quality cancer pathology images by generative adversarial networks[7] evaluates the use of Progressive GAN (ProGAN) to synthesize high-resolution pathology images for 10 cancer types. ProGAN starts with low-resolution image and iteratively adds details as training continues. A survey of pathologists confirmed that synthetic images were visually hard to distinguish from real ones and achieved comparable classification accuracy. Furthermore, adding synthetic images to training datasets enhanced the performance of deep learning models, highlighting their potential for AI training in rare cancer diagnosis, quality assurance, and competence testing.

3 Solution Design

In this section we specify details of our approaches - data preparation, design of our experiments and evaluation.

3.1 Data preprocessing

For training, we used a publicly available dataset. The MITOS-ATYPIA-14 dataset[1] contains an annotated set of breast cancer biopsy (Figure 1) collected by the team of Professor Frédérique Capron at Pitié-Salpêtrière Hospital in Paris, France. Frames are used for nuclear atypia scoring by three levels. Starting from score 1, which means low grade atypia, to moderate grade atypia annotated as score 2 and high grade atypia labeled as score 3.

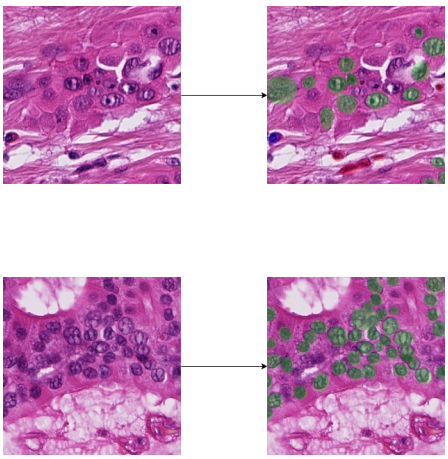

Figure 1: MITOS-ATYPIA-14 tissue sample. Regions highlighted with green color are nuclei.

These frames have resolution 1539 × 1376 pixels. Firstly, each frame was cut to several smaller images according to the resolution of output images, in our case 128 or 256, without overlapping. Each frame has a corresponding class, which was inherited from the original picture. Segmentation masks are needed for a model's input, therefore we needed to create these masks to get information about sizes and positions of nuclei in each sample. Utilizing the HD-Yolo algorithm, this process facilitates nuclei segmentation across whole slides. It includes a pretrained object detection and segmentation model[9] that processes image patches, identifies and segments nuclei. This pretrained model outputs segmentation masks with information about nuclei locations and shapes. These masks are used as input in our second generative model.

3.2 Classification network

Multiple methods were investigated in order to efficiently synthesize annotated histological data. To compare performance of different models and quality of original and enriched datasets, we trained a classification network to assign atypia class to every image. In this case, we used EfficientNet, more specifically EfficientNetV2[11], which is a family of convolutional neural networks, providing a good parameter efficiency and training speed compared to other known architectures, while maintaining similar accuracy. For nuclei synthesis we used denoising diffusion probabilistic models[12] with U-net architecture, but with two distinct approaches.

3.3 Training model per atypia class

In generative neural networks, diffusion models—which draw inspiration from non-equilibrium thermodynamics—have shown great potential, especially for image generation. These models simulate the diffusion process, in which realistic samples are produced by reconstructing

data in reverse after it has been gradually converted into noise by a forward process. Diffusion models, in contrast to conventional generative models like GANs, use a Markov chain of iterative refinements, which enables them to accurately represent complex data distributions. The computational inefficiency of denoising diffusion probabilistic models (DDPMs), that require thousands of repeated steps to produce a single high-quality sample, is an important drawback. Despite promising better output quality, this method is much slower than GANs, which makes it impractical for real-time applications. Although this drawback, diffusion models have become useful in domains where accuracy and variety of data are essential, such as medical imaging.

The first model based on U-net architecture[10] generates nuclei images based on the atypia class. Since this model has no input and was trained on a specific class of images, a separate model was trained for each of the three atypia classes. Results from these models were evaluated in our first experiment, which involved incorporating synthetic data into the training set of a classification model and comparing performance metrics.

First, raw images are split into smaller samples, and annotations are assigned. A classification model is trained on the original dataset, and its performance metrics are recorded. Simultaneously, a generative model is trained to synthesize new samples, which are then used to create an enriched dataset. A second classification model is trained on this expanded dataset, and its performance is compared to the original. The goal is to determine whether synthetic images improve classification accuracy and model robustness. Detailed diagram of this approach is in the Figure 2.

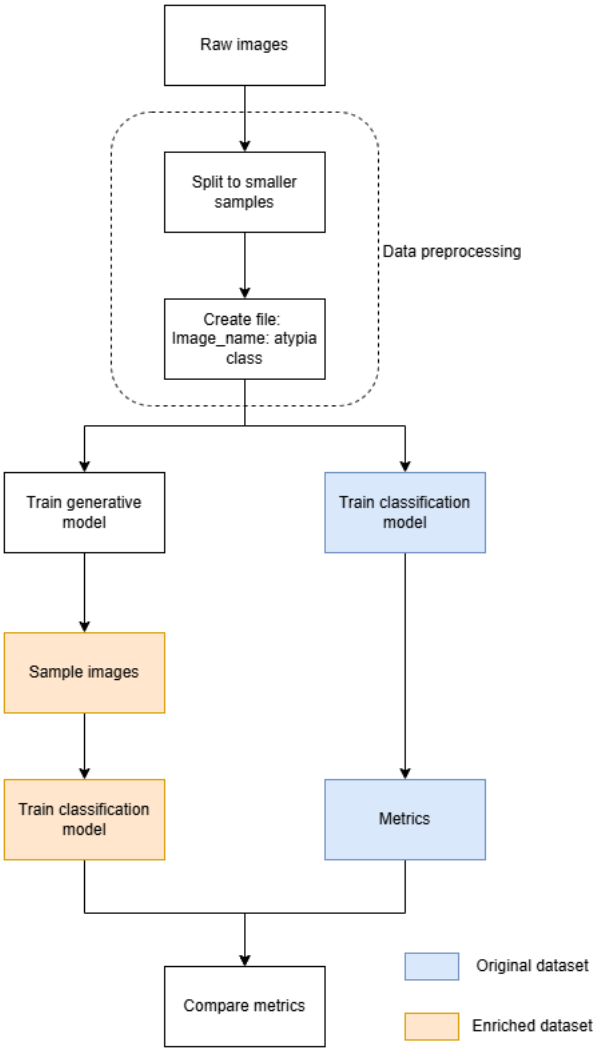

Figure 2: This experiment evaluates the impact of synthetic data on classification performance.

3.4 Segmentaion guided synthesis

Another approach for synthesizing annotated images is to train a network that takes a segmentation mask as input to influence shape, size, and position on a frame[6] (Figure 3. The model learns to reconstruct images based on provided segmentation maps rather than producing images unconditionally, ensuring that structures match the intended annotations. The diffusion model is guided to adjust anatomical details by concatenating the segmentation mask with the input at each denoising step during training. This allows the network to generate realistic histological images while preserving spatial relationships dictated by the mask. Table 1 presents the evaluation metrics. The Frechet Inception Distance (FID) was used to assess the similarity between synthetic and real images, while the Intersection over Union (IoU) measured the consistency between the input mask and the mask generated from the synthetic image.

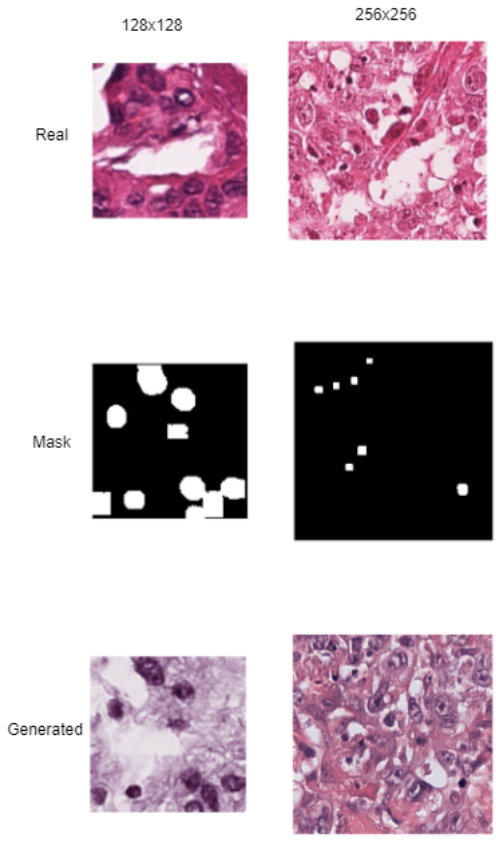

Figure 3: Segmentation-guided image generation: From original sample we created segmentation mask, which was provided to network. Last image shows result sample from the same mask.

4 Results

Results from the dataset enrichment experiment are presented in Figure 4. The experiment began with approximately 70,000 images, and in each of the 14 iterations, 8,000 synthetic images were incrementally added to the dataset. Evaluation metrics were recorded at the end of each training phase. The results indicate an improvement in performance, with an optimal balance observed when synthetic samples constituted arround 50% of the whole training set.

In the Figure 5 is a complete pipeline for diffusion guided by segmentation masks. Firstly, we created segmentation masks for all input samples. Then, we trained a model to synthesize an image with nuclei to match the input mask. We trained models with image sizes of 128x128 and 256x256. Evaluation was conducted using Frechet Inception Distance and Intersection over Union, as further detailed in Figure 5.

Resolution	Epochs	Batch Size	FID↓	IoU↑
128x128	300	8	133.12	0.182
256x256	300	8	91.13	0.174

Table 1: Comparison of training results for different resolutions.

5 Conclusion and Future Work

This study's main goals were to create models for the synthesis of annotated histology data and determine how artificial data augmentation affected classification accuracy. The study focuses on nuclear characteristics, such as enlarged nuclear size and irregular shapes, as important markers of cellular abnormality. We selected diffusion models and used the MITOS-ATYPIA-14 datasets due to their ability to produce high-quality synthetic images. Nuclear atypia was classified using the EfficientNetV2 classification network, which was used as a benchmark to measure the quality of the dataset. We investigated two synthesis strategies: one that used segmentation masks to control nuclear morphology, and another that used diffusion models trained per atypia class. Based on the experiments, the presence of synthetic samples had a limited but noticeable impact on classification performance. This impact may have been caused by subjective pathologist annotations and small inter-class differences. While highlighting the necessity of careful dataset annotation, these findings also demonstrate the potential of synthetic histological data to enhance deep learning applications for pathology.

In the near future, we plan to analyze entire slide images from the Faculty of Medicine, Comenius University Bratislava. In addition to these samples, we have received annotations from three different pathologists. These annotations define various cellular features, including nuclear

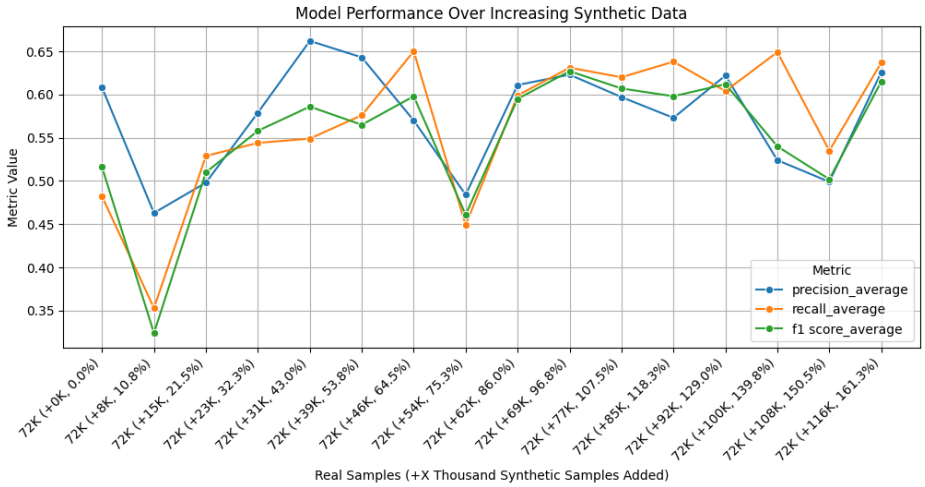

Figure 4: The first experiment, focused on synthesizing images based on atypia class, was evaluated using a classification network. Initially, the network was trained exclusively on real samples. Subsequently, synthetic samples were incrementally introduced into the training set.

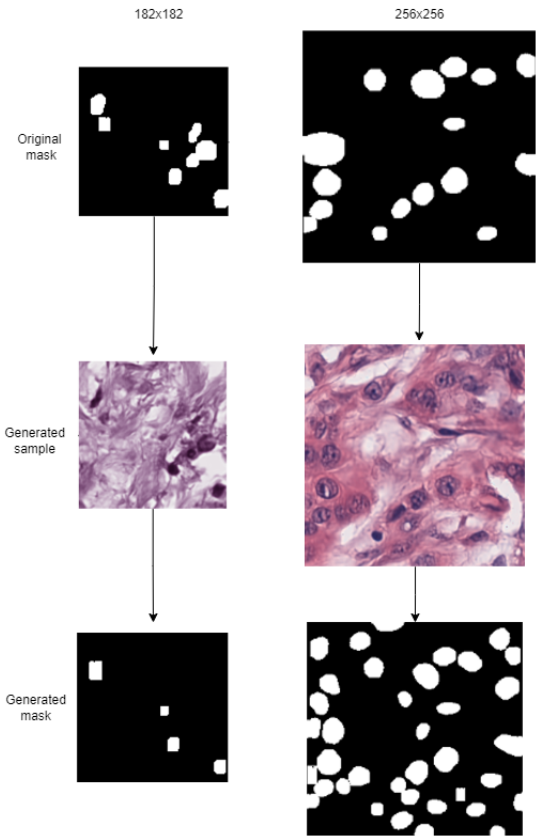

Figure 5: Segmentation-guided image generation - evaluation.
A sample was synthesized using our pre-trained model from input mask. From this sample, another mask was created. By comparison of two masks, metric Interception over Union was used to determine the model's ability to place nuclei on specific locations.

deformations (pleomorphism), cell division (mitosis), and the formation of trabeculae.

Our first step is to extract the annotated regions and group the data based on the type of annotation. In addition, we will use the remaining regions to obtain regular samples for comparison. After preprocessing, our aim was to train a model capable of synthesizing individual nuclei, both regular and exhibiting varying degrees of atypia.

References

[1] Challenges mitos-atypia-14 dataset. https://mitos-atypia-14.grand-challenge.org/Dataset/.

[2] Cong Cong, Sidong Liu, Antonio Di Ieva, Maurice Pagnucco, Shlomo Berkovsky, and Yang Song. Colour adaptive generative networks for stain normalisation of histopathology images. *Medical Image Analysis*, 82:102580, 2022.

[3] Gagandeep B. Daroach, Savannah R. Duenweg, Michael Brehler, Allison K. Lowman, Kenneth A. Iczkowski, Kenneth M. Jacobsohn, Josiah A. Yoder, and Peter S. LaViolette. Prostate cancer histology synthesis using stylegan latent space annotation. In Linwei Wang, Qi Dou, P. Thomas Fletcher, Stefanie Speidel, and Shuo Li, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*, pages 398–408, Cham, 2022. Springer Nature Switzerland.

[4] Srijay Deshpande, Fayyaz Minhas, and Nasir Rajpoot. Synthesis of annotated colorectal cancer tissue images from gland layout, 2024.

[5] Lucy Fortson. Chapter 10 - from green peas to steve: Citizen science engagement in space science. In Amy Paige Kaminski, editor, *Space Science and Public Engagement*, pages 185–219. Elsevier, 2021.

[6] Nicholas Konz, Yuwen Chen, Haoyu Dong, and Maciej A. Mazurowski. Anatomically-controllable medical image generation with segmentation-guided diffusion models. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2024.

[7] Adrian B Levine, Jason Peng, David Farnell, Mitchell Nursey, Yiping Wang, Julia R Naso, Hezhen Ren, Hossein Farahani, Colin Chen, Derek Chiu, Aline Talhouk, Brandon Sheffield, Maziar Riazy, Philip P Ip, Carlos Parra-Herran, Anne Mills, Naveena Singh, Basile Tessier-Cloutier, Taylor Salisbury, Jonathan Lee, Tim Salcudean, Steven JM Jones, David G Huntsman, C Blake Gilks, Stephen Yip, and Ali Bashashati. Synthesis of diagnostic quality cancer pathology images by generative adversarial networks. *The Journal of Pathology*, 252(2):178–188, 2020.

[8] Aman Rana, Gregory Yauney, Alarice Lowe, and Pratik Shah. Computational histological staining and destaining of prostate core biopsy rgb images with generative adversarial neural networks. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 828–834, 2018.

[9] Ruichen Rong, Hudanyun Sheng, Kevin W Jin, Fangjiang Wu, Danni Luo, Zhuoyu Wen, Chen Tang, Donghan M Yang, Liwei Jia, Mohamed Amgad, et al. A deep learning approach for histology-based nucleus segmentation and tumor microenvironment characterization. *Modern Pathology*, page 100196, 2023.

[10] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.

[11] Mingxing Tan and Quoc V. Le. Efficientnetv2: Smaller models and faster training, 2021.

[12] Phil Wang. denoising-diffusion-pytorch. `https://github.com/lucidrains/denoising-diffusion-pytorch`, 2023.