DIGITAL STAINING IN CERVICAL CYTOLOGY

Generative adversarial networks for digital staining

Andreia Mariana Ferreira Carreira

Supervisor:

Professor António Raimundo, PhD, Iscte – Instituto Universitário de Lisboa

Date: 28/02/2026

Co-funded by the European Union

DIGITAL STAINING IN CERVICAL CYTOLOGY

Generative adversarial networks for digital staining

Andreia Mariana Ferreira Carreira

Supervisor:

Professor António Raimundo, PhD, ISCTE- Instituto Universitário de Lisboa

Date: 28/02/2026

Co-funded by the European Union

Dedication / Acknowledgements

I would like to express my gratitude to my supervisor, Professor António Raimundo, for welcoming me with enthusiasm from our very first meeting. I am especially grateful for his patience in addressing my many questions and, above all, for continually encouraging me to step beyond my comfort zone throughout this project.

I would also like to thank Mónica for her kindness, unwavering support, and ongoing motivation during this journey. Her care and encouragement made a genuine difference, particularly during the most challenging moments.

My thanks also go to João Cambeiro for his support and practical guidance in identifying the best solutions to obtain the data required for this research.

Finally, I would like to acknowledge the Joaquim Chaves Saúde group, with special thanks to Dr. Herminia Pereira for her invaluable support and collaboration.

Abstract

Cervical cancer is the fourth most common malignant neoplasm in women worldwide. The diagnosis relies on Papanicolaou test, which includes the examination in a light microscope of a stained slide of exfoliated cervical cells. The aim of this dissertation is to transform unstained cytology into digitally stained cervical cytology. A dataset was obtained by acquired whole-slide images from the same sample before and after Papanicolaou staining. Whole-slide images of normal cervical cells were digitalised and subdivided into image patches for model training and evaluation. Four model configurations were evaluated: CycleGAN, pix2pix, CycleGAN-Turbo and pix2pix-Turbo. Among the non-Turbo approaches, CycleGAN produced a more coherent morphology and staining patterns than pix2pix. Overall, pix2pix-Turbo produced the most realistic results, better preserving cellular structure and staining appearance under imperfect image registration, due to its supervised approach and pre-trained diffusion backbone. A quantitative assessment of pix2pix-Turbo yielded a SSIM value of 0.76 and a PSNR value of 18.18. However, these values should be interpreted with caution given the structural mismatches between the output and ground truth patches. A Visual Turing Test questionnaire involving 12 participants and 60 images (30 real and 30 synthetic) indicated that participants were able to distinguish between real and synthetic patches.

In summary, digital staining in cervical cytology seems feasible, but further validation is required before its routine diagnostic use. Future work should include larger and more diverse datasets, including pathological cases, with improved acquisition and registration procedures methods.

Keywords: cytology; digital staining; PAP smear; generative adversarial networks; cycleGAN; pix2pix

JEL classification codes: C45; l10

Index

Dedication / Acknowledgements ii
Abstract iv
Index vi
Index of Figures ix
Index of Tables x
Index of Graphs xi
Glossary xiii
Introduction 1
Chapter 1 2
Organisation of the Dissertation 2
1.1 Research Questions 2
1.2 Objectives 2
1.3 Methodology and Organisation of the Dissertation 2
Chapter 2 Literature Review 4
2.1 Literature review research 4
2.2 Cervical Cytology 4
2.3 Papanicolaou Stain – Traditional Process 6
2.4 Digital Cytology 7
2.5 Computer Vision 8
2.6 Image Enhancement 9
2.6.1 Contrast Limited Adaptive Histogram Equalization 9
2.7 Machine Learning Algorithms 10
2.7.1 Digital Staining 12
2.7.2 Supervised Learning 12
2.7.3 Unsupervised Learning 12
2.7.4 Hallucinations 13
2.8 Digital Staining Evaluation 14

2.8.1 Quantitative Metrics 14
2.8.2 Qualitative Metrics 14
2.9 Business Understanding 15
Chapter 3 Methodology 16
3.1 Data Understanding 16
3.1.1 Samples 16
3.1.2 Datasets 16
3.2 Data Preparation 18
3.2.1 Implementation of CLAHE 18
3.3 Data Modelling 19
3.3.1 pix2pix 19
3.3.2 CycleGAN 20
3.3.3 CycleGAN-Turbo and pix2pix-Turbo 21
3.4 Data Evaluation 22
3.4.1 Structural Similarity Index 23
3.4.2 Peak Signal-to-Noise Ratio 23
3.4.3 Visual Turing Test 23
Chapter 4 Results 24
4.1 First Experiment – CycleGAN (without pre-processing) 24
4.2 Second Experiment – CycleGAN (with CLAHE pre-processing) 26
4.3 Third Experiment – pix2pix (with CLAHE pre-processing) 27
4.4 Fourth Experiment – CycleGAN-Turbo (with CLAHE pre-processing) 28
4.5 Fifth Experiment - pix2Pix-Turbo (with CLAHE pre-processing) 30
4.6 Evaluation 31
4.6.1 Structural Similarity Index – pix2pix-Turbo 31
4.6.2 Peak Signal-to-Noise Ratio – pix2pix-Turbo 31

4.6.3 Visual Turing Test 31

Chapter 5 Discussion 35

Conclusion 41

Limitations and Restrictions 42

Future Work 43

Bibliographic References and Sources 45

Appendix A 52

Procedure for Pap stain 52

Index of Figures

Figure 1.1. Diagram illustrating the phases of the CRISP-DM. Reprinted from CRISP-DM Process Diagram by Kenneth Jensen (2012) 3
Figure 2.1. Different types of cervical cells. Adapted from Jiang et al. (2023). 5
Figure 2.2. Original retinal fundus image. Adapted from Bhat & Tarun Patil M S, (2014).....10
Figure 2.3. Enhanced retinal fundus image. Using modified CLAHE. Adapted from Bhat & Tarun Patil M S, (2014).....10
Figure 3.1. Illustration of overview datasets acquisition. Author's own work (2026) 18
Figure 3.2. Original image. Author's own work (2026) 18
Figure 3.3.Enhanced image using CLAHE. Author's own work (2026) 18
Figure 3.4. Overview of pix2pix architecture with cytology images. Author's own work (2026) 20
Figure 3.5. Overview of CycleGAN goal. Adapted from Zhu et al., (2020). 20
Figure 3.6. Workflow of CycleGAN with cytological images. The figure illustrates the forward and backward cycle consistency and constrains. Adapted from Salido et al., (2023) with cytological images. 21
Figure 3.7. Workflow of CycleGAN-Turbo and pix2pix-Turbo with cytological images. Adapted from Parmar et al., (2024) 22

Index of Tables

Table 4.1. Summary of the algorithm configuration for the first experiment. 24
Table 4.2. Results of first experiment using CycleGAN without pre-processing.....25
Table 4.3. Summary of the algorithm configuration for the second experiment. 26
Table 4.4. Results of digital staining using CycleGAN with pre-processed images. 26
Table 4.5. Summary of the algorithm configuration for the third experiment.....27
Table 4.6. Results of digital staining using pix2pix with pre-processed images.....28
Table 4.7. Summary of the algorithm configuration for the fourth experiment. 29
Table 4.8. Results of digital staining using CycleGAN-Turbo with pre-processed images. 29
Table 4.9. Summary of the algorithm configuration for the fifth experiment.....30
Table 4.10. Results of digital staining using pix2pix-Turbo with pre-processed images.....30
Table 4.11. Confusion Matrix.....32
Table 5.1. Patches with missing cells. 35
Table 5.2. A detailed analysis of the CycleGAN output.....36
Table 5.3. Example of the pix2pix output 37
Table 5.4. Comparison of the same patches.....38
Table 5.5. Example of output obtained using pix2pix-Turbo with the respective ground truth. 39
Table 5.6. Example of output obtained using CycleGAN-Turbo.....39

Index of Graphs

Graph 4.1. Accuracy, sensitivity, and specificity between pathologist and cytotechnicians ...32

Graph 4.2. Accuracy-confident and error-not confident between pathologist and cytotechnologist. 33

Graph 4.3. D' between pathologist and cytotechnologist. 34

Glossary

3D	Three Dimensional
ANNs	Artificial Neural Networks
BS	Block Size
CADx	Computer-Aided Diagnosis
cGAN	Conditional Generative Adversarial Network
CL	Clip Limit
CLAHE	Contrast Limited Adaptive Histogram Equalization
CMV	Cytomegalovirus
CNNs	Convolutional Neural Networks
CRISP-DM	Cross-Industry Standard Process for Data Mining
CUT	Contrastive Unpaired Translation
CV	Computer Vision
CycleGAN	Cycle-consistent Generative Adversarial Network
DL	Deep Learning
DP	Digital Pathology
EA	Eosin Azure
GANs	Generative Adversarial Networks
H&E	Haematoxylin and Eosin
HPV	Human Papillomavirus
LBC	Liquid- Based Cytology
LoRA	Low-Rank Adaptation
ML	Machine Learning
MRI	Magnetic Resonance Imaging
MSE	Mean Squared Error
NILM	Negative for Intraepithelial Lesion or Malignancy
Pap	Papanicolaou
PFFN	Parallel Feature Fusion Network
PSNR	Peak Signal-to-Noise Ratio
ROI	Regions Of Interest
SSIM	Structural Similarity Index
VTT	Visual Turing Test
WHO	World Health Organization
WSI	Whole Slide Imaging
WSR	Wilcoxon signed- rank

Introduction

According to the World Health Organization (WHO, 2025) cervical cancer is the fourth most common malignant neoplasm in women worldwide. The diagnosis relies on the Papanicolaou (Pap) test. This test consists of a collection of exfoliated cervical cells, which are subsequently fixed, stained and then examined under a light microscope.

The staining technique enables the visualisation of the changes in the nuclear structures and cytoplasmic features, which allows the diagnosis. However, cytology-based screening is constrained by suboptimal sensitivity and a substantial proportion of unsatisfactory or inadequate specimens. Furthermore, this visual screening process is time-consuming, subject to variable accuracy, and influenced by human subjectivity, which can lead to inconsistent results and, in some cases, serious diagnostic errors.

The advent of digital cytology with Computer-Aided Diagnosis (CADx) systems has helped automated screening, as well as in professionals training and analysing cellular samples, particularly in detecting cervical cancer (Wilbur, 2011). By applying sophisticated image analysis and machine learning methods, these systems assist pathologists in interpreting cytological images, further contributing to the precision and efficiency of diagnostic processes (Saha et al., 2016; Kalita et al., 2024)

Machine Learning (ML) systems in cytology are based on chemical staining, which is expensive, produces hospital waste, is time-consuming and varies from laboratory to laboratory.

Bai et al. (2023) conducted a review of various deep learning-based virtual staining methods in histology, including Generative Adversarial Networks (GANs) for label-free virtual staining and stain-to-stain conversion. This demonstrates the increasing importance of generative models in digital pathology. In the field of cytopathology, Teramoto et al. (2021) proposed a CycleGAN-based method for converting stains between Giemsa- and Papanicolaou-stained cytological images. Meanwhile, Wang et al. (2024) introduced CytoGAN for transferring stains between H&E- and Papanicolaou-stained endometrial cytopathology images, paying particular attention to preserving cellular structure. However, these studies used chemically stained samples for stain-to-stain transformation and did not focus on cervical-vaginal cytology. Therefore, to the best of our knowledge, GANs-based digital staining remains under explored in cervical-vaginal cytology, particularly regarding approximating Papanicolaou stain images and reducing dependence on conventional chemical staining. This constitutes the focus of this dissertation.

Chapter 1

Organisation of the Dissertation

This chapter contains the research questions, objectives and methodology and organisation of this dissertation.

1.1 Research Questions

The aim of the current study is to transform unstained cytology into digitally stained cytology. To address this aim, the following research questions are posed:

- How can we create the necessary datasets for training the algorithms, and what precautions should be taken in this process?
- Which digital staining model produces results most similar to chemical staining?
- Which deep learning computer vision model is most effective in digital staining?
- Can digitally stained images be considered equivalent to chemically stained images, and can they be used for diagnostic purposes in the future?

1.2 Objectives

The main objectives of this dissertation are:

- Build a cytopathology dataset with unstained cells and stained cells.
- Transform unstained cytology into digitally stained cytology.
- Evaluate the reliability of digitally stained cytology.
- Assess the diagnostic viability of digitally stained cytology.

1.3 Methodology and Organisation of the Dissertation

This dissertation is guided by the CRISP-DM (Cross-Industry Standard Process for Data Mining) by Chapman, (2000) which provides a comprehensive and structured framework for the systematic examination and interpretation of data. In line with the methodological principles of CRISP-DM, the study employs an experimental research design, involving a series of processes and iterative refinement of parameters in accordance with the results obtained, as illustrated in **Erro! A origem da referência não foi encontrada.**.



Figure 1.1. Diagram illustrating the phases of the CRISP-DM. Reprinted from CRISP-DM Process Diagram by Kenneth Jensen (2012)

The Introduction and Organisation of the Dissertation (Chapter 1) provide the foundation for a thorough understanding of the problem context.

The Literature Review (Chapter 2) establish a conceptual and contextual foundation for the dissertation. They outline the study's main objectives, discuss the key theoretical concepts relevant to the research topic and identify the potential contributions to the academic field.

Subsequently, the Methodology (Chapter 3) presents a comprehensive account of the resources employed throughout the dissertation. Section 3.1 covers Data Understanding and provides a detailed description of the cytology samples and datasets creation. Data Preparation is the focus of Section 3.2. Section 3.3 specifics Data Modelling and contains a detailed description of the algorithms used.

The experiments and the results of those experiments are described in the Results (Chapter 4).

The Discussion (Chapter 5) involves analysing and discussion of the results aims of assessing the extent to which the models fulfil the objectives of the dissertation.

Finally, the Conclusion, Limitations and Future Work sections examine the strengths and weaknesses of the dissertation, offering a comprehensive reflection that aims to provide insights into future studies.

Chapter 2

Literature Review

This chapter contains a review of the fundamental concepts of digital staining in cervical cytology. Firstly, there will be an introduction to cervical cytology, including staining, and digital techniques currently used in this field. Secondly, we will present the key principles of computer vision, image enhancement, and different types of machine learning algorithms. Within this, we will focus on the most used algorithms for digital staining. We will also provide an overview of the different types of metrics used to evaluate digital staining. Lastly, we will present the business understanding.

2.1 Literature review research

A narrative literature review was conducted by searching in the PubMed, Scopus, and IEEE Xplore databases. The search terms used were "virtual staining", "digital staining", "stain transfer", "cervical cytology", "Pap smear", “Papanicolaou”, “GAN”, “CycleGAN”, “pix2pix”, "deep learning".

Studies were considered eligible when they addressed deep learning-based virtual staining, GAN-based stain transformation, digital staining in histology or cytology, digital cytology, or computational approaches relevant to cervical cytology image analysis. Priority was given to original research articles and recent review articles. Studies were excluded when they were unrelated to biomedical images, did not involve digital or virtual staining, or were not available in English. The selected studies were used to contextualise the field, compare existing approaches and define the specific gap addressed by this dissertation.

2.2 Cervical Cytology

Cervical cancer is strongly associated with persistent infection by high-risk types of the sexually transmitted human papillomavirus (HPV). The HPV vaccines were introduced in the past decade and can protect against approximately 70-90% of cervical cancers (Cheng et al., 2020).

In cervical cytology, the normal cellular elements are squamous cells, glandular cells, and certain normal microorganisms of the genital tract (Nayar & Wilbur, 2015). Figure 2.1 illustrates the different types of normal cervical cells, squamous (a) superficial, b) intermediate, c) parabasal, d) basal) as well as glandular (e), f) endocervical and g) endometrial). The squamous cells are superficial cells originating from the outermost layer of cervical epithelium with a markedly condensed nucleus and an abundant eosinophilic cytoplasm. Keratohyalin granules may also be present within the cytoplasm, indicating the production of high-molecular-weight keratin (Nayar & Wilbur, 2015). The intermediate cells are typically located within the middle or intermediate layer of the squamous epithelium. Compared with

superficial cells, the nucleus is larger, with a cross-sectional area of around 35 μm^2 and a finely granular chromatin texture.

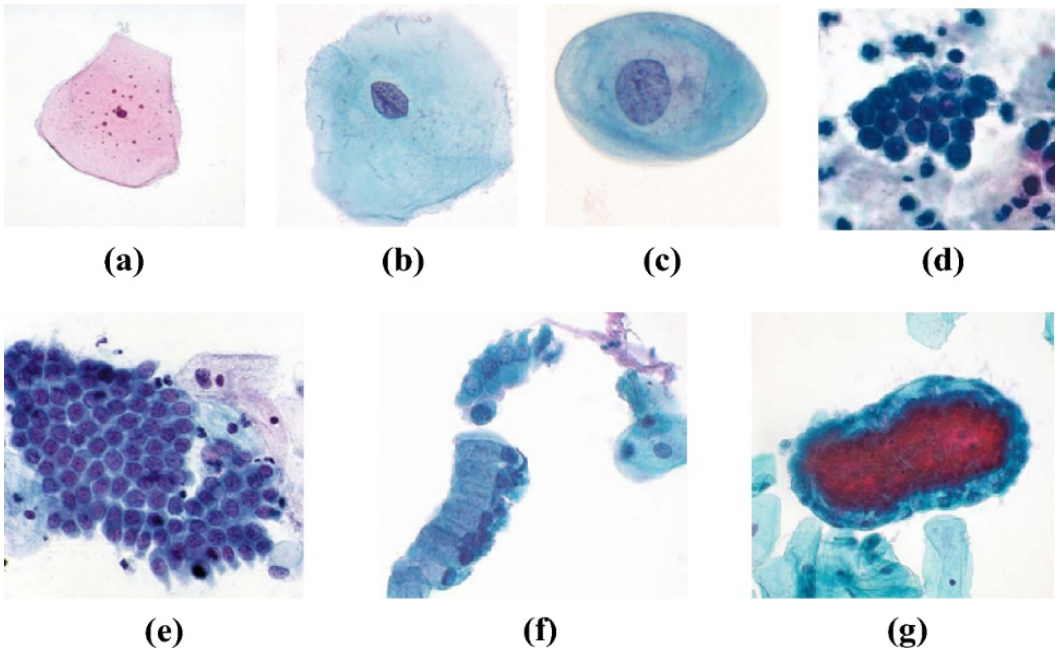


Figure 2.1. Different types of cervical cells. Adapted from Jiang et al. (2023).

The nucleus is frequently elongated and may display a longitudinal groove. In cervical cytology, the intermediate cell nucleus is considered the standard reference for nuclear size in other cell types. The parabasal cells together with immature squamous metaplastic cells represent the least mature elements observed in a cervical cytology specimen. In hormonally mature epithelium, they are usually absent, as they originate from the deeper epithelial layers that are rarely collected in cytology samples from premenopausal women. When hormonal stimulation is lacking, these cells form part of a relatively thin, atrophic epithelium. Parabasal cells are more common in postmenopausal and postpartum conditions. Their nuclei are larger than those of intermediate cells, with a cross-sectional area of approximately 50 μm^2. The cytoplasmic volume is reduced, resulting in a higher nuclear-to-cytoplasmic ratio compared with intermediate or superficial cells and the cytoplasm itself is denser and more granular in texture (Nayar & Wilbur, 2015).

Endocervical glandular cells display considerable variation in nuclear size, with an average diameter of around 50 μm^2, slightly exceeding that of an intermediate squamous cell. The nuclei possess finely granular, evenly distributed chromatin and contain small nucleoli. The cytoplasm appears diffusely vacuolated or granular. These cells demonstrate polarity, with the nuclei positioned at one pole and mucus located at the opposite pole of the cytoplasm. Endometrial glandular cells are shed spontaneously and may originate from either epithelial or stromal tissue and may appear as single cells or in clusters. Compared with endocervical cells, endometrial cells are generally smaller,

possessing nuclei with an area equal to or slightly less than that of intermediate squamous parabasal cells (around 35 μm^2) and typically demonstrate a higher nuclear-to-cytoplasmatic ratio. Their chromatin is usually dense and uneven, sometimes containing degenerative debris. Nucleoli are not usually prominent but may be more easily recognised in liquid-based samples owing to improved preservation. The cytoplasm is usually sparse and may appear either dense or vacuolated (Solomon & Nayar, 2015). A variety of infectious agents and cellular alterations can be identified through cervical cytology: *Trichomonas vaginalis*, *Candida species*, *Gardnerella vaginalis*, *Atopobium vaginae*, *Megasphaera spp., Prevotella spp*., *Sneathia spp*, *Herpes simplex virus* (HSV), *Cytomegalovirus* (CMV), and *Actinomyces.*

Cervical cancer primarily originates in the transformation zone of the cervix, typically in squamous cells (AleksioskaPapestiev et al., 2018). The Pap test, invented by George N. Papanicolaou and H.F. Traut in the 1940s, provided a robust and low-complexity method of screening for cervical cancer (Chantziantoniou et al., 2017). According to Marth et al., (2017) the Pap test has served for decades as the primary approach for cervical cancer screening, reducing incidence by 60-90% and mortality by 90%. However, cytology-based screening is a time-consuming process, constrained by suboptimal sensitivity estimated at approximately 50% and a substantial proportion of unsatisfactory or inadequate specimens, which can lead to inconsistent results and, in some cases, serious diagnostic errors.

2.3 Papanicolaou Stain – Traditional Process

The Pap test encompasses the collection of exfoliated cervical cells, which are subsequently fixed, microscopically examined and interpreted based on their morphological attributes. The staining technique employed affords a polychromatic visualisation of nuclear structures and cytoplasmic features. The Pap stain facilitates the evaluation of alterations in nuclear chromatin indicative of necrotic processes, the assessment of cellular degeneration and the determination of the maturation status of squamous epithelial cells (Gill, 2013).

The Pap stain is composed of five distinct dyes, which are distributed across three separate solutions: Haematoxylin, Orange G and Eosin Azure (EA). Haematoxylin and Orange G are each prepared individually, while the third solution, known as EA, is a mixture containing Bismarck brown Y, Light Green SF yellowish and Eosin Y. This stain was designed to achieve three principal aims: to provide clear nuclear definition, to preserve the natural transparency of cells and to enable reliable cytoplasmic differentiation. Haematoxylin served as nuclear dye. The issue of transparency was addressed using alcohol-based dye solutions, which penetrate cells more slowly than aqueous preparations, thus maintaining their inherent clarity and limiting the disruptive effect of salts on

staining quality. Cytoplasmic contrast was obtained with counterstains, Orange G and EA. Despite their utility, the EA formulations presented chemical incompatibilities and lacked precise standardization, often producing variable results. Even decades after its introduction, the Pap stain remains fundamental in cytology, though inconsistencies in reagents, protocols, and outcomes (Gill, 2013)

2.4 Digital Cytology

Digital cytology and histology, key parts of Digital Pathology (DP), evolved from telepathology in the late 20th century. It allowed for remote diagnosis, automated screening, education and training, and quality control (Wilbur, 2011; Kim et al., 2024). Early systems faced issues like slow internet and synchronization errors. Virtual microscopy later emerged to solve these problems by enabling scanned slides to be stored and accessed remotely.

Cytology possesses distinct characteristics that differentiate it from histology and necessitate tailored technological approaches within the context of digital pathology. One of the main distinctions lies in the variety of stains employed, including modified Giemsa/Diff- Quick, Pap stain and Haematoxylin and Eosin (H&E), whereas histology primarily relies on H&E staining. Another distinction is that the cytological material may contain background elements that obscure diagnostic cells, such as blood, inflammatory exudate, cellular debris and exogenous substances (e.g., lubricant), complicating interpretation. Finally, the diverse cellular configurations observed in cytology slides demand the application of sophisticated image processing methods (Giansanti, 2025)

Cytology preparations can be categorized as either conventional smears or Liquid-Based Cytology (LBC). Each type has its own unique features, depending on how the cells are distributed. Conventional smears typically show three-dimensional (3D) cells clusters with uneven distribution and variable thickness, while LBC preparations offer a more uniform distribution of cells, consistent thickness, and simpler topography. These features pose additional challenges for Whole Slide Imaging (WSI) in cytology. Image acquisition and analysis require cytology- specific solutions, including adapted cell detection algorithms, higher-resolution scanning, Z-stacking or volumetric (extended focus) applications, as well as intuitive tools for efficient slide navigation and annotation. WSI scanners may need to perform Z-stacking. This technique captures multiple focal planes to adequately visualize the uneven thickness and 3D cellular structures of cytology samples. However, this process significantly increases scanning time, file size, and bandwidth required for data transmission (Marletta et al., 2021).

Despite the technical and practical challenges, digital cytology presents considerable benefits. Once cytological slides are scanned, the reliance on physical glass slides is eliminated, allowing for greater mobility of data, easier sharing between professionals and simplified digital archiving. In fact, digital cytology has the potential to enhance workflow efficiency by reducing manual tasks, enabling

quicker case distribution, supporting fully electronic reporting, and facilitating streamlined long-term storage and retrieval of cases (House et al., 2013). The digital format also opens new opportunities for remote learning and enables the application of artificial intelligence (AI) tools for image analysis. Various models that use AI in cytology have already been published, including neural networks, support vector machines, and k-nearest neighbours, as well as ensembles that combine these approaches (Victória Matias et al., 2021).

According to the review article by Emery et al. (2025) laboratories currently rely on AI mainly through digital cytology platforms designed to support screening workflows, particularly in cervical and urinary cytology. Tools such as BD FocalPoint, TechCyte, Hologic Genius and AIxURO use whole-slide imaging together with deep learning algorithms, mostly Convolutional Neural Network (CNN), to scan slides automatically, pick out areas that appear atypical and present them to the cytotechnologist or pathologist. In practice, The AI is responsible for tasks such as identifying suspicious cells, locating relevant regions, and analysing nuclear features, which helps speed up the screening process and reduces workload. All these solutions work on slides that have already been stained traditionally.

2.5 Computer Vision

Computer Vision (CV) is a domain of computer science concerned with enabling computers to process and interpret visual information in a manner analogous to human perception. Using CV algorithms, computers can perform a wide variety of tasks, such as object detection, facial recognition, pattern identification, image classification, and partitioning of images into meaningful components. When applied to digital images or videos sequences, these algorithms enable systems to construct a representation of the visual content, which can then serve as the basis for automated decision-making or action.

In general terms, CV techniques may be distinguished into two principal categories. The first encompasses low-level algorithms, which operate at the pixel level to derive elementary features such as edges, contours, and basic geometric shapes. The second comprises high-level algorithms, which are concerned with the recognition of objects and the representation of more complex structural relationships within an image (Mugesh, 2023).

Victória Matias et al. (2021) present a systematic review of the state of the art in applying computer vision techniques to cytology. They emphasise the importance of these methodologies in facilitating medical diagnosis. The authors identify three main areas of cytological image analysis- classification, object detection, and segmentation – each of which has different approaches based on manual feature extraction. The review reveals an increasing prevalence of deep learning techniques, including architectures such as VGG, ResNet, Inception, DenseNet, U-Net, Mask R-CNN, Faster R-CNN and

YOLOv3. These techniques enable a range of applications, from pixel-by-pixel segmentation to large-scale cell classification and offer high performance. Additionally, it mentions hybrid methods that combine classic preprocessing techniques with deep learning models, leveraging the robustness of neural networks and the simplicity of traditional filters and algorithms. Giarnieri & Scardapane, (2023) note that more recent models, such as EfficientNet, MobileNet and InceptionNetv3, offer advantages in computational efficiency while maintaining high levels of accuracy, even when dealing with the morphological variability typical of cytological samples.

However, significant challenges remain, including the scarcity of sufficiently large and annotated databases, the diversity of laboratory techniques that introduce variability in the data and the reduced transparency of deep learning algorithms. Thus, the application of computer vision in cytopathology represents a paradigm shift, with the potential to make diagnostic processes more efficient and consistent. However, consolidation still depends on advances in standardization, model interpretability, and validation in real clinical settings.

2.6 Image Enhancement

Typically, in computer vision, images or videos are processed before being fed into a classifier. This image enhancement step improves image quality and enables feature extraction. This step can highlight the characteristics that distinguish between different object classes and can also reduce computation time (Fan et al., 2020)

According to the review article by Seoni et al. (2024), the most used image enhancement techniques in medical images are grayscale normalisation, resampling, colour normalisation, denoising, and contrast enhancement. These techniques form the basis of AI methods. By harmonising input image data, they facilitate accurate and reliable analysis, interpretation, and diagnosis. The methods enable AI models to focus effectively on the crucial aspects of an image.

2.6.1 Contrast Limited Adaptive Histogram Equalization

Contrast Limited Adaptive Histogram Equalization (CLAHE) is a technique belonging to the contrast enhancement category. It is a recognised technique used to improve the contrast of Magnetic Resonance Imaging (MRI) scans (Halloum & Ez-Zahraouy, 2025). The purpose of this technique is to adjust the contrast levels of an image, resulting in improved visibility and differentiation of objects or structures (Seoni et al., 2024). For example, Bhat & Tarun Patil M S (2014) demonstrate greatest improvement in enhancing medical images using CLAHE, as measured by Structural Similarity Index (SSIM).

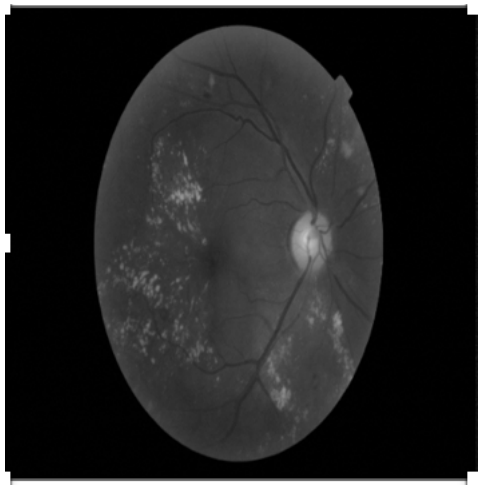


Figure 2.2. Original retinal fundus image. Adapted from Bhat & Tarun Patil M S, (2014)

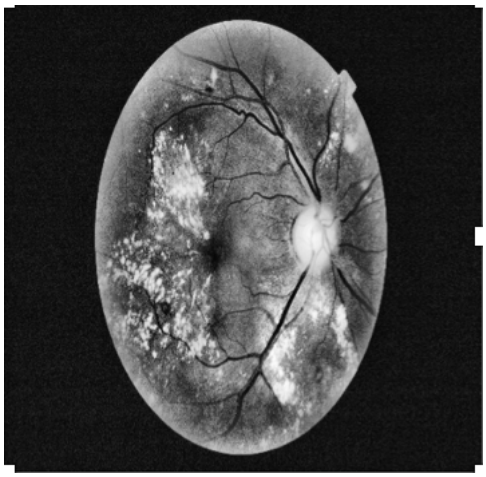


Figure 2.3. Enhanced retinal fundus image. Using modified CLAHE. Adapted from Bhat & Tarun Patil M S, (2014)

In the CLAHE technique, there are two parameters that can be controlled to improve the image, namely block size (BS) and clip limit (CL). Images become brighter if CL increases and darker if CL decreases (Halloum & Ez-Zahraouy, 2025).

In their article, Verma & Barthwal, (2025) applied four pre-processing methods — wavelet denoising, CLAHE, background correction and Laplacian sharpening — to cytological images. CLAHE was used to enhance image quality by adjusting pixel intensities within small regions while minimising noise amplification. However, according to Setiawan et al. (2013) , CLAHE can amplify noise in low-contrast areas that are not necessarily related to image quality. This can result in the introduction of optical artefacts, including small coloured spots.

2.7 Machine Learning Algorithms

According to Mohammed et al (2016) ML forms part of AI and seeks to endow machines with the capacity to handle tasks competently by relying on advanced software solutions. At the heart of this process are statistical learning methods, which act as the essential building blocks of machine intelligence.

The underlying idea is relatively simple: instead of replicating the cognitive mechanisms of human learning, computer programs are designed to enable machines to learn from data and apply this acquired knowledge to tasks such as prediction. The result of this process is a model that establishes a mapping between input variables and expected outputs. While some models are sufficiently transparent to be interpreted, others operate as “black boxes”, with internal decision processes that

are not readily understood. Such models can be regarded as approximations of the real-world phenomena they are intended to simulate (Mohammed et al., 2016).

Mohammed et al (2016) describes the four types of algorithms: supervised, unsupervised, semi-supervised and reinforcement learning.

Deep Learning (DL) is a part of ML that uses artificial neural networks (ANNs), designed to mimic how the biological neural network of the human brain works. An ANN begins with an input layer that receives raw data, which is then passed through one or more hidden layers. Each hidden layer processes and changes the data using nonlinear transformations until it reaches the output layer, where the prediction is produced. The term 'deep' refers to the use of multiple hidden layers in these models (Goodfellow et al., 2020). According to Dey (2022), a range of deep learning models, particularly convolutional neural networks (CNNs), have been applied to the detection and classification of cells in cytology. CNNs are a specific type of ANN designed primarily for processing images, videos, and spatial data (LeCun et al., 2015). Another type of deep learning model used in cytology is the Generative Adversarial Networks (GANs). GANs have been applied to image enhancement tasks such as stain normalisation, virtual staining, and the removal of ink/markers. Moreover, GANs can be trained to produce images of rare diseases, which helps improve generalisation, reduce overfitting, and strengthen model robustness (Homeyer et al., 2022). GANs are formed by two distinct neural networks that compete in an adversarial structure: one generates synthetic samples and the other evaluates them, distinguishing them from real samples. These networks are usually referred to as generators and discriminators, depending on the role they play in the method.

During training, GANs rely on both real data and synthetic samples produced by the generator. The discriminator's role is to distinguish between these two sets, whereas the generator seeks to deceive the discriminator by generating artificial examples that closely resemble the training data distribution. Consequently, the generator's main objective is to approximate this distribution so that its outputs cannot be recognized as synthetic. Training begins with noise sampled from a known statistical distribution (such as Gaussian or uniform), which is progressively transformed into synthetic instances. Throughout the process, the generator adjusts its parameters in response to errors detected by the discriminator, thereby improving the realism of its outputs. In parallel, the discriminator refines its ability to differentiate real from synthetic data by estimating the probability of authenticity for each sample. When it fails to classify a synthetic instance correctly, a negative loss is applied, reinforcing the adversarial learning dynamic (Yilmaz & Korn, 2024).

2.7.1 Digital Staining

Digital staining is an artificial intelligence-driven method that substitutes the traditional staining process (Salido et al., 2023). This transformation is learnt using GANs models that have been trained on datasets including chemically stained, unstained and other stained cytological slides, depending on the goal. In other words, the process of converting unstained cells into stained cells is known as label-free virtual staining.

Conversely, the process of converting stained cells into different stains is referred to as stain-to-stain transformation (Bai et al., 2023). These models may be broadly classified into two categories according to the availability of paired images: supervised and unsupervised learning (Lin et al., 2025). Supervised learning relies on pixel-wise spatial correspondence between source and target images, ensuring direct alignment during training, whereas unsupervised learning infers the transformation from the intrinsic content of the images, thereby obviating the requirements for explicit pixel-level alignment (Latonen et al., 2024).

2.7.2 Supervised Learning

In Latonen et al. (2024), pix2pix is identified as the most used supervised learning method in digital stain. As a supervised learning method, pix2pix requires paired images during the training phase, consisting of an original image and its corresponding transformed version, although such pairs are often challenging to obtain in practical applications (Pegah Salehi and Abdolah Chalechale, 2020). Pix2pix generates output images in accordance with the provided input, a process implemented by introducing the same input image as a conditioning factor to both the generator and the discriminator (K. S. Park et al., 2023). This model has an excellent performance but obtaining paired images is extremely complicated and time consuming.

Some studies have shown the effectiveness of the pix2pix model in histopathological image processing. Levy et al. (2020) applied it for synthetic stain generation and cell segmentation, achieving high accuracy and strong correlation with true stains. Likewise, Pegah Salehi and Abdolah Chalechale, (2020) employed a pix2pix based approach to stain normalisation that outperformed traditional methods while preserving tissue morphology.

2.7.3 Unsupervised Learning

The cycle-consistent generative adversarial network (CycleGAN) model performs image-to-image translation through style transfer without assuming explicit spatial or pixelwise alignment. It uses two coupled generator- discriminator pairs (dual networks) to learn the mapping that produces realistic synthetic staining images. It is considered an unsupervised model because it does not require paired

images (Zhu et al., 2020). In Latonen et al., (2024) CycleGAN is the most widely used unsupervised learning method in digital staining. Numerous studies indicate that CycleGAN based virtual staining methods can generate images that are visually comparable to genuine stained samples. For example, Gadermayr et al. (2019) demonstrated that CycleGAN based image translation can enable stain independent segmentation in kidney histology by transferring structural information across different staining domains. Tissue morphology was preserved and reliance on annotated data reduced, while the results were comparable to those of supervised models.

The study of Salido et al. (2023) compared pix2pix, CycleGAN and contrastive unpaired translation (CUT) models for virtual H&E staining of label-free breast tissue using multispectral microscopy images. Among them, CycleGAN produced images visually consistent with real H&E stains.

Moreover, some studies have introduced modifications or extensions to the original CycleGAN architecture, such as Meng et al. (2021) who proposed a deep learning based virtual staining method for ovarian cancer tissue using GANs. Their improved CycleGAN and Parallel Feature Fusion Network (PFFN) generated realistic H&E-like images from unstained autofluorescence samples while preserving cellular detail. An accuracy of 97% was obtained by pathologists' visual assessment, and 95% accuracy was based in deep learning.

Nevertheless, Cohen et al. (2018) cautioned that "these translated images should not be used for interpretation (e.g. by doctors) without proper tools to verify the translation process. We illustrate this problem using dramatic examples of tumours being added and removed from MRI images" (p. 535).

2.7.4 Hallucinations

As Sagheer et al. (2026) describe, although GANs have considerable potential for generating realistic medical images and improving diagnostic tools, they still face major challenges. One such challenge is hallucinations. These involve the creation of artificial elements or the omission of clinically relevant features in generated images, which can result in diagnostic errors. Hallucinations frequently arise because of mode collapse during training, leading to inaccurate or misleading structures in synthetic outputs, particularly in image-to-image translation tasks.

Huang et al. (2025) further highlight this concern when discussing hallucinations in the context of virtual tissue staining and digital pathology. They note that hallucinations can range from subtle structural or colour inconsistencies to entirely fabricated content. In virtual histopathology and virtual staining, these hallucinations can be categorised as either technical or unrealistic, such as blurred regions, folded tissue, or abnormal staining, or realistic, which are more concerning as they may appear visually plausible. For example, artificial tissue components may replace true structures, which could mislead pathologists and lead to diagnostic errors or altered tumour grading.

From an ethical perspective, this is particularly important in pathology, since hallucinated structures could compromise the reliability of diagnoses and endanger patient safety if synthetic images are used to inform clinical decisions. Therefore, when using GAN-generated images in diagnostic contexts, rigorous validation, expert oversight and transparent reporting of the limitations of these models are required.

2.8 Digital Staining Evaluation

GANs produce synthetic images similar to the real images with which they were trained. Quantitative and/or qualitative metrics are used to measure this similarity. Two of the quantitative metrics are SSIM and PSNR and one of the qualitative metrics pathologist assessments (Bai et al., 2023).

2.8.1 Quantitative Metrics

2.8.1.1 Structural Similarity Index

Structural Similarity Index (SSIM) is a metric that measures how similar one image is to another. This metric compares the luminance, contrast, and structure between two images. To compute SSIM, first the image is converted to greyscale (luminance). Then a window is slid over the image and an SSIM value is calculated at each position, producing an 'SSIM map'. The overall SSIM index is the average of all these local values (Zhou Wang et al., 2004). If the SSIM global index is close to 1, the image is more similar to the compared image and if the SSIM global index is closer to -1, the image differs more from the compared image (Jim Nilsson and Tomas Akenine-Möller, 2020).

2.8.1.2 Peak Signal-to-Noise Ratio

In image evaluation, Peak Signal-to-Noise Ratio (PSNR) is used as an indicator of quality, comparing the peak signal level to the error introduced by compression or processing. PSNR is calculated from the mean squared error (MSE) between a reference signal (e.g., the original image) and a reconstructed one (e.g., synthetic image). As image quality improves, MSE approaches zero and the PSNR increases (in theory, without bound); in contrast, lower PSNR values indicate that the images differ significantly (Hore & Ziou, 2010).

2.8.2 Qualitative Metrics

2.8.2.1 Visual Turing Test

Several studies use the Visual Turing Test (VTT) (Jang et al., 2023; Park et al., 2021; Shin et al., 2023). This method is used to assess the realism and quality of images that have been synthesised or

processed by AI. The test determines if these images are indistinguishable from real images to human observers (Hong et al., 2022). The VTT contains the same number of synthetic and real images, and the experts must determine for each whether it is real or synthetic (Levine et al., 2020; Pozzi et al., 2024; Falahkheirkhah et al., 2023).

2.9 Business Understanding

As previously mentioned, chemical cytological staining relies on specific chemical methods, such as the Papanicolaou or Diff-Quik stains. These methods require reagents and equipment, as well as strict standardization and manual intervention. These processes can introduce variability, increase operating costs, and extend the time taken to prepare slides.

With advances in artificial intelligence methods and digital microscopy technologies, the concept of digital stain has emerged, which consists of applying virtual staining to unstained cytological images. This technique has the potential to replace or complement the chemical staining process, allowing for consistent, replicable, and fully digitally adjustable staining. However, it is essential to ensure that digital staining algorithms preserve the morphological integrity of cells and maintain colour fidelity. Most importantly, they must not introduce artefacts that could compromise diagnostic. These requirements must be met before they can be used in clinical practice.

This dissertation aims to assess and evaluate a computational model that can apply realistic digital staining to cytological images and reproduce the visual characteristics of chemical stains used in laboratories. To achieve the mains objectives, we need to build a cytopathological dataset with unstained cells and stained cells, transforming unstained cytology into digitally stained cytology using GANs, assess the use of digitally stained cytology in clinical practice and evaluate the reliability of digitally stained cytology.

This digital staining process can help to reduce laboratory times and chemical waste, as well as reducing laboratory technician’s exposure to hazardous chemicals and standardising staining.

Chapter 3

Methodology

This chapter provides a detailed description about image acquisition, dataset construction, pre-processing, training models, and its evaluation.

3.1 Data Understanding

This section describes how the samples and datasets were obtained.

3.1.1 Samples

Four cervical cytology slides from four patients who had Negative for Intraepithelial Lesion or Malignancy (NILM) cytology were used.

All cervical samples were collected by gynaecologists and were prepared using ThinPrep 5000 system (Hologic, Inc.) according to the manufacturer's instructions.

To obtain unstained slides the smear was fixed with 96% ethanol for 15 minutes. The slides were then dehydrated with 96%, absolute ethanol and xylene. Coverslips were attached with a mounting medium (Surgipath, Leica). After that Whole Slides Imaging (WSI) were acquired using a VENTANA DP600 scanner (Roche Diagnostics, Basel, Switzerland) configured with three focal planes (-1, 0, +1 μm from the central plane) at 20x magnification.

To obtain the stained slides, the unstained slides after digitalization were immersed in xylene overnight to remove the coverslip and the mounting medium. Then, it was rehydrated with absolute ethanol, 96% ethanol and dipped in distilled water. The Papanicolaou staining protocol was then applied as shown in Appendix A.

Following staining, WSIs were acquired using a VENTANA DP600 scanner (Roche Diagnostics, Basel, Switzerland) configured with three focal planes (-1, 0, +1 μm from the central plane) at 20x magnification.

In total, the dataset contains 8 WSIs, paired unstained / stained WSIs from the same four patients.

3.1.2 Datasets

Anonymous cytological images were provided by Joaquim Chaves Saúde in the department of Pathological Anatomy and used exclusively for research purposes with a protocol between ISCTE and Joaquim Chaves Saúde.

The proprietary VENTANA Image Viewer software (Roche Diagnostics, Basel, Switzerland) automatically blended the Z-stack layers, by combining multiple images captured at different focal lengths into a single image with greater depth of field. This thereby eliminating manual processing and

preserving the optimal focus areas. With this process, file sizes were in the range of 5 to 7 GB per WSI, excluding areas of high cell density from the digitalisation where the −1/+1 planes were unable to achieve complete focus and areas of low cell density without adequate focal references.

Since each WSI is very large and cannot be directly input to the deep learning model, the VENTANA Image Viewer program (Roche Diagnostics, Basel, Switzerland) was used to select the Regions Of Interest (ROIs) in the WSI at 20X magnification and then cropped into uniform dimensions (1916x892 pixels). The ROIs were selected to reflect typical cervical cytology variability, including different cell types, cells in overlapping areas, areas with low and high cellularity. Subsequently 512x512-pixel square patches were generated via cropping. All patches obtained by cropping were used. In all experiments, cytology patches of 512 × 512 pixels at 20× magnification were initially extracted because this patch size provided a suitable field of view to include individual cervical cells, small cell groups, and part of the surrounding background. This was important to preserve relevant cytological information, such as nuclear morphology, cytoplasmic boundaries, cell arrangement, and staining context. Although the standard CycleGAN and pix2pix pipelines subsequently resized and cropped the images during training, the initial use of 512 × 512-pixel patches ensured consistency across datasets and compatibility with the CycleGAN-Turbo and pix2pix-Turbo models, which were trained directly at this resolution.

Dataset 1 has 1094 training patches and 278 test patches. The training patches are divided into TrainA, which contains 547 patches of unstained cells, and TrainB, which contains 547 patches of stained cells. The test patches are divided into TestA, which has 139 patches with unstained cells and TestB, which has 139 patches with stained cells. The patches from TrainA and TrainB are visual aligned but not pixel-to-pixel aligned, the same applies to patches Test A and TestB. All patches were stored in RGB format.

Figure 3.1 shows that the dataset 2 is identical to dataset 1, except that the TrainA patches and TestA patches have been converted into images enhanced with CLAHE.

The train-test split was performed at patch level using different, non-overlapping patches. Although training and test patches originated from the same four patients, no identical patch was included in both sets. Since all cases were NILM and normal cervical cells do not show marked patient-specific morphological differences, according to an experienced cytotechnologist, this strategy was considered acceptable for this exploratory study.

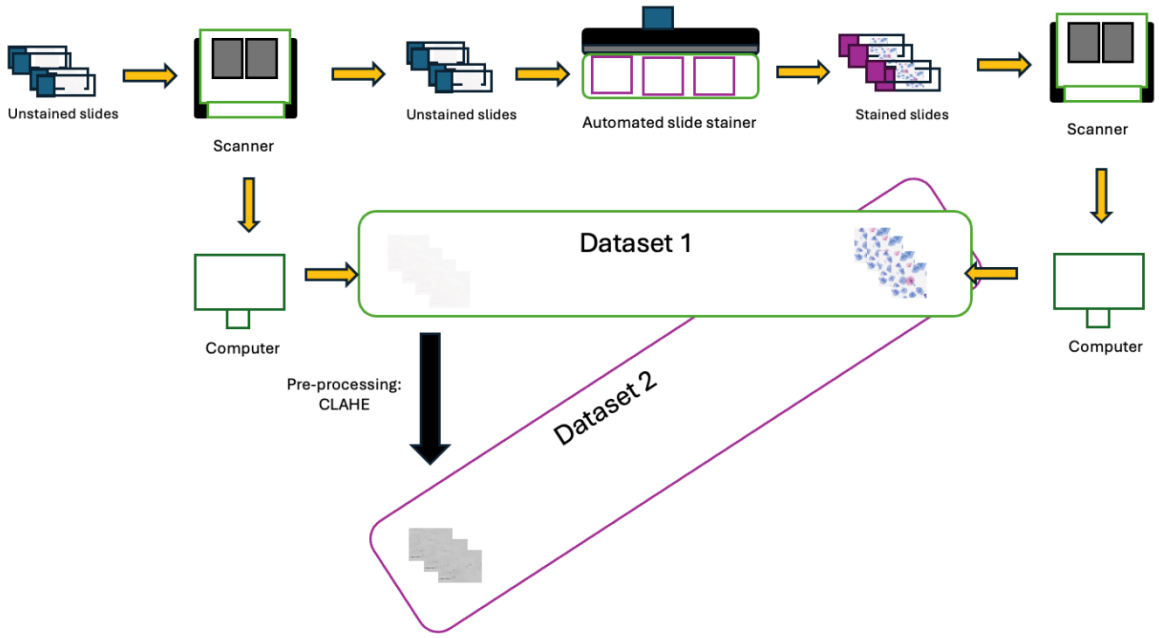


Figure 3.1. Illustration of overview datasets acquisition. Author's own work (2026)

3.2 Data Preparation

This section describes how the preprocessing was applied.

3.2.1 Implementation of CLAHE

Contrast Limited Adaptive Histogram Equalization (CLAHE) was used to provide a contrast limit. It works by dividing the image into small regions and enhancing the contrast within each region by adjusting its local intensity distribution towards a chosen histogram profile. The processed small regions were then blended smoothly using bilinear interpolation to avoid visible boundaries.

Figure 3.2. Original image. Author's own work (2026)

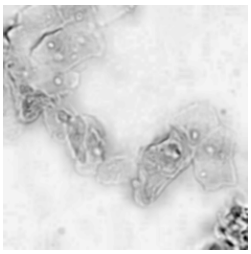


Figure 3.3.Enhanced image using CLAHE. Author's own work (2026)

Dataset 2 required the unstained cytology patches from TrainA and TestA to be converted into images contrast-enhanced images. This process increased the contrast, emphasising structural features such as nuclear contours and cell morphology. Figure 3.2 has not undergone any pre-processing, whereas Figure 3.3 has undergone CLAHE processing. Increasing the contrast of the cells and other elements the discriminator guides the generator more effectively, reducing solutions based solely on texture and improving the morphological fidelity of synthetic images.

The analysis of patches pre-processed using CLAHE was compared with that of the raw images. No additional features were detected, in contrast to what was previously described in the article of (Setiawan et al., 2013)

3.3 Data Modelling

This section provides a detailed explanation of the architecture of each algorithm.

The pix2pix and CycleGAN models were derived from the base PyTorch implementations, as proposed by Isola et al. (2017) and (Zhu et al. 2020). The pix2pix-turbo and CycleGAN-turbo were implemented using PyTorch framework, as proposed by Parmar et al., (2024).

3.3.1 pix2pix

The pix2pix model was implemented as a conditional generative adversarial network (cGAN) comprising two components: a generator and a discriminator as illustrated in Figure 3.4. The input to the generator is an unstained cytology patch and the objective is to output the corresponding stained cytology patch. The discriminator compares the similarity between the output from the generator and the ground truth, and it will give feedback to generator if the image is real or fake. The discriminator provides feedback to the generator, which in turn uses this information to improve the generated images. Therefore, this type of model requires training data in the form of pairs consisting of unstained images and their corresponding stained images. Model optimisation relies on paired training data and combines adversarial learning with a similarity-driven objective, encouraging the generated image to match the reference (ground truth) stain patch (Isola et al., 2017).

In our dataset, each unstained image is associated with a corresponding stained image, fulfilling the paired-data requirement. However, these pairs are not perfectly aligned at pixel level, which might affect the results.

The pix2pix generator is a generator with skips connections. In cases of image-to-image translation, this generator mitigates the loss of information caused by the bottleneck by adopting skip connections inspired by the U-Net architecture (Isola et al., 2017). The discriminator is a PatchGAN, which operates on local *N* x *N* image patches. It predicts realism for each patch rather than for the full

image and combines the responses across the image to produce the discriminator output (Isola et al., 2017).

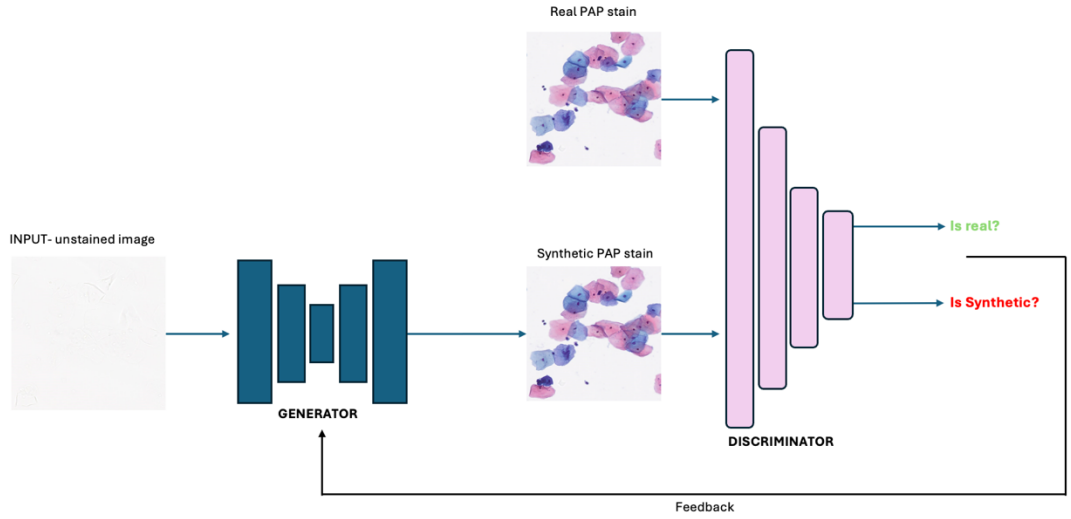


Figure 3.4. Overview of pix2pix architecture with cytology images. Author's own work (2026)

3.3.2 CycleGAN

In this work, a CycleGAN architecture was implemented, which facilitates unpaired image-to-image translation between two distinct visual domains. In this context, the two domains correspond to unstained (domain X) and stained (domain Y) cytology. The objective is to learn one mapping function, G: X→Y and F: Y→X, such that an unstained cytology image can be transformed into a realistically stained version and vice versa. Figure 3.5 provides an overview of the CycleGAN goal, G and F are the generators and D_x and D_y are the discriminators. X and Y represent the images used for training.

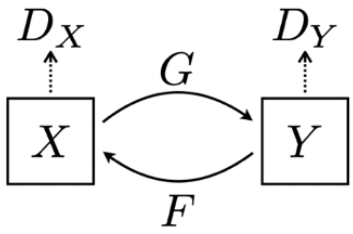


Figure 3.5. Overview of CycleGAN goal. Adapted from Zhu et al., (2020).

This dual structure enables the model to operate without requiring paired or spatially aligned images. This is important in cytology, as perfectly corresponding stained and unstained versions of the same microscopic field are very difficult to obtain.

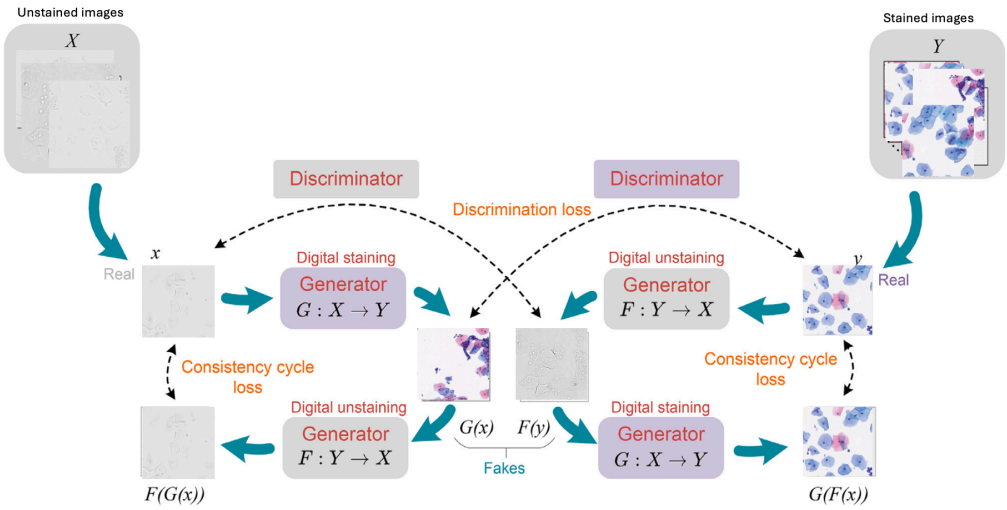


Figure 3.6. Workflow of CycleGAN with cytological images. The figure illustrates the forward and backward cycle consistency and constrains. Adapted from Salido et al., (2023) with cytological images.

To ensure that the generators retain the essential structural information of the image while learning the mappings, CycleGAN introduces the cycle consistency, which is particularly important in cytology where nuclear morphology, cytoplasmic boundaries and the nucleus-cytoplasm ratio have diagnostic significance, as can be seen in Figure 3.6. This constraint stipulates that when an image is translated from one domain to another and then mapped back to its original domain, the result should approximate the original image.

3.3.3 CycleGAN-Turbo and pix2pix-Turbo

The architectures of CycleGAN-Turbo and pix2pix-Turbo are the same. The main difference is whether training is performed with paired (pix2pix-turbo) or unpaired (CycleGAN-turbo).

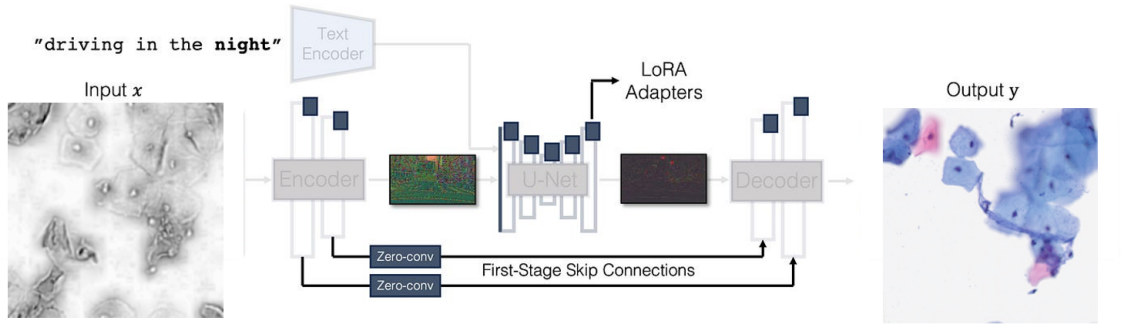


Figure 3.7. Workflow of CycleGAN-Turbo and pix2pix-Turbo with cytological images. Adapted from Parmar et al., (2024)

The generator architectures of CycleGAN-Turbo and pix2pix-Turbo consolidate three distinct modules derived from the original latent diffusion models into a unified, end-to-end framework with a limited number of trainable parameters. The key elements are the low-rank adaptation (LoRA) weights, skip connections and the trainable first layer, as illustrated in Figure 3.7. The LoRA weights are essential for enabling the backbone model to adapt to new conditions. The skip connections are important to keep the high-frequency details of the input image. The model's ability to adapt to new domains and tasks can be improved further by trainable first layer (Parmar et al., 2024).

The Figure 3.7 shows that LoRA layers are added to various layers in the U-Net, the skip connections are introduced between the encoder and decoder via zero-convolutions and the text encoder is introduced in the first convolutional layer of the U-Net.

3.3.3.1 Unpaired Training

CycleGAN-Turbo is designed to learn two translation processes: one that converts images from the source domain into the target domain and another that performs the reverse conversion. These translations are guided by the corresponding textual captions for each domain (Parmar et al., 2024).

3.3.3.2 Paired Training

In paired training scenarios, pix2pix-Turbo is designed to learn a single translation pathway that converts images from the source domain into the corresponding target domain, guided by the associated textual captions (Parmar et al., 2024).

3.4 Data Evaluation

It was selected the best result in terms of cytoplasms, nucleus and staining, both for quantitative and qualitative assessment. This section describes how each evaluation was performed.

3.4.1 Structural Similarity Index

The SSIM was calculated using the test patches (output) and their corresponding ground truth patches. Then, the average of all results per image was calculated. The closer the value is to 1, the more realistic the image will be.

3.4.2 Peak Signal-to-Noise Ratio

The PSNR was calculated by comparing the target patch with the ground truth patch. Then, the average of all results was calculated. Higher values indicate better image quality. The range of PSNR values is from 0 to positive infinity.

3.4.3 Visual Turing Test

The questionnaire, which was conducted on Google Forms, consisted of presenting 60 images, displayed randomly, 30 of which were real and 30 were synthetic. To avoid duplicate responses, those who completed the questionnaire were asked to write down their email address. The email collection was anonymised.

Participants were asked about their area of expertise. Specifically, it was asked whether they were pathologists or cytotechnicians. They were asked about their years of experience.

Two questions were asked about each image. 'Is the image real or synthetic?' and 'How confident are you in your answer?'

Participants were not informed of the proportion of synthetic and real images and answered without having been trained to distinguish a synthetic image. To prevent time pressure from influencing decision-making, no time limit was set, ensuring that responses primarily reflected the ability to distinguish real images from synthetic ones.

The mean accuracy, sensitivity, and specificity of the 12 participants were calculated. Wilcoxon signed- rank (WSR) test was used to assess if participants were able to distinguish real from synthetic image. Mann- Whitney U test was used to compare accuracy between cytotechnicians and pathologists. The Spearman and Pearson test were used to evaluate the impact of years of experience in general and according to professional area. D-prime was used to assess whether the participant is guessing by chance or not stratified by professional group.

The R version 4.5.2 was used for the statistical analyses with the significance level of $p < 0.05$.

Chapter 4

Results

This chapter presents the experimental results obtained in the five experiments conducted for this dissertation. Several experiments were conducted with different algorithms, and the results of the experiments are described below. The summary of the algorithm configuration for each experiment is presented in table form, with the most relevant topics from the experiment. The results table contain the input, output, and ground truth, making it easier to compare the obtained results. All images were chosen at random from set test.

The model that produced the best visual quality was selected by an experienced cytotechnologist, based on the definition of the cytoplasm, nuclear morphology, and staining appearance. Only this selected model was subsequently evaluated using SSIM, PSNR, and the Visual Turing Test (VTT). The results of these evaluations are presented in Section 4.6.

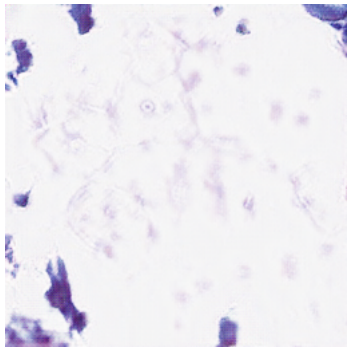
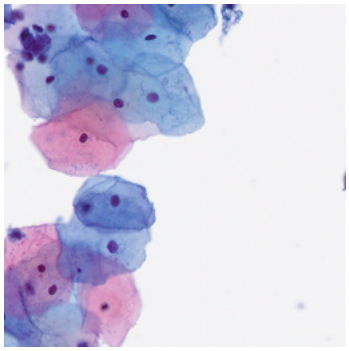
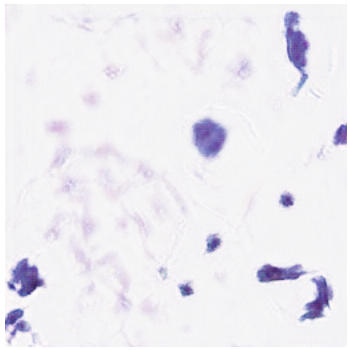
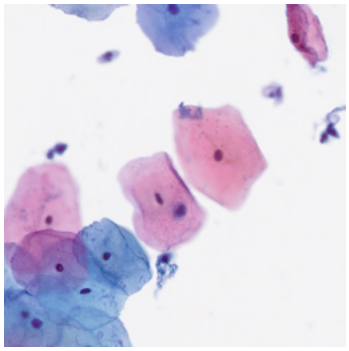
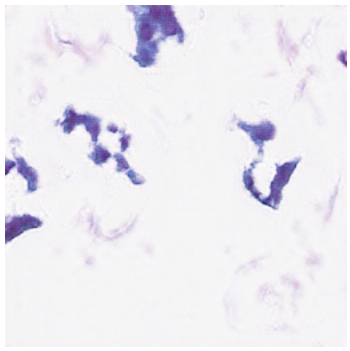
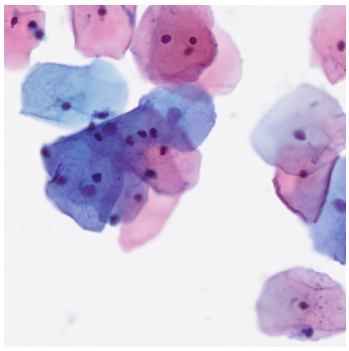
4.1 First Experiment – CycleGAN (without pre-processing)

This experiment was performed on NVIDIA Geforce RTX 3070, with Python 3.11. Table 4.1 contains a summary of the CycleGAN configurations, and Table 4.2 contains the images obtained from this experiment.

Table 4.1. Summary of the algorithm configuration for the first experiment.

Topic	Summary
Dataset	Dataset1: 1,094 training patches and 278 test patches
Model	CycleGAN
Training mode	Unaligned
Translation direction (A⇒B)	A=unstained cytology patches; B=stained cytology patches
Initial resizing	286 x 286 pixels
Random cropping	256 x 256 pixels
Data augmentation	Random horizontal flips
Generators	ResNet-based
Discriminators	Basic PatchGAN
Batch size	1
Optimiser	Adam
Initial learning rate	0.0002

Table 4.2. Results of first experiment using CycleGAN without pre-processing.

Input	Output	Ground Truth
		
		
		

Examining Table 4.2, the output patches does not appear realistic, there are no cellular structures clearly distinguishable and many artefacts are observed, some of them are bluish and some are pinkish. The digital staining cannot be reliably assessed. As can be seen, the input patches are noticeably clear, and it is difficult to distinguish the cellular structures. In the ground truth patches, a clean background and cells can be observed, some clustered and with different cytoplasm colours (bluish/pinkish variations). Most cytoplasms are well defined. The cell nuclei are clearly visible, some more pyknotic than others.

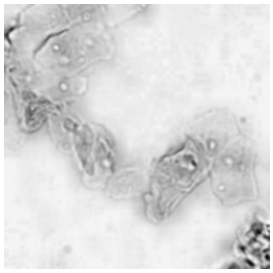
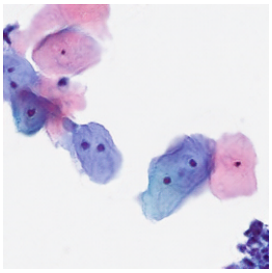
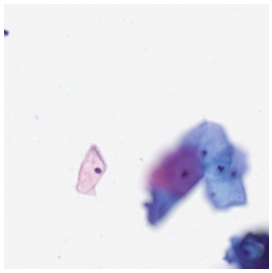
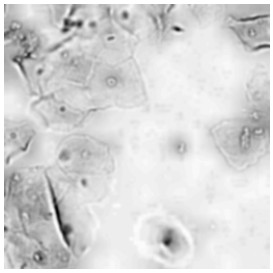
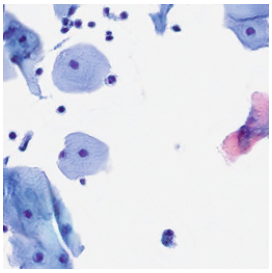
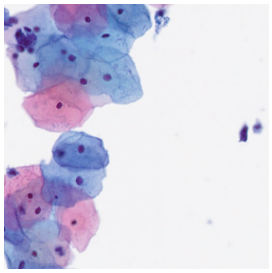
4.2 Second Experiment – CycleGAN (with CLAHE pre-processing)

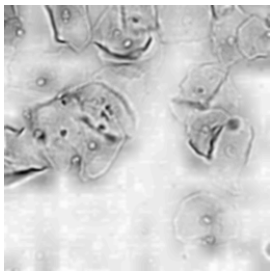
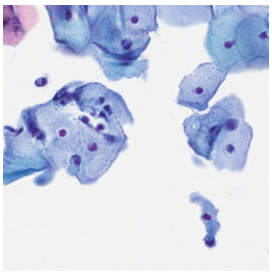
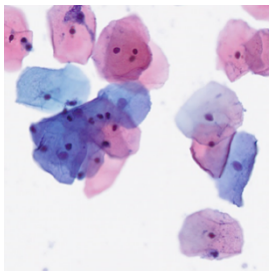
This experiment was performed on NVIDIA Geforce RTX 3070, with Python 3.11. Table 4.3 contains a summary of the CycleGAN configurations, and Table 4.4 contains the images obtained from this experiment.

Table 4.3. Summary of the algorithm configuration for the second experiment.

Topic	Summary
Dataset	Dataset2: 1,094 training patches and 278 test patches
Model	CycleGAN
Training mode	Unaligned
Translation direction (A⇒B)	A=unstained cytology patches; B=stained cytology patches
Initial resizing	286 x 286 pixels
Random cropping	256 x 256 pixels
Data augmentation	Random horizontal flips
Generators	ResNet-based
Discriminators	Basic PatchGAN
Batch size	1
Optimiser	Adam
Initial learning rate	0.0002

Table 4.4. Results of digital staining using CycleGAN with pre-processed images.

Raw patches	Input	Output	Ground Truth
			
			

As can be seen in Table 4.4, the image appears realistic and the cellular structures are clearly distinguishable. The nuclear shape and size are consistent with the real cytological images. However, not all the cytoplasmic membranes are, some of them are incomplete. Preprocessing with CLAHE shows input images with greater contrast, making it easier to see where the cells are and distinguish between cytoplasm and nuclei. Although the digital staining appears more realistic, a comparison of the output of the patches with the ground truth reveals that there are more pink cells in the ground truth.

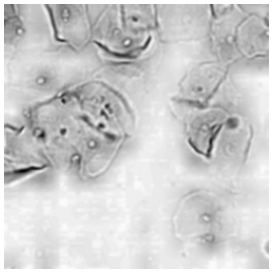
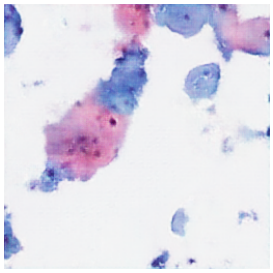
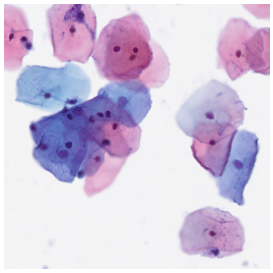
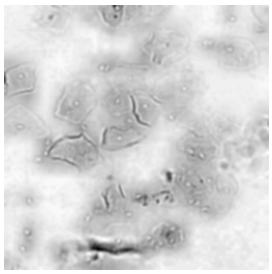
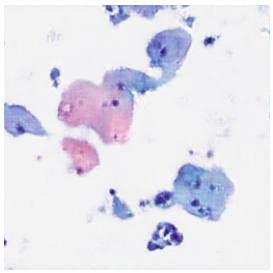
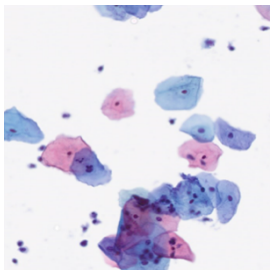
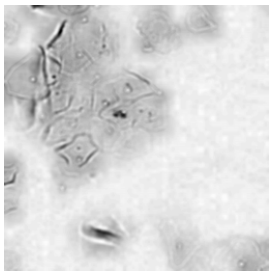
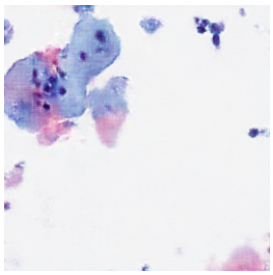
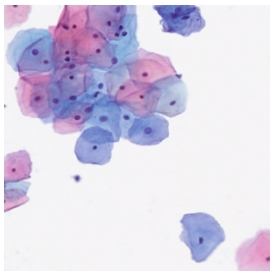
4.3 Third Experiment – pix2pix (with CLAHE pre-processing)

This experiment was performed on NVIDIA Geforce RTX3070, with Python 3.11. Table 4.5 contains a summary of the pix2pix configurations, and Table 4.6 contains the images obtained from this experiment.

Table 4.5. Summary of the algorithm configuration for the third experiment.

Topic	Summary
Dataset	Dataset2: 1,094 training patches and 278 test patches
Model	pix2pix
Training mode	Aligned
Translation direction (A⇒B)	A=unstained cytology patches; B=stained cytology patches
Initial resizing	286 x 286 pixels
Random cropping	256 x 256 pixels
Data augmentation	Random horizontal flips
Generator	U-Net-based
Discriminator	PatchGAN
Batch size	1
Optimiser	Adam
Initial learning rate	0.0002

Table 4.6. Results of digital staining using pix2pix with pre-processed images.

Raw patches	Input	Output	Ground Truth
			
			
			

Examining Table 4.6, the output images do not appear realistic. Many cells in the input do not appear in the output. Additionally, there is a lack of cytoplasm membranes in the cells that do appear. Cellular structures such as nuclear membranes, cytoplasmic membranes, and nucleus are not clearly distinguishable. The shape and size of the nucleus are inconsistent with real cytological images. It is difficult to evaluate digital staining here, as it is difficult to determine the location of the cytoplasm membranes.

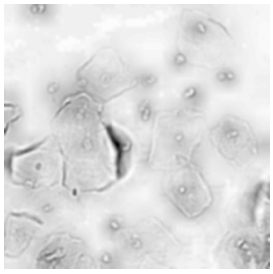
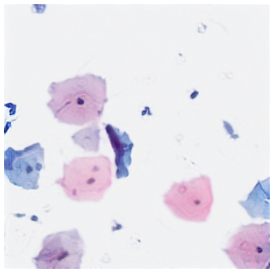
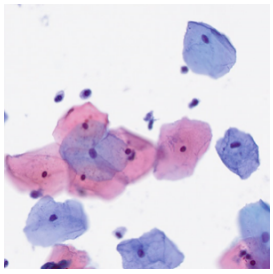
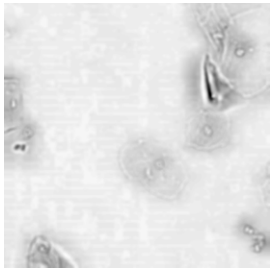
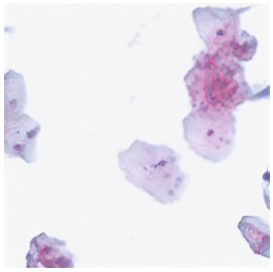
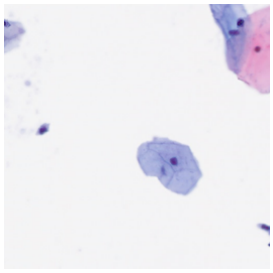
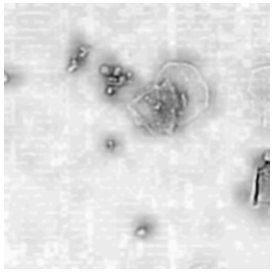
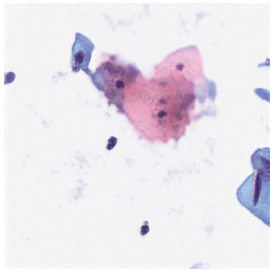
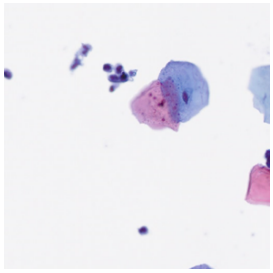
4.4 Fourth Experiment – CycleGAN-Turbo (with CLAHE pre-processing)

This experiment was performed on NVIDIA A100 80GB PCIe, with Python 3.13. Table 4.7 contains a summary of the CycleGAN-Turbo configurations, and Table 4.8 contains the images obtained from this experiment.

Table 4.7. Summary of the algorithm configuration for the fourth experiment.

Topic	Summary
Dataset	Dataset2: 1,094 training patches and 278 test patches
Model	CycleGAN-Turbo
Training mode	Unaligned
Translation direction (A⇒B)	A=unstained cytology patches; B=stained cytology patches
Resolution	512 x 512 pixels
Translation model	One-step diffusion image-translation models (A→B and B→A) using SD-Turbo with LoRA adapters
Trainable adapters	LoRA rank: U-Net=64; VAE=4
Batch size	1
Optimiser	AdamW
Initial learning rate	0.0001

Table 4.8. Results of digital staining using CycleGAN-Turbo with pre-processed images.

Raw patches	Input	Output	Ground Truth
			
			
			

As can be seen on Table 4.8, the images do not appear realistic. This means that cellular structures are partially distinguishable. The shape and size of the nucleus are consistent with real cytological images. When comparing the input with the ground truth patches, it can be observed that some cells are missing in the ground truth patches, and vice versa. It is difficult to evaluate the digital staining in output patches. However, cells are stained in a much lighter pink than the ground truth cells.

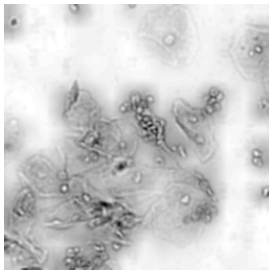
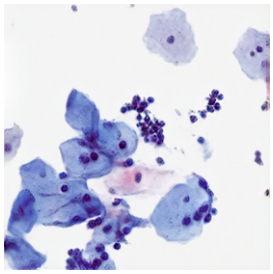
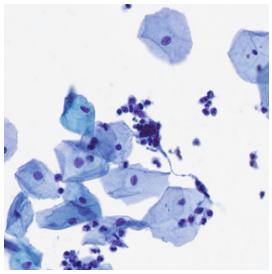
4.5 Fifth Experiment - pix2Pix-Turbo (with CLAHE pre-processing)

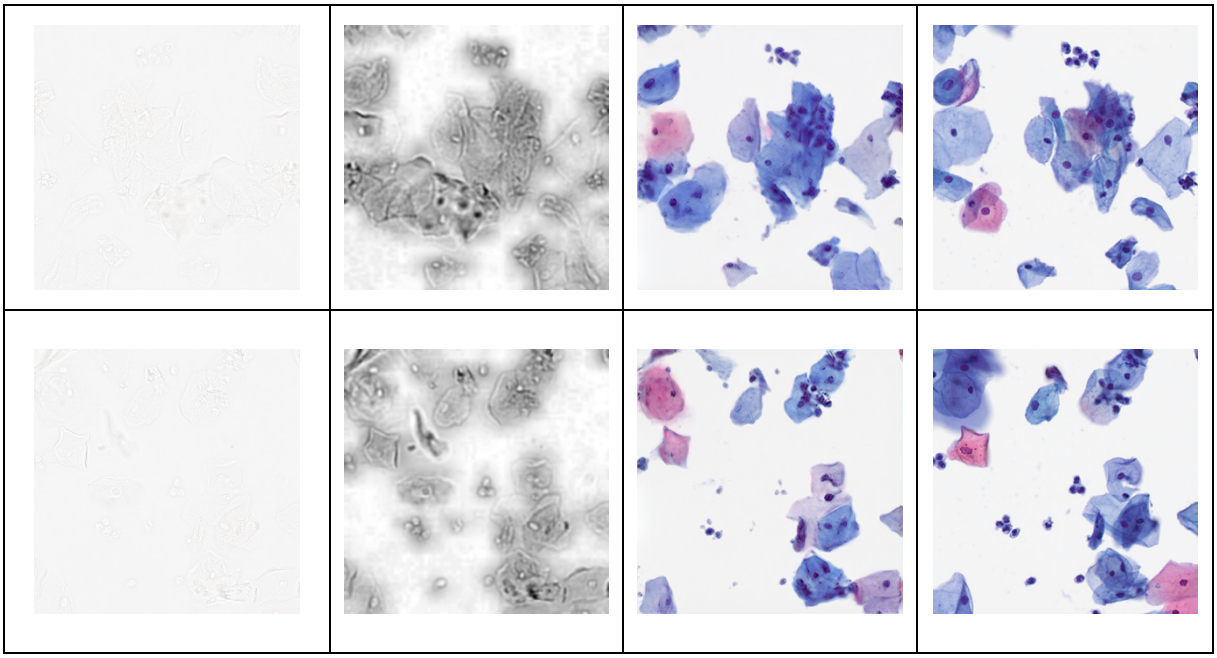
This experiment was performed on NVIDIA A100 80GB PCIe, with Python 3.13. Table 4.9 contains a summary of the pix2pix-Turbo configurations, and Table 4.10 contains the images obtained from this experiment.

Table 4.9. Summary of the algorithm configuration for the fifth experiment.

Topic	Summary
Dataset	Dataset2: 1,094 training patches and 278 test patches
Model	pix2pix-Turbo
Training mode	Aligned
Translation direction (A⇒B)	A=unstained cytology patches; B=stained cytology patches
Resolution	512 x 512 pixels
Translation model	One-step diffusion image-translation model with LoRA adapters
Trainable adapters	LoRA rank: U-Net=128; VAE=4
Batch size	4
Optimiser	AdamW
Initial learning rate	0.0001

Table 4.10. Results of digital staining using pix2pix-Turbo with pre-processed images.

Raw patches	Input	Output	Ground Truth
			



Examining Table 4.10, the output patches appear very realistic, it means that cellular structures are clearly distinguishable. The shape and size of the nuclei are consistent with the real cytological images. However, some of the nuclei appear less distinct in the output. Regarding the evaluation of digital staining, the cells are stained similarly to the ground truth.

4.6 Evaluation

In this section, the model selected by the cytotechnologist that produced the best results in terms of cellular characteristics and staining was evaluated using SSIM, PSNR and VTT.

4.6.1 Structural Similarity Index – pix2pix-Turbo

The SSIM was calculated using the test patches (ground truth and output) and the mean of SSIM was 0.76.

4.6.2 Peak Signal-to-Noise Ratio – pix2pix-Turbo

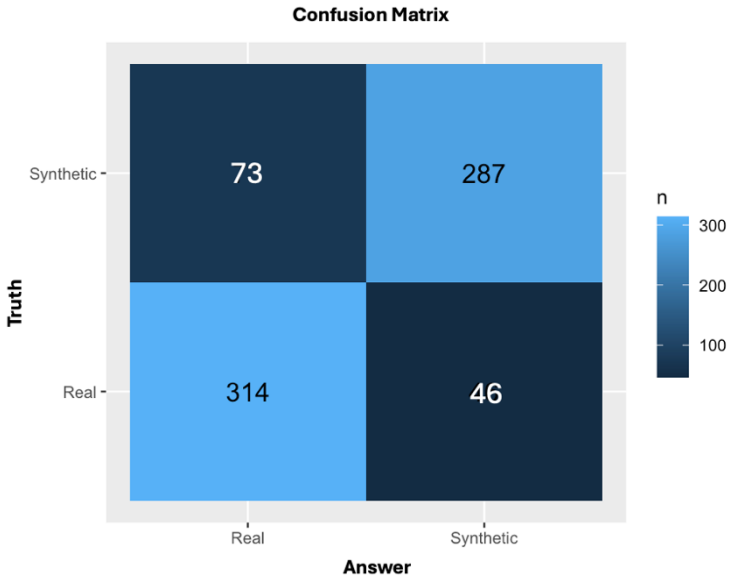
The mean PSNR obtained was 18.18, which was calculated using the test patches (ground truth and output).

4.6.3 Visual Turing Test

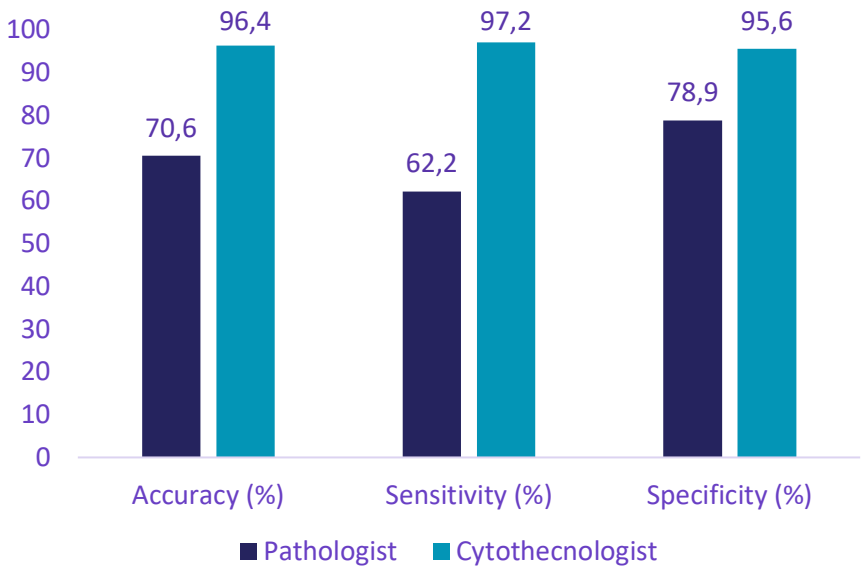
The mean accuracy, sensitivity, and specificity of the 12 participants were 83.5%, 79.7% and 87.2%, respectively. Table 4.11 shows the confusion matrix. When “Synthetic” was considered the positive

class, the sensitivity for identifying synthetic images as such was 79.7%, while the specificity for identifying real images as such was 87.2%. The overall accuracy was 83.5%.

Table 4.11. Confusion Matrix

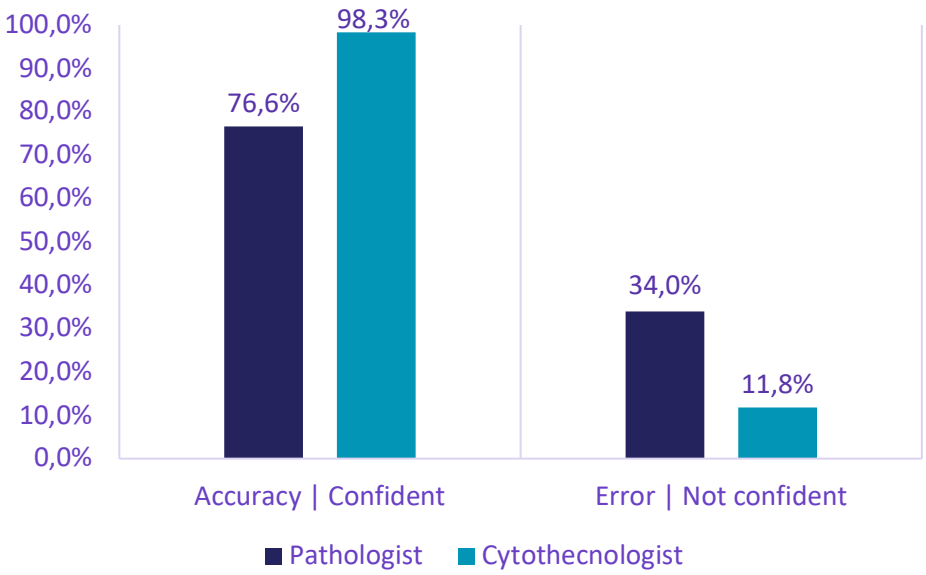


According to the Wilcoxon signed-rank test there was no significant statistical difference between the accuracy in distinguishing “real” and “synthetic” images with a p-value of 0.677. However, as can be seen in Graph 4.1 cytotechnologists had better accuracy when compared to pathologists, with a Mann-Whitney p-value of 0.0047.



Graph 4.1. Accuracy, sensitivity, and specificity between pathologist and cytotechnicians

Among pathologists, the Graph 4.2 shows, that being “not confident” is associated with errors (34%). Regarding cytotechnologists, reporting “confident” almost means correct (98.3%), but “not confident” does not necessarily mean they were incorrect (the error rate is 11.8%).

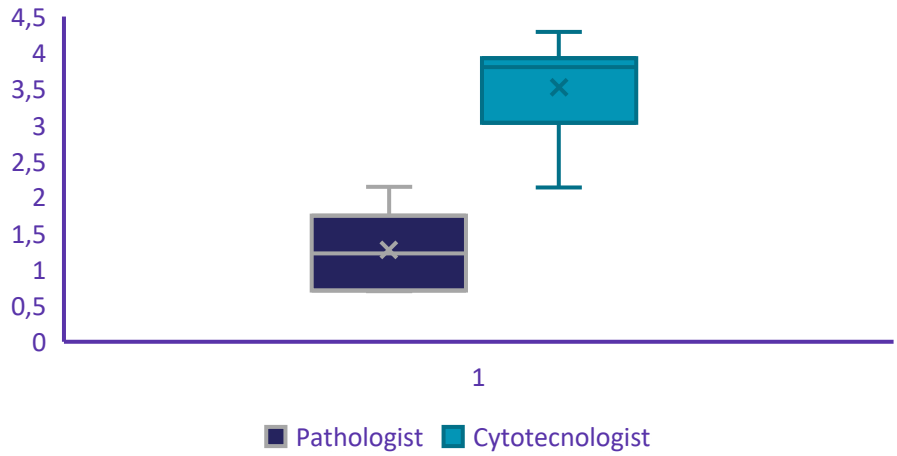


Graph 4.2. Accuracy-confident and error-not confident between pathologist and cytotechnologist.

In these samples, years of experience showed no significant association with accuracy (Spearman p=0.229, p=0.475; Pearson r=0.164, p=0.610). In a model controlling for professional area, experience remained non-significant (β=0.0039, p=0.201).

4.6.3.1 D-Prime

Graph 4.3 shows that d-prime indicated good global discrimination in guessing if the image was real or synthetic (d’ > 1.96). Cytotechnologists revealed a substantially better ability to discriminate whether the image was real or synthetic (d’=3.52 vs. pathologists with d’=1.27) with a p-value in Welch t-test of 0.000192 and in the Mann-Whitney U test, with a p-value of 0.008.



Graph 4.3. D’ between pathologist and cytotechnologist.

Chapter 5

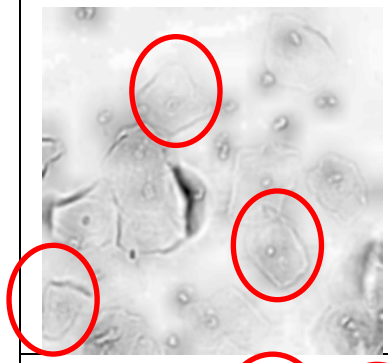
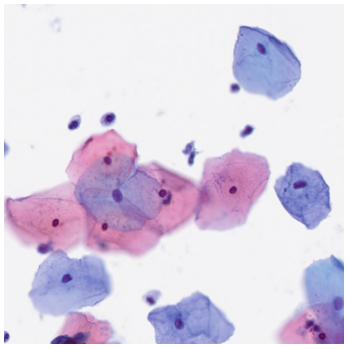
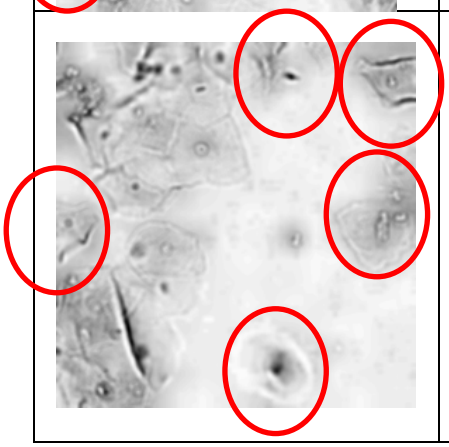
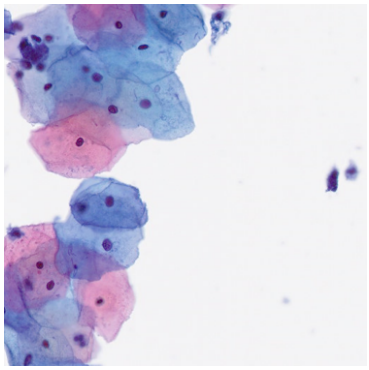
Discussion

Cervical cytology presents many challenges for AI, and this dissertation revealed further issues regarding missing cells, the need for pre-processing, image acquisition and proper alignment.

In the first experiment, the CycleGAN did not produce outputs with well-defined cellular structures when very bright input images were used. Instead, the generated images contained artefacts, which hindered the visual interpretation. These findings indicated that pre-processing was necessary. It involved the use of the CLAHE method. This method improved the definition of the cell contours and nuclei, and it was used in the subsequent experiments.

In the second experiment, images with CLAHE pre-processing were used with the same algorithm, as in the first experiment (CycleGAN), achieving significantly improved results. This made it possible to obtain outputs with well-defined cells and realistic cellular features.

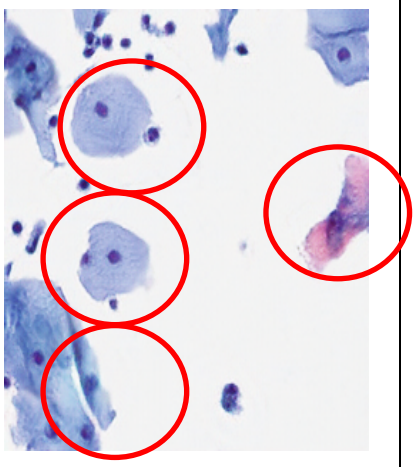
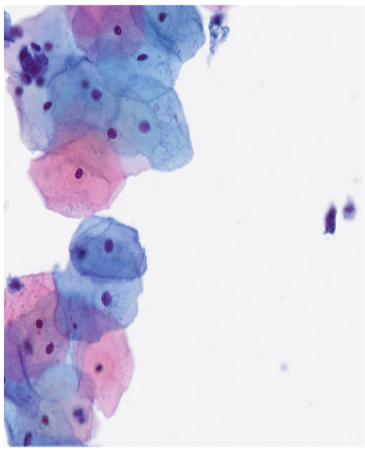
Table 5.1. Patches with missing cells.

Input	Ground Truth
	
	

A cell loss of approximately 29% was observed between input patches and ground truth patches. This indicates that the number of cells in the input and ground truth are not the same. As can be seen in Table 4.1 the input patches are marked with red circles, indicating the locations of cells that are absent in the ground truth patches.

It is unclear whether this loss was due to the staining itself or to the process of removing the coverslip from the slide, which both are abrasive methods for the cells, even though the coverslip came off on its own. This result may impact clinical practice when samples have an insufficient number of cells for evaluation after staining. The Bethesda system stipulates that there must be more than 5000 cells on each LBC for a diagnosis to be possible (Nayar & Wilbur, 2015). The cytologies selected for this study contained sufficient cells for a diagnosis, before staining and after staining. Nevertheless, the lack of cells also impacts the calculation of SSIM and PSNR, as elements present in the output patches will not be present in the ground truth patches. Therefore, SSIM and PSNR should be interpreted with caution to provide guidance on quality of output patches.

Table 5.2. A detailed analysis of the CycleGAN output.

Input	Output-CycleGAN	Ground Truth
		

In Table 5.2 it can be observed incomplete cytoplasmatic membranes, and cell shapes that differ from those observed in the input patches. The red circles showing where the incomplete cells are.

Although the digital staining appears more realistic, a comparison with the ground truth reveals a discrepancy in colour distribution: the cells in the output images have more pronounced blue tone,

whereas the ground truth patches contain blue and pink cytoplasms. This suggests that the staining does not fully correspond to the ground truth patches.

In the third experiment, the pix2pix algorithm was used, which requires aligned pixel-to-pixel inputs. In their article on histology, Yang et al. (2022) discuss one of the main challenges in training deep neural networks for staining transfer is the scarcity of suitable training data. Since slides are typically stained with a single stain, and the washing steps required for an additional staining can compromise tissue integrity, it is challenging to collect paired images with different stains. In this study, there was no transfer of staining, but staining happened after obtaining the WSI of the unstained slide. The slide was also washed several times during the staining process, and cells are more susceptible to washing than tissues. This aspect directly affects supervised algorithms that require pixel-to-pixel paired images. This data dilemma may explain the differences observed both pixel-to-pixel and when comparing the outputs with the ground truth.

However, although they appeared aligned to the naked eye, the scanner used did not allow pixel-to-pixel alignment. This is because the scanner does not always start in the same position on the slides. Although coordinates (x, y) are recorded when images are collected in the scanner programme, these do not correspond to images of stained and unstained cells.

Table 5.3. Example of the pix2pix output

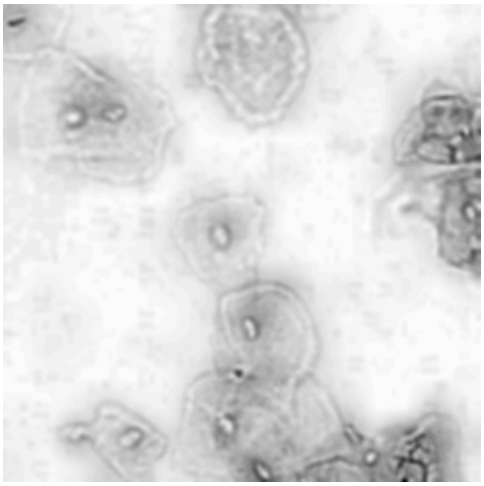
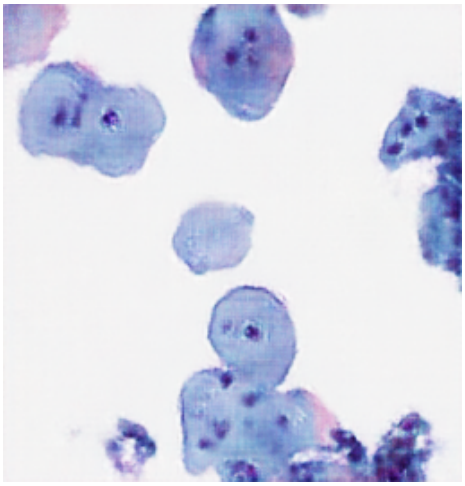
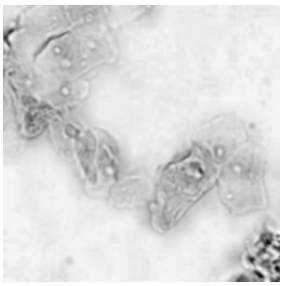
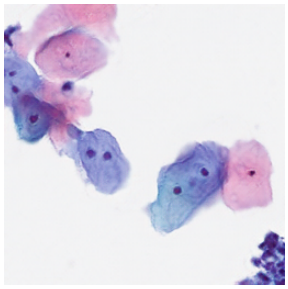
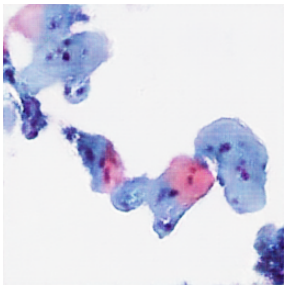
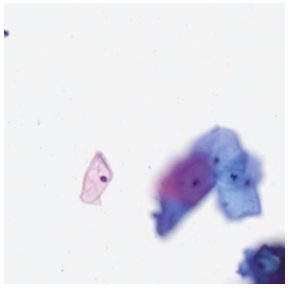
Input	Output-pix2pix
	

Table 5.3 shows an example of the outputs obtained with this model. The output has blue cytoplasm and pink edges, which do not correspond to traditional staining methods. Cells produced using this model appear pixelated.

Table 5.4. Comparison of the same patches

Input	Output- CycleGAN	Output-pix2pix	Ground truth
			

When comparing the results obtained with the CycleGAN and pix2pix model, CycleGAN is superior. The Table 5.4 shows the key differences between pix2pix and CycleGAN outputs. The cells obtained with CycleGAN are more clearly defined, with visible nuclei and distinct nuclear membranes. However, there are some cells in which this is not the case. One possible explanation is that the input patches in the pix2pix model are not perfectly aligned, which causes it to perform worse than the CycleGAN model, which does not require paired images.

As can be seen in Table 5.4, the generated outputs tend to preserve the cellular morphology visible in the input image rather than reproducing the ground truth image. This behaviour is expected because the models generate the digitally stained image from the structural information available in the input patch. Therefore, when there are differences between the input and the ground truth, such as pixel misalignment or cell loss after staining, the output may appear morphologically closer to the input than the ground truth. This suggests that CLAHE helped expose relevant cellular contours and nuclear details, guiding the models towards preserving input morphology while learning a Papanicolaou staining appearance.

For the fourth and fifth experiments, CycleGAN-turbo and pix2pix-Turbo were trained, respectively. Table 5.5 shows an example of the outputs obtained using pix2pix-Turbo. In this scenario, where pairs exist (with pixel-to-pixel imperfect registration), pix2pix-turbo provides a more realistic image (b) and preserves structure more effectively. In contrast, CycleGAN-Turbo, operating in unpaired mode, maintains ambiguity and potential unwanted structural changes, as shown in Table 5.6.

Table 5.5. Example of output obtained using pix2pix-Turbo with the respective ground truth.

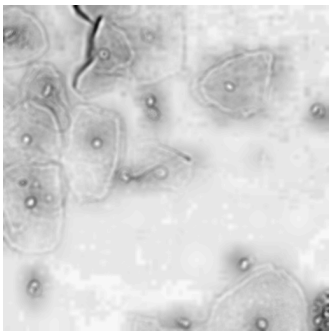
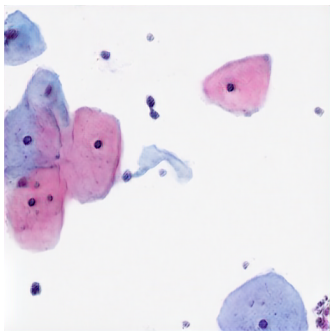
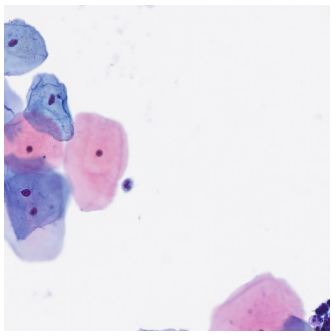
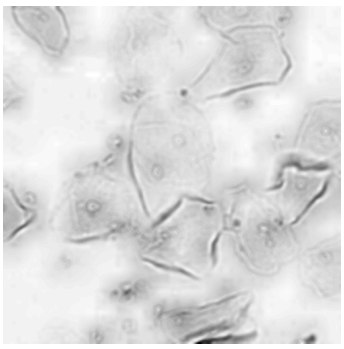
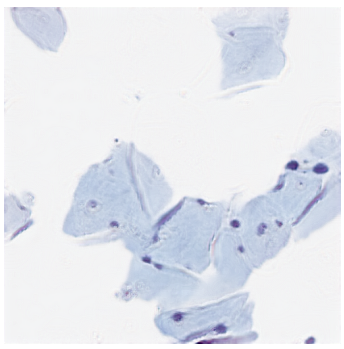
Input	Output-pix2pix-Turbo	Ground truth
		

Table 5.6. Example of output obtained using CycleGAN-Turbo.

Input	Output-CycleGAN-Turbo
	

Comparing the results of all experiments, pix2pix-Turbo produced better visual results. This can be explained by the fact that pix2pix-Turbo is supervised and uses a pre-trained diffusion backbone. In pix2pix, being a GAN that is trained from scratch, the misalignment is more susceptible than pix2pix-Turbo. The versions without Turbo cannot be trained with higher-resolution images. Training can initially start with 512 x 512 pixels images, but these are first resized to 286 x 286 pixels and then randomly cropped to 256 x 256 pixels. This causes a loss of quality during training.

The Turbo versions require higher-resolution images (512 x 512 pixels) than non-Turbo versions and use pre-trained diffusers as part of their algorithms, whereas non-Turbo versions are trained from scratch. However, when comparing the results of pix2pix-Turbo and CycleGAN-Turbo, pix2pix-Turbo performed better. The fact that pix2pix-Turbo uses a pre-trained diffuser, may reveal that 100% pixel-to pixel alignment is not needed.

Training is performed from X to Y and from Y to X in CycleGAN-Turbo, with cycle consistency added. However, this model requires more computing power to train it than the same batch size as

pix2pix-Turbo. Therefore, CycleGAN-Turbo was trained with batch size of 1 and pix2pix-Turbo with a batch size of 4, both maintaining an image resolution of 512 x 512 pixels.

At the end of the experiments, the results from pix2pix-Turbo were selected to perform quantitative and qualitative evaluation, because these results were the ones that most closely reflected reality. Aiming for a quantitative way to rate performance, SSIM and PSNR were calculated in pix2pix-Turbo with lower values than previously reported with CytoGAN by Wang et al. (2024) when transferring from PAP smear staining to H&E (0.76 and 18.18 in this dissertation vs. 0.993 and 34.5)

The absence of cells impacts the calculation of SSIM and PSNR, since elements present in the output patches are not present in the ground truth patches. Therefore, SSIM and PSNR values cannot be relied upon to provide guidance on the quality of the output patches.

To evaluate the reliability of digitally stained cytology, a VTT questionnaire was used with pix2pix-Turbo images. Twelve participants assessed 60 images, 30 were real and 30 were synthetics. Only 20.28% of participants said that the image is “real” when a “synthetic” one was shown and 12.78% stated “synthetic” when it was “real”. This could be explained by some improvements that need to be done in what concern image quality and cellular definition. No statistically significant difference between false positive and false negatives are related with the lack of training in detecting synthetic images because none of the participants were previously exposed to it.

Cytotechnologists had a better accuracy and confidence when compared to pathologists, regardless of the years of experience. This is grounded in some core differences in their daily work; cytotechnologists usually do the screening of the slides searching for the quality of the sample, lesions and they point out the area of interest with pathology for the pathologist to make the diagnosis. So, pathologists rarely see slides without disease. This allows cytotechnologists to regularly train their ability to distinguish cellular characteristics, as can be seen in the difference of discrimination capacity between professional areas (d´1.27 in pathologists Vs 3.52 in cytotechnologists).

Although GAN-based image translation models may introduce hallucinated features that could have diagnostic implications, no such hallucinations were identified in the present study. The dataset consisted exclusively of cervical cytology images without pathological alterations, and the generated images were assessed for artificial cellular features that could alter the diagnosis. No hallucinated cells or morphological alterations capable of modifying the cytological diagnosis were detected. Nevertheless, these findings should be interpreted in the context of the dataset used. As the model was not trained or evaluated using pathological samples, its ability to preserve abnormal cellular features is unknown. Further studies including representative images of different cytological abnormalities are therefore required to determine whether the model can accurately reproduce pathological features that are relevant to diagnosis without introducing, modifying or omitting cellular characteristics that could affect clinical interpretation.

Conclusion

This dissertation explored digital staining for cervical cytology, addressing four guiding objectives: dataset creation, similarity to chemical staining, model effectiveness, and the prospect of future equivalence between digital and chemical staining. These objectives were partially achieved, as the results demonstrated progress across all four areas, although important limitations remain to be addressed.

Some limitations were encountered during the building of the cervical cytology dataset, such as pixel-to-pixel misalignment and the loss of 29% of cells during the traditional staining process. The percentage of cells lost was calculated by comparing the average number of missing cells in the ground truth patches with the number of cells in the input patches.

Throughout the experiments, it was possible to transform unstained patches into outputs that resemble stained cytology to varying degrees. The best performing models were CycleGAN and pix2pix-Turbo. The latter showed superior performance.

The early CycleGAN results were affected by artefacts and insufficient definition, suggesting that raw unstained inputs might not be suitable for consistent conversion. In other words, the reduced contrast of raw images causes CycleGAN to produce outputs without well-defined cellular structures.

However, introducing CLAHE improved the visibility of cellular contours and the clarity of nuclei, leading to more realistic outputs. This suggests that pre-processing is an essential step for this task rather than an optional enhancement.

The reliability assessment in this dissertation combined quantitative metrics and human evaluation. Although SSIM and PSNR were computed for the selected pix2pix-Turbo results (0.76 and 18.18 respectively), the observed cell loss and imperfect correspondence between ‘paired’ patches limits the interpretability of these metrics. Additionally, the VTT questionnaire revealed that participants sometimes misclassified real and synthetic patches, indicating that while generated images can be convincing, they are not consistently indistinguishable. These findings suggest a considerable degree of visual realism but also indicate that synthetic images cannot yet be considered fully relevant in clinical contexts.

Based on the evidence gathered, this dissertation does not yet support the claim that digitally stained cytology is equivalent to chemically stained cytology for diagnostic purposes. The dataset consisted of negative cytology, and the evaluation focused on visual plausibility and perceptual realism rather than diagnostic endpoints, such as lesion detection or grading.

In conclusion, the results suggest that digital staining may be feasible under controlled conditions. However, they also indicate that current limitations in cytology prevent definitive claims of equivalence with chemically stained slides.

Limitations and Restrictions

This dissertation produced promising results, but several limitations and practical constraints were identified during its development.

The first major limitation relates to the digitalisation process of the cytology slides. During scanning, some regions of the whole-slide images exhibited inconsistent focus, including blurred areas and cells appearing in different focal planes. This is particularly relevant in cytology, where samples are not always perfectly in monolayer and may contain overlapping cellular material. Consequently, some image regions may not provide consistent morphological information for model training or visual assessment. With the benefit of hindsight, it would have been useful to benchmark the different scanners and scanning settings (e.g. illumination conditions, z-stacking configuration, focus strategies and pixel alignment) prior to finalising the data acquisition process and constructing the datasets. Another limitation of the scanner was that it was not possible to align pairs of images pixel-to-pixel. This aspect is particularly important for supervised algorithms such as pix2pix and pix2pix-Turbo.

Secondly, the dataset size used for training and evaluation is an important restriction. For example, in their article, Rivenson et al., (2019) used several datasets comprising over 2,000 training patches and over 500 validation patches.

A third limitation concerns the interpretation of quantitative similarity metrics, specifically SSIM and PSNR. In this case, these metrics could not be considered reliable because the ground truth and input images did not contain the same cellular content. As SSIM and PSNR are based on pixel-wise structural correspondence, discrepancies in cellular presence can result in distorted values.

Regarding the expert questionnaire study, it was not possible to include examples from all experiments (CycleGAN, pix2pix and CycleGAN-Turbo) without the survey becoming excessively long. This would likely result in lower participation rates. Therefore, it was decided that only output images from experiment using the pix2pix-Turbo model would be included, since this model performed better than the others. Thus, feasibility and participant engagement were prioritised over comprehensive experimental coverage. This restricts the interpretation of the survey outcomes since the Visual Turing Test does not compare all model variants.

Finally, the project was constrained by computational resources. Training and testing these kinds of algorithmic models is a computationally demanding. Access to suitable hardware, such as GPUs, was limited, reducing the number of experiments, hyperparameter searches and repetitions that could be performed.

All these limitations and restrictions should be considered when interpreting the results and generalising the conclusions of this dissertation, and they point to clear directions for future improvements.

Future Work

Based on the findings of this dissertation, there are several ways in which the technical performance and clinical relevance of digital staining for cervical cytology can be improved.

Firstly, the digitisation process should be improved by benchmarking scanners and scanning settings that are more suitable for cytology slides, such as pixel-to-pixel alignment, focus strategies, illumination, and other techniques.

Secondly, the process could be improved by adding additional pre-processing steps to address artifacts specific to cytology, such as Laplacian sharpening. This step increases image sharpness by highlighting intensity transitions (edges) (Verma & Barthwal, 2025).

Thirdly, an important extension is the inclusion of pathological cytology cases (e.g. ASC-US, LSIL and HSIL). This enables assessment of whether digital staining preserves the subtle diagnostic cues required for clinically meaningful interpretation. In parallel, the dataset should be expanded in terms of size.

Fourthly, in addition to the Visual Turing Test, a dedicated validation procedure should be included in the evaluation of generated images for the detection of hallucinated features. Complementary expert-based evaluations could be conducted, in which cytotechnologists or cytopathologists would compare source and generated images, identifying potential discrepancies at the cellular level. Attention should particularly be given to erroneous additions or removals of cells, changes in nuclear or cytoplasmic morphology, and when pathological samples are included, the omission of abnormal cells or the artificial generation of features suggestive of malignancy. Each discrepancy could then be classified according to its potential effect on cytological interpretation, including whether it could lead to a false negative, false positive or altered diagnosis. This task-specific validation would provide a more clinically meaningful assessment of model reliability.

Finally, the focus of evaluation should shift from SSIM/PSNR to metrics that more accurately reflect cytology quality. This should be achieved by combining perceptual/feature-based measures with morphology-aware analyses.

Bibliographic References and Sources

AleksioskaPapestiev, I., Chibisheva, V., Micevska, M., & Dimitrov, G. (2018). Prevalence of Specific Types of Human Papilloma Virus in Cervical Intraepithelial Lesions and Cervical Cancer in Macedonian Women. *Medical Archives*, *72*(1), 26. https://doi.org/10.5455/medarh.2018.72.26-30

Bai, B., Yang, X., Li, Y., Zhang, Y., Pillar, N., & Ozcan, A. (2023). Deep learning-enabled virtual histological staining of biological samples. *Light: Science & Applications*, *12*(1), 57. https://doi.org/10.1038/s41377-023-01104-7

Bhat, M., & Tarun Patil M S. (2014). Adaptive clip limit for contrast limited adaptive histogram equalization (CLAHE) of medical images using least mean square algorithm. *2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies*, 1259–1263. https://doi.org/10.1109/ICACCCT.2014.7019300

Chantziantoniou, N., Donnelly, A. D., Mukherjee, M., Boon, M. E., & Austin, R. M. (2017). Inception and Development of the Papanicolaou Stain Method. *Acta Cytologica*, *61*(4–5), 266–280. https://doi.org/10.1159/000457827

Cheng, L., Wang, Y., & Du, J. (2020). Human Papillomavirus Vaccines: An Updated Review. *Vaccines*, *8*(3), 391. https://doi.org/10.3390/vaccines8030391

Cohen, J. P., Luck, M., & Honari, S. (2018). *Distribution Matching Losses Can Hallucinate Features in Medical Image Translation* (pp. 529–536). https://doi.org/10.1007/978-3-030-00928-1_60

Dey, P. (2022). Artificial neural network in diagnostic cytology. *Cytojournal*, *19*, 27. https://doi.org/10.25259/Cytojournal_33_2021

Emery, J. F., Le, M.-K., Levy, J., Vaickus, L., & Liu, X. (2025). Machine learning, deep learning, and artificial intelligence as applied to the field of cytopathology: a comprehensive review. *Journal of the American Society of Cytopathology*. https://doi.org/10.1016/j.jasc.2025.07.004

Falahkheirkhah, K., Tiwari, S., Yeh, K., Gupta, S., Herrera-Hernandez, L., McCarthy, M. R., Jimenez, R. E., Cheville, J. C., & Bhargava, R. (2023). Deepfake Histologic Images for Enhancing Digital Pathology. *Laboratory Investigation*, *103*(1), 100006. https://doi.org/10.1016/j.labinv.2022.100006

Fan, R., Li, X., Lee, S., Li, T., & Zhang, H. L. (2020). Smart Image Enhancement Using CLAHE Based on an F-Shift Transformation during Decompression. *Electronics*, *9*(9), 1374. https://doi.org/10.3390/electronics9091374

Gadermayr, M., Gupta, L., Appel, V., Boor, P., Klinkhammer, B. M., & Merhof, D. (2019). Generative Adversarial Networks for Facilitating Stain-Independent Supervised and Unsupervised Segmentation: A Study on Kidney Histology. *IEEE Transactions on Medical Imaging*, *38*(10), 2293–2302. https://doi.org/10.1109/TMI.2019.2899364

Giansanti, D. (2025). Advancements in Digital Cytopathology Since COVID-19: Insights from a Narrative Review of Review Articles. *Healthcare*, *13*(6), 657. https://doi.org/10.3390/healthcare13060657

Giarnieri, E., & Scardapane, S. (2023). Towards Artificial Intelligence Applications in Next Generation Cytopathology. *Biomedicines*, *11*(8), 2225. https://doi.org/10.3390/biomedicines11082225

Gill, G. W. (2013). *Papanicolaou Stain* (pp. 143–189). https://doi.org/10.1007/978-1-4614-4933-1_10

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, *63*(11), 139–144. https://doi.org/10.1145/3422622

Halloum, K., & Ez-Zahraouy, H. (2025). Enhancing Medical Image Classification through Transfer Learning and CLAHE Optimization. *Current Medical Imaging Formerly Current Medical Imaging Reviews*, *21*. https://doi.org/10.2174/0115734056342623241119061744

Homeyer, A., Geißler, C., Schwen, L. O., Zakrzewski, F., Evans, T., Strohmenger, K., Westphal, M., Bülow, R. D., Kargl, M., Karjauv, A., Munné-Bertran, I., Retzlaff, C. O., Romero-López, A., Sołtysiński, T., Plass, M., Carvalho, R., Steinbach, P., Lan, Y.-C., Bouteldja, N., … Zerbe, N. (2022). Recommendations on compiling test datasets for evaluating artificial intelligence solutions in pathology. *Modern Pathology*, *35*(12), 1759–1769. https://doi.org/10.1038/s41379-022-01147-y

Hong, K.-T., Cho, Y., Kang, C. H., Ahn, K.-S., Lee, H., Kim, J., Hong, S. J., Kim, B. H., & Shim, E. (2022). Lumbar Spine Computed Tomography to Magnetic Resonance Imaging Synthesis Using Generative Adversarial Network: Visual Turing Test. *Diagnostics*, *12*(2), 530. https://doi.org/10.3390/diagnostics12020530

Hore, A., & Ziou, D. (2010). Image Quality Metrics: PSNR vs. SSIM. *2010 20th International Conference on Pattern Recognition*, 2366–2369. https://doi.org/10.1109/ICPR.2010.579

House, J. C., Henderson-Jackson, E. B., Johnson, J. O., Lloyd, M. C., Dhillon, J., Ahmad, N., Hakam, A., Khalbuss, W. E., Leon, M. E., Chhieng, D., Zhang, X., Centeno, B. A., & Bui, M. M. (2013). Diagnostic digital cytopathology: Are we ready yet? *Journal of Pathology Informatics*, *4*(1), 28. https://doi.org/10.4103/2153-3539.120727

Huang, L., Li, Y., Pillar, N., Keidar Haran, T., Wallace, W. D., & Ozcan, A. (2025). A robust and scalable framework for hallucination detection in virtual tissue staining and digital pathology. *Nature Biomedical Engineering*, *9*(12), 2196–2214. https://doi.org/10.1038/s41551-025-01421-9

Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2017). Image-to-Image Translation with Conditional Adversarial Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 5967–5976. https://doi.org/10.1109/CVPR.2017.632

Jang, M., Bae, H., Kim, M., Park, S. Y., Son, A., Choi, S. J., Choe, J., Choi, H. Y., Hwang, H. J., Noh, H. N., Seo, J. B., Lee, S. M., & Kim, N. (2023). Image Turing test and its applications on synthetic chest radiographs by using the progressive growing generative adversarial network. *Scientific Reports*, *13*(1), 2356. https://doi.org/10.1038/s41598-023-28175-1

Jim Nilsson and Tomas Akenine-Möller. (2020). Understanding SSIM. *ArXiv*.

Kalita, M., Mahanta, L. B., Das, A. K., & Nath, M. (2024). A new deep learning model with interface for fine needle aspiration cytology image-based breast cancer detection. *Indonesian Journal of Electrical Engineering and Computer Science*, *34*(3), 1739. https://doi.org/10.11591/ijeecs.v34.i3.pp1739-1752

Kenneth Jensen. (2012, April 26). *CRISP-DM Process Diagram*. Https://Commons.Wikimedia.Org/Wiki/File:CRISP-DM_Process_Diagram.Png.

Kim, D., Sundling, K. E., Virk, R., Thrall, M. J., Alperstein, S., Bui, M. M., Chen-Yost, H., Donnelly, A. D., Lin, O., Liu, X., Madrigal, E., Michelow, P., Schmitt, F. C., Vielh, P. R., Zakowski, M. F., Parwani, A. V., Jenkins, E., Siddiqui, M. T., Pantanowitz, L., & Li, Z. (2024). Digital cytology part 1: digital cytology implementation for practice: a concept paper with review and recommendations from the American Society of Cytopathology Digital Cytology Task Force. *Journal of the American Society of Cytopathology*, *13*(2), 86–96. https://doi.org/10.1016/j.jasc.2023.11.006

Latonen, L., Koivukoski, S., Khan, U., & Ruusuvuori, P. (2024). Virtual staining for histology by deep learning. *Trends in Biotechnology*, *42*(9), 1177–1191. https://doi.org/10.1016/j.tibtech.2024.02.009

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Levine, A. B., Peng, J., Farnell, D., Nursey, M., Wang, Y., Naso, J. R., Ren, H., Farahani, H., Chen, C., Chiu, D., Talhouk, A., Sheffield, B., Riazy, M., Ip, P. P., Parra-Herran, C., Mills, A., Singh, N., Tessier-Cloutier, B., Salisbury, T., … Bashashati, A. (2020). Synthesis of diagnostic quality cancer pathology images. *BioRxiv*. https://doi.org/10.1101/2020.02.24.963553

Levy, J., Jackson, C., Sriharan, A., Christensen, B., & Vaickus, L. (2020). Preliminary Evaluation of the Utility of Deep Generative Histopathology Image Translation at a Mid-sized NCI Cancer Center. *Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies*, 302–311. https://doi.org/10.5220/0009427603020311

Lin, W., Hu, Y., Zhu, R., Wang, B., & Wang, L. (2025). Virtual staining for pathology: Challenges, limitations and perspectives. *Intelligent Oncology*, *1*(2), 105–119. https://doi.org/10.1016/j.intonc.2025.03.005

Marletta, S., Treanor, D., Eccher, A., & Pantanowitz, L. (2021). Whole-slide imaging in cytopathology: state of the art and future directions. *Diagnostic Histopathology*, *27*(11), 425–430. https://doi.org/10.1016/j.mpdhp.2021.08.001

Marth, C., Landoni, F., Mahner, S., McCormack, M., Gonzalez-Martin, A., & Colombo, N. (2017). Cervical cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology*, *28*, iv72–iv83. https://doi.org/10.1093/annonc/mdx220

Meng, X., Li, X., & Wang, X. (2021). A Computationally Virtual Histological Staining Method to Ovarian Cancer Tissue by Deep Generative Adversarial Networks. *Computational and Mathematical Methods in Medicine*, *2021*, 1–12. https://doi.org/10.1155/2021/4244157

Mohammed, M., Khan, M. B., & Bashier, E. B. M. (2016). *Machine Learning*. CRC Press. https://doi.org/10.1201/9781315371658

Mugesh, S. (2023). *Hands-On ML Projects with OpenCV : Master Computer Vision and Machine Learning Using OpenCV and Python* (First). Orange Education Pvt Ltd.

Nayar, R., & Wilbur, D. C. (Eds.). (2015). *The Bethesda System for Reporting Cervical Cytology*. Springer International Publishing. https://doi.org/10.1007/978-3-319-11074-5

Park, H. Y., Bae, H.-J., Hong, G.-S., Kim, M., Yun, J., Park, S., Chung, W. J., & Kim, N. (2021). Realistic High-Resolution Body Computed Tomography Image Synthesis by Using Progressive Growing Generative Adversarial Network: Visual Turing Test. *JMIR Medical Informatics*, *9*(3), e23328. https://doi.org/10.2196/23328

Park, K. S., Moon, J. B., Cho, S.-G., Kim, J., & Song, H.-C. (2023). Applying Pix2pix to Translate Hyperemia in Blood Pool Image into Corresponding Increased Bone Uptake in Delayed Image in Three-Phase Bone Scintigraphy. *Nuclear Medicine and Molecular Imaging*, *57*(2), 103–109. https://doi.org/10.1007/s13139-022-00786-y

Parmar, P. T. , N. S. , & Z. J. (2024). One-Step Image Translation with Text-to-Image Models. *ArXiv*.

Pegah Salehi and Abdolah Chalechale. (2020). Pix2Pix-based Stain-to-Stain Translation: A Solution for Robust Stain Normalization in Histopathology Images Analysis. *ArXiv*.

Peter Chapman. (2000). *CRISP-DM 1.0: Step-by-step data mining guide*. Https://Api.Semanticscholar.Org/CorpusID:59777418.

Pozzi, M., Noei, S., Robbi, E., Cima, L., Moroni, M., Munari, E., Torresani, E., & Jurman, G. (2024). Generating and evaluating synthetic data in digital pathology through diffusion models. *Scientific Reports*, *14*(1), 28435. https://doi.org/10.1038/s41598-024-79602-w

Rivenson, Y., Liu, T., Wei, Z., Zhang, Y., de Haan, K., & Ozcan, A. (2019). PhaseStain: the digital staining of label-free quantitative phase microscopy images using deep learning. *Light: Science & Applications*, *8*(1), 23. https://doi.org/10.1038/s41377-019-0129-y

Sagheer, S. V. M., Nimitha, U., Ameer, P. M., Parayangat, M., Abbas, M., & Arunachalam, K. P. (2026). A Survey of Generative Adversarial Networks for Medical Images. *Computer Modeling in Engineering & Sciences*, *146*(2), 1–10. https://doi.org/10.32604/cmes.2025.067108

Saha, M., Mukherjee, R., & Chakraborty, C. (2016). Computer-aided diagnosis of breast cancer using cytological images: A systematic review. *Tissue and Cell*, *48*(5), 461–474. https://doi.org/10.1016/j.tice.2016.07.006

Salido, J., Vallez, N., González-López, L., Deniz, O., & Bueno, G. (2023). Comparison of deep learning models for digital H&E staining from unpaired label-free multispectral microscopy images. *Computer Methods and Programs in Biomedicine*, *235*, 107528. https://doi.org/10.1016/j.cmpb.2023.107528

Seoni, S., Shahini, A., Meiburger, K. M., Marzola, F., Rotunno, G., Acharya, U. R., Molinari, F., & Salvi, M. (2024). All you need is data preparation: A systematic review of image harmonization techniques in Multi-center/device studies for medical support systems. *Computer Methods and Programs in Biomedicine*, *250*, 108200. https://doi.org/10.1016/j.cmpb.2024.108200

Setiawan, A. W., Mengko, T. R., Santoso, O. S., & Suksmono, A. B. (2013). Color retinal image enhancement using CLAHE. *International Conference on ICT for Smart Society*, 1–3. https://doi.org/10.1109/ICTSS.2013.6588092

Shin, K., Lee, J. S., Lee, J. Y., Lee, H., Kim, J., Byeon, J.-S., Jung, H.-Y., Kim, D. H., & Kim, N. (2023). An Image Turing Test on Realistic Gastroscopy Images Generated by Using the Progressive Growing of Generative Adversarial Networks. *Journal of Digital Imaging*, *36*(4), 1760–1769. https://doi.org/10.1007/s10278-023-00803-2

Teramoto, A., Yamada, A., Tsukamoto, T., Kiriyama, Y., Sakurai, E., Shiogama, K., Michiba, A., Imaizumi, K., Saito, K., & Fujita, H. (2021). Mutual stain conversion between Giemsa and Papanicolaou in cytological images using cycle generative adversarial network. *Heliyon*, *7*(2), e06331. https://doi.org/10.1016/j.heliyon.2021.e06331

Verma, G., & Barthwal, A. (2025). Attention-enhanced deep learning for cervical cytology: combining convolutional networks with multi-head attention and fuzzy logic. *Polish Journal of Radiology*, *90*, 414–430. https://doi.org/10.5114/pjr/207475

Victória Matias, A., Atkinson Amorim, J. G., Buschetto Macarini, L. A., Cerentini, A., Casimiro Onofre, A. S., De Miranda Onofre, F. B., Daltoé, F. P., Stemmer, M. R., & von Wangenheim, A. (2021). What is the state of the art of computer vision-assisted cytology? A Systematic Literature Review. *Computerized Medical Imaging and Graphics*, *91*, 101934. https://doi.org/10.1016/j.compmedimag.2021.101934

Wang, R., Yang, S., Li, Q., & Zhong, D. (2024). CytoGAN: Unpaired staining transfer by structure preservation for cytopathology image analysis. *Computers in Biology and Medicine*, *180*, 108942. https://doi.org/10.1016/j.compbiomed.2024.108942

WHO. (2025, December 2). *Cervical cancer*. Https://Www.Who.Int/News-Room/Fact-Sheets/Detail/Cervical-Cancer.

Wilbur, D. C. (2011). Digital Cytology: Current State of the Art and Prospects for the Future. *Acta Cytologica*, *55*(3), 227–238. https://doi.org/10.1159/000324734

Yang, X., Bai, B., Zhang, Y., Li, Y., de Haan, K., Liu, T., & Ozcan, A. (2022). Virtual Stain Transfer in Histology via Cascaded Deep Neural Networks. *ACS Photonics*, *9*(9), 3134–3143. https://doi.org/10.1021/acsphotonics.2c00932

Yilmaz, B., & Korn, R. (2024). A Comprehensive guide to Generative Adversarial Networks (GANs) and application to individual electricity demand. *Expert Systems with Applications*, *250*, 123851. https://doi.org/10.1016/j.eswa.2024.123851

Zhou Wang, Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, *13*(4), 600–612. https://doi.org/10.1109/TIP.2003.819861

Zhu J, P. T. I. P. E. A. (2020). Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. *ArXiv*.

Appendix A

Procedure for Pap stain

Step	Stain/Reagent	Time/Procedure
1	Tap water	1 minute
2	Haematoxylin	45 seconds
3	Tap water	3 minutes
4	Bluing	1 minute
5	Tap water	15 seconds
6	Alcohol 96%	45 seconds
7	Orange G	2 minutes
8	Alcohol 96%	30 seconds
9	EA 50	4 minutes
10	Alcohol 96%	45 seconds
11	Alcohol 100%	45 seconds
12	Alcohol 100%	1 minute
13	Xilol	2 minutes