TRANSLATING AI TO DIGITAL PATHOLOGY WORKFLOW: DEALING WITH SCARCE DATA AND HIGH VARIATION BY MINIMISING COMPLEXITIES IN DATA AND MODELS

Govini Tantrige Upeka Vianthi Somaratne

Submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy of Information Technology

Murdoch University

July 2021

Abstract

The recent conversion to digital pathology using Whole Slide Images (WSIs) from conventional pathology opened the doors for Artificial Intelligence (AI) in pathology workflow. The recent interests in machine learning and deep learning have gained a high interest in medical image processing. However, WSIs differ from generic medical images. WSIs are complex images which can reveal various information to support different diagnosis varying from cancer to unknown underlying conditions which were not discovered in other medical investigations. These investigations require expert knowledge spending a long time for investigations, applying different stains to the WSIs, and comparing the WSIs. Differences in WSI differentiate general machine learning methods that are applied for medical image processing. Co-analysing multi-stained WSIs, high variation of the WSIs from different sites, and lack of labelled data are the main key interest areas that directly influence in developing machine learning models that support pathologists in their investigations. However, most of the state-of-the-art machine learning approaches cannot be applied in the general clinical workflow without using high compute power, expert knowledge, and time. Therefore, this thesis explores avenues to translate the highly computational and time intensive model to a clinical workflow. Co-analysing multi-stained WSIs require registering differently stained WSI together. In order to get a high precision in the registration exploring non-rigid and rigid transformation is required. The non-rigid transformation requires complex deep learning approaches. Using super-convergence on a small Convolutional Neural Network model it is possible to achieve high precision compared to larger auto-encoders and other state-of-the-art models. High variation of the WSIs from different sites heavily effect machine learning models in their predictions. The thesis presents an approach of using a pre-trained model by using only a small number of samples from the new site. Therefore, re-training larger deep learning models are not required which saves expert time for re-labelling and computational power. Finally, lack of labelled data is one of the main issues in training any supervised machine learning or deep learning model. Using a Generative Adversarial Networks (GAN) is an approach which can be easily implemented to avoid this issue. However, GANs are time and computationally expensive. These are not applicable in a general

translating ai to digital pathology workflow: dealing with scarce data and high variation by minimising complexities in data and models

i

clinical workflow. Therefore, this thesis presents an approach using a simpler GAN that can generate accurate sample labelled data. The synthetic data are used to train classifier and the thesis demonstrates that the predictive model can generate higher accuracy in the test environment. This thesis demonstrates that machine learning and deep learning models can be applied to a clinical workflow, without exploiting expert time and high computing power.

Table of Contents

List of Figures

translating ai to digital pathology workflow: dealing with scarce data and high variation by minimising complexities in data and models

v

List of Tables

translating ai to digital pathology workflow: dealing with scarce data and high variation by minimising complexities in data and models

vii

List of Abbreviations

AI - Artificial Intelligence
ANHIR - Automatic Non-rigid Histological Image Registration
AUC - Area Under Curve
CAD - Computer-Aided Design
CAE - Convolutional Stacked Auto-Encoder
CLL - Chronic Lymphocytic Leukemia
CNN - Convolutional Neural Network
COAD - COlon ADenocarcinoma
CT - Computed Tomography Scan
DCGAN – Deep Convolutional Generative Adversarial Network
DL - Deep Learning
ER - Estrogen Receptor
FC – Fully-Connected
FL - Follicular Lymphoma
GAN - Generative Adversarial Network
GPU - Graphics Processing Unit
H&E - Hematoxylin and Eosin
IDC - Invasive Ductal Carcinoma
IHC - Immunohistochemistry
MCL - Mantle Cell Lymphoma
ML - Machine Learning
MOPS - Multi-Scale Oriented Patch
MRI - Magnetic Resonance Imaging
NIA - National Institute in Aging
PR - Progesterone Receptor
ROI - Region of Interest
SIFT - Scale Invariant Feature Transform
SVM - Support vector machines
t-SNE - t-Distributed Stochastic Neighbour Embedding
TL - Transfer Learning
TRE- Target Registration Error
WA - Western Australia
WHO - World Health Organisation
WSI - Whole Slide Image

Declaration

I declare that this thesis is my own account of my research and contains as its main content work which has not previously been submitted for a degree at any tertiary education institute.

G.T.U.V. Somaratne

14/07/2021

translating ai to digital pathology workflow: dealing with scarce data and high variation by minimising complexities in data and models

ix

Acknowledgements

For paving the way to complete my thesis, I would like to pay my gratitude to Associate Professor Kevin Wong, my principal supervisor. His guidance and moral support help me see beyond my horizons. I would like to equally pay my gratitude to my co-supervisors, Associate Professor Hamid Laga and Dr. Alex Wang, for their time and effort in helping me reach the pinnacle. I would like to thank Jeremy Parry, Consultant Pathologist, Western Diagnostic Pathology for giving me deep insights into medical pathology. I am also grateful to Associate Professor Ferdous Sohel, who was the Advisory Chair, for his guidance and support.

Thank you, Murdoch University for the International Tuition Fee Scholarship to complete my PhD studies. Thank you, Lakmal Engineering, Colombo, for the scholarship which supported my candidature at Murdoch University. I would also like to thank the Discipline of Information Technology, Mathematics and Statistics and the graduate research office for their kind support during my studies. My dear colleagues at Murdoch University, thank you for creating an inspiring and motivating environment to work.

My dear family, thank you for supporting me during my PhD. I am grateful to my parents, my sister, and my in-laws, who have been always supportive to me. Their encouragement and advice shaped my growth mindset from the very beginning of my PhD journey.

Finally, I thank my husband, the guardian of my life supporting me to be the best of myself and thank you my little angel daughter, for your love, consideration, and positive attitude. I dedicate my work to my husband and my little daughter.

List of Publications Related to the Thesis

C1	Somaratne, Upeka V., Kok Wai Wong, Jeremy Parry, Ferdous Sohel, Xuequn Wang, and Hamid Laga. "Improving follicular lymphoma identification using the class of interest for transfer learning." In 2019 Digital Image Computing: Techniques and Applications (DICTA), pp. 1-7. IEEE, 2019.
J1	Somaratne, Upeka V., Jeremy Parry, Kok Wai Wong, and Hamid Laga. " Super-Convergence for Non-Rigid Registration using Deep Convolutional Neural Networks for WSIs." Journal of Pathology Informatics (Under Preparation)
J2	Somaratne, Upeka V., Kok Wai Wong, Jeremy Parry, and Hamid Laga. "The Use of Generative Adversarial Networks with Multi-site data in One-Class Follicular Lymphoma Classification." Neural Computing and Applications. (Under Review, 2022)

translating ai to digital pathology workflow: dealing with scarce data and high variation by minimising complexities in data and models

xi

Summary of Contributions of the Thesis

Chapter	Contribution	Publications
Chapter 2 Background	This Chapter presents a comprehensive literature review regarding factors translating AI in digital pathology, with the main factors being addressed in Chapters 3 to 5.	C1, J1, J2
Chapter 3 Super-Convergence for Non-Rigid Registration using Deep Convolutional Neural Networks for WSIs	Proposed a simplified approach for non-rigid registration of multi-stained WSIs.	J1
Chapter 4 Improving follicular lymphoma identification using the class of interest for transfer learning	Proposed an approach to handle inter-site differences in WSIs using small dataset from new sites.	C1
Chapter 5 The Use of Generative Adversarial Networks with Multi-site data in One-Class Follicular Lymphoma Classification	Proposed an approach to handle lack of data from new sites to handle inter-site differences in WSIs using synthetic data	J2

Chapter 1: **Introduction**

Digital pathology has increased the opportunities improve the pathology workflow. Before the introduction of digital pathology, the pathology workflow, consisted of specimen retrieval, tissue preparation on the glass slide, courier the glass slides, manual slides analysis in the pathology laboratory and generating manual patient diagnosis report. Whereas the digital pathology workflow includes specimen retrieval, tissue preparation, creation of digital slides, computerised analysis of digital slides and generating automated patient diagnosis report [1, 2].

1.1   DIGITAL PATHOLOGY

Digital pathology is a growing field in the medical domain. The process of digitising the histopathology specimens using whole-slide scanners and analysing digital slides using computational approaches is known as digital pathology [3]. Slide digitisation and computational analysis have helped to improve the efficiency and effectiveness in the pathology workflows [3].

In digital pathology workflow, the glass slide is scanned creating a digitised version of the slide which can be viewed and analysed in any magnification using a computer. The digitised slides are known as Whole Slide Images (WSIs) which are high resolution images which capture the entire glass slide. Figure 1.1 shows the efficient process of creating WSIs compared to using glass slides and manual analysis. In traditional pathology workflow, glass slides are manually annotated and manually transported to analyse using a microscope with a camera by which the specimen can be viewed as a video via computer software. In digital pathology, the glass slides are scanned using whole slide image scanners to generate digitised version of the glass slide [4].

WSIs are Giga-pixel images that are file sizes up to 10+GB with multiple levels of zoom levels for a single slide as shown in Figure 1.2 [5]. The zoom levels can vary from 3 to 7 levels depending on the imaging method. Each zoom level in a WSI

provides a different level of information about the slide. The low zoom levels with resolutions around $1,000 \times 1,000$ pixels would provide information about regions which may need further examination, and higher zoom levels ranging from $35,000 \times 45,000$ pixels to $100,000 \times 100,000$ pixels at full resolution would provide more details of the regions with cells and spatial arrangement characteristics to support diagnosis.

In the conventional pathology workflow, the glass slide storage requires a standard storage facility but due to increased slide volume and storage, management of slides are challenged as the slide quality and chances of loss of slides are increased [6]. Hospitals in remote locations need to transport slides to locations where experts are available and thus is a complex process. On the other hand, performing manual glass slide sample analysis and generating diagnosis for a patient require multiple expert opinions. Multiple expert opinion is becoming a scarce resource, as its challenged by the increased volume potential new cases [7, 8].

Compared to glass slides, storage of WSIs can be performed by using disks or cloud storage or combination of storage options. The digitised slides can be shared easily, which can be viewed and analysed using visualisation software locally and remotely by different stakeholders including pathologists, patients, research institutes, etc. [2, 9]. Use of WSIs helps to minimise the analytical and diagnostic step problems as the pathologist could use the zoom levels of WSI to analyse the WSI in different levels and perform a proper diagnosis. Giovanni et al., (2021) presents the business and monetary benefits of digital pathology workflow [4]. Enabling automated image analysis techniques implemented in the digital pathology flow will bring benefits in the diagnosis, prognosis, risk stratification and therapeutic selection [4].

Overall, the digitisation of histopathology slides has resulted in improved pathology workflow by minimising risks of handling glass slides and errors in managing and storing slides, improving visualisation with zooming and annotation functionalities, increased cost effectiveness by eliminating steps and bringing benefits of globalisation by sharing digitised slides. Furthermore, advancements in AI have shown promising outcomes in medical image processing. AI applications for WSIs have the potential of better time efficiency with faster data processing, analysis, and diagnosis [10].
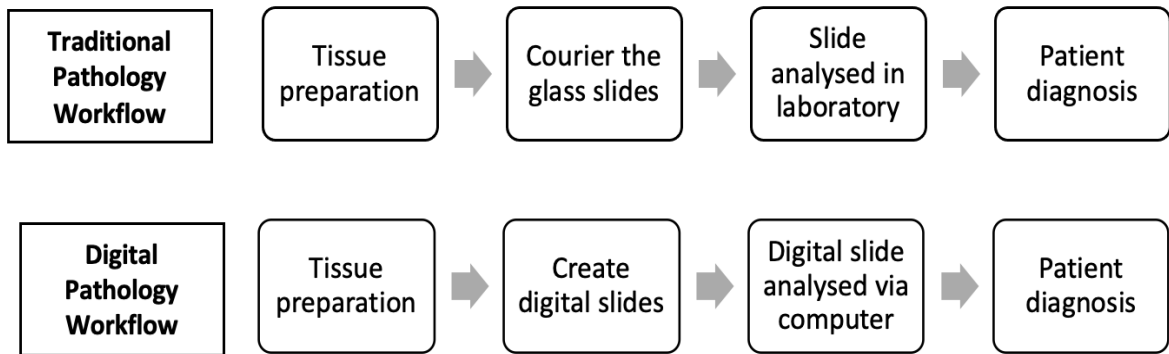
Figure 1.1 – A typical slide preparation process
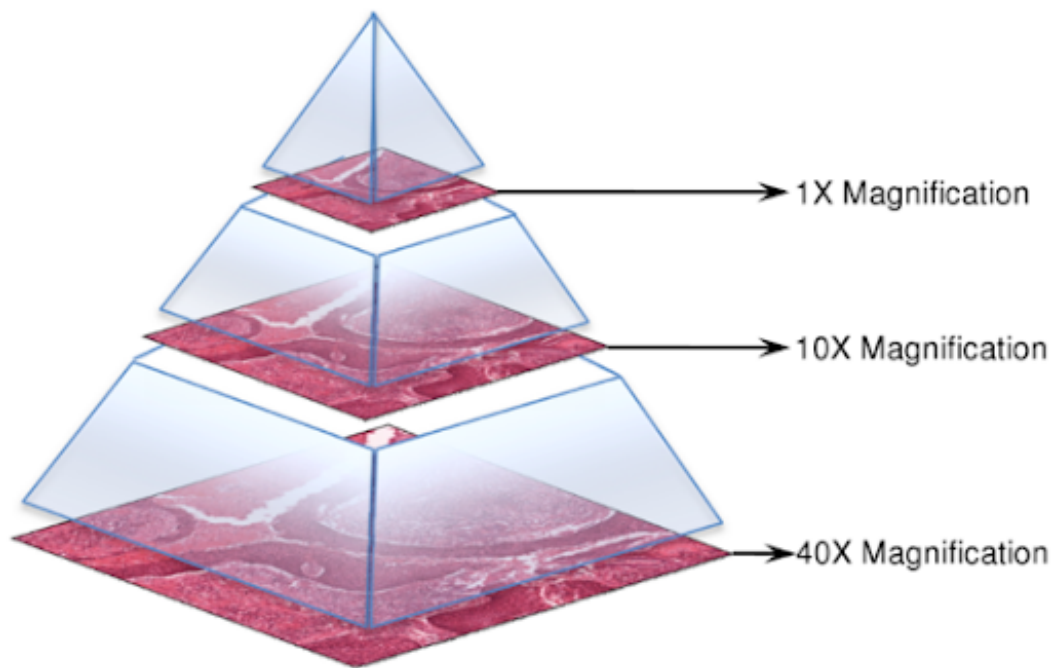
Figure 1.2 - An illustration of a whole slide Image at different levels ranging from resolutions at 1x, 10x, and 40x magnifications. [5]

1.2 ARTIFICIAL INTELLIGENCE FOR DIGITAL PATHOLOGY

AI is used in many domains and has shown high promise in accuracy and performance. Breakthroughs in AI have resulted in various products over the years. Drug discovery, stock market predictions, fraud detection, healthcare and education are among domains which have benefited from the advancements in AI. The ML and DL subcategories are the most popular approaches. ML focuses on learning patterns from historical data and features. DL is suitable for complex features with large amounts of high variation data. The capability of DL to learn hidden patterns of data is highly effective. Two main categories which ML and DL techniques can be categorised are, supervised and unsupervised. In supervised methods the model relies on the availability of known output, while unsupervised approaches are suitable when the output is not labelled. In medical domain both approaches are used. However, AI for cancer diagnosis in digital pathology was started to be explored in a slower pace compared to other medical imaging types such as, Computed Tomography Scan (CT) and Magnetic Resonance Imaging (MRI).

Assistance from automated methods for early cancer diagnosis can support to identify abnormal cells faster and accurately to handle the high number of new cancer cases reported. The computational techniques to analyse the WSIs and generate diagnosis have evolved rapidly over the recent years. Early computer aided design (CAD) systems to automate pathology workflow were challenged due to the unavailability of digitised data and later with insufficient computational resources [2, 7, 11, 12].

Recent availability of WSI and improved computational hardware and computational algorithms have brought histopathology forward despite the past lag compared to radiology and cardiology image analysis [13]. Though the early methods were not successful, the improvements to histopathology specimen scanning, computational hardware and machine learning (ML) which is a sub-category of AI has shown the potential of clinically applicable computerised methods [14]. ML requires training data for a given task in order to learn patterns in data. The training data supports the ML models to adjust to the training data. This allows ML models to learn the variations in the data and to adjust to a given task. ML models learn by themselves without any

specific manual tailoring to the training data. ML approaches have shown promising results for simple data [7, 15].

However, ML is affected by barriers in improving achievements in medical images compared to ML performance for natural images [13]. Complex data require higher processing and learning of data. WSI data are high resolution images and therefore consist of complex hidden features. Simplicity of ML methods is an advantage but the inability to identify and use large number of features including hidden features is a disadvantage. The limitation can be commonly identified as disadvantage of dimensionality due to the complexities in developing algorithms for high dimensionality. Support vector machines (SVM), and Random Forrest are ML methods that depends on feature selection and feature extraction techniques. Xing, et al., (2016) discuss the differences of SVM and Random Forrest in relying on hand-crafted features compared to Convolutional Neural Networks (CNNs). CNN automatically learn hierarchies of hidden features [10]. This need for pre-defined features leads to require more complex approaches to process the hidden patterns in high resolution images where a CNN can outperform by automated feature learning.

DL has progressed in the recent years to develop practical applications to various areas of research [16]. The further improvements in computational power, availability of larger amounts of data and the methods with capabilities of automatically identifying unseen features to learn patterns in data resulted in impressive progress in developing applications [9]. Among the various research fields, the medical field has benefitted from the DL techniques [17]. The improvements in medical imaging devices resulted in availability of data in various formats supporting automated diagnosis process [7].

The digital pathology diagnosis tasks can be identified as classification, segmentation, registration, detection and grading [7, 14, 18]. A DL classification method proposed by Miyoshi et al., (2020) differentiates WSI patches from different magnifications. A comparison of the DL model performance with pathologists' performances shows the potential in DL supporting malignant lymphoma diagnosis [19]. Prior research shows progress in segmenting WSIs to identify locations of interested regions. The authors review segmentation methods for colorectal cancer using an improved U-shape network with VGGNet as the backbone [20]. U-shape networks consist of an encoder

and a decoder and VGGNet is a pre-trained CNN using a dataset that consist of a large number of natural images. WSI image registration to align multi stained WSIs with DL features for rigid and non-rigid registration has shown progress in a multistep approach [21]. Furthermore, detection and grading methods have shown promising results. Yusuf et al., (2020) showed promising performance in detecting cells using pre-trained models for invasive ductal carcinoma (IDC) [22]. The mitotic count of invasive breast cancer is required for histological grading. A study which developed a DL based technique for mitotic counting concluded that CNN based counting is comparable to observer's manually calculated results and that it is a reliable method to assist in grading [23].

However, despite the numerous successful performances, research shows that there is still room to improve in order to apply in clinical flow [24]. The specific process of preparing WSIs, complex process of diagnosis, high variation in patient samples, and feature rich high dimensional WSIs leads to gaps in translating AI to digital pathology workflow. Identifying the gaps in AI research translation to digital pathology is crucial for future improvements of the workflow for an efficient and accurate diagnosis outcome [25].

1.3 CHALLENGES IN TRANSLATING AI TO DIGITAL PATHOLOGY

WSI processing is a challenging and unique task in image processing, because of the variation in each imaging type (size, colour, hue etc.). This depends on the tissue sample preparation standards, imaging modality, image acquisition methods [26]. This limits the practical aspect of using one method that fits all for WSI processing. Compared to radiology images, WSIs are different and complex resulting barriers in translating analysis methods to WSI from other medical image types [27]. There are numerous methods with high variation that are applied for digital pathology. However, each method has own unique capabilities and flaws. Therefore, identifying the challenges which creates a gap between translating AI methods and digital pathology workflow is crucial. Identifying the challenges and factors will help to ensure a smoother translation of AI models to the digital pathology workflow.

DL methods have shown capability of improving the results and efficiency for digital pathology workflow [8]. However, prior literature identifies pre-analytical factors, analytical factors, and post - analytical factors as main categories of barriers translating AI methods to digital pathology. David F. Steiner et al., (2020) reviews challenges in detail and potential improvements to the digital pathology workflow to handle the shortage of trained pathologists and handle increasing number of new cases, etc. [25]. Hamid Reza et al., (2018) discuss challenges and opportunities in AI for digital pathology [24]. Challenges which limit the employment of AI methods to the digital pathology flow are identified as lack of labelled data, pervasive variability, non-boolean nature of diagnostic tasks, dimensionality obstacles, Turing test dilemma, uni-task orientation, computational expenses, adversarial attacks, lack of transparency and interpretability, and realism of AI.

Opportunities to mitigate the challenges have been discussed in prior literature. Many of the challenges results in lack of annotated data where techniques which use least amount of annotated data, or no annotations are an opportunity. Attempts to minimise the gaps in translating AI to digital pathology have been developed. A DL based confidence scoring method to handle high inter-pathologist discordance have been developed using multi-site data which is one such attempt [28]. Methods which use one class data, which are weakly labelled data with labels available only for one of the classes and other techniques using partially labelled data are beneficial as well as methods which use the unlabelled data. However, further research and development to align AI and digital pathology needs to be conducted based on impact factors identified. There are a few main obstacles which multiple prior literature has identified as barriers of implementation of AI methods in digital pathology workflow.

The WSIs are Giga-pixel images with very high zoom levels and resolutions with $1,000 \times 1,000$ pixels to $100,000 \times 100,000$ pixels. Figure 1.2 shows an example of a WSI in the low resolution and a section in high resolution. Annotating and analysing WSI is a complex task which requires expert knowledge and time [15, 29, 30]. Therefore, it requires a long time and is very costly. Along the annotation difficulties, the uni-task oriented solutions which are trained only to perform one task and high variability in data due to differences in slide preparation process are commonly identified pre-analytical factors. Capturing the variability in data is important to train

models, but such variation is not readily available in data due to annotation complexities. Pathologists co-analyse WSI samples in multiple types of stains to identify different tissues and cells. The co-analysis adds more complexity to the annotating process. Variation in data from different locations which use different scanners and standards also result in a challenge to translate AI applications to the digital pathology workflow. Prior research shows the importance of having a large amount of labelled data which capture the high variation of WSIs to achieve promising outcome from trained models. The limitations in available training data increase the need of methods to utilise the limited labelled data to maximise by augmentation or models requiring minimal labelled data. Furthermore, the dimensionality obstacle of WSIs needs to be considered along with the affordability of computational expenses. Therefore, these factors affect the performance of the models threatening the realism of AI applications to the digital pathology workflow. The above challenges lead to the problems focused on this thesis.

1.3.1 Problem Statements

1. **Co-Analysing multi-stained data**

   Tissue structures are complex and consist of several types of tissue including nuclei, small holes, vacuoles, cells and different spatial arrangements. Pathologists use biomarkers which are known as stains to help differentiate tissue types and spatial arrangements [31]. The analysis requires co-image analysis for the same tissue with many stain types which result in many WSIs for the same sample. Co-analysis requires WSI registration in order to identify corresponding tissue and cells identified in multi-stained WSIs [21]. Figure 1.3 shows WSIs stained using different types of stains to identify different cells. Expert pathologists use different stains to identify regions of interest on the WSIs, however, the process is time consuming. Therefore, registering the WSIs using different stains is important to support pathologists [32]. The main challenge is when the WSIs are re-stained and scanned which result in movement in tissue sample and changes in placing it compared to the original WSI resulting rigid and non-rigid deformations [33]. The rigid transformations do not change the shape or size of the content of the image, but non-rigid transformations are
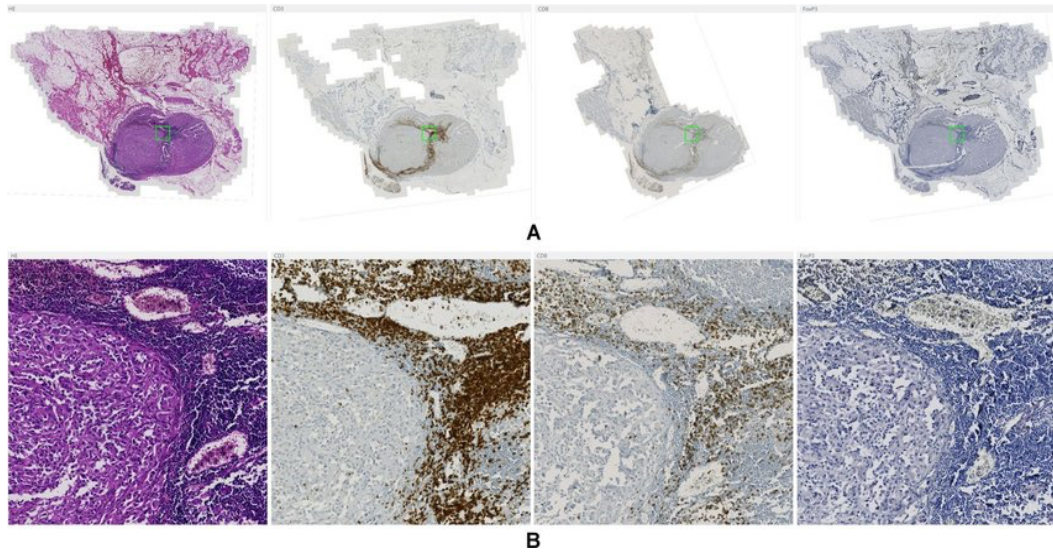
Figure 1.3 - Four different stain types applied to the same WSI at low resolution (A)
and to the same WSI patch at high resolution (B)
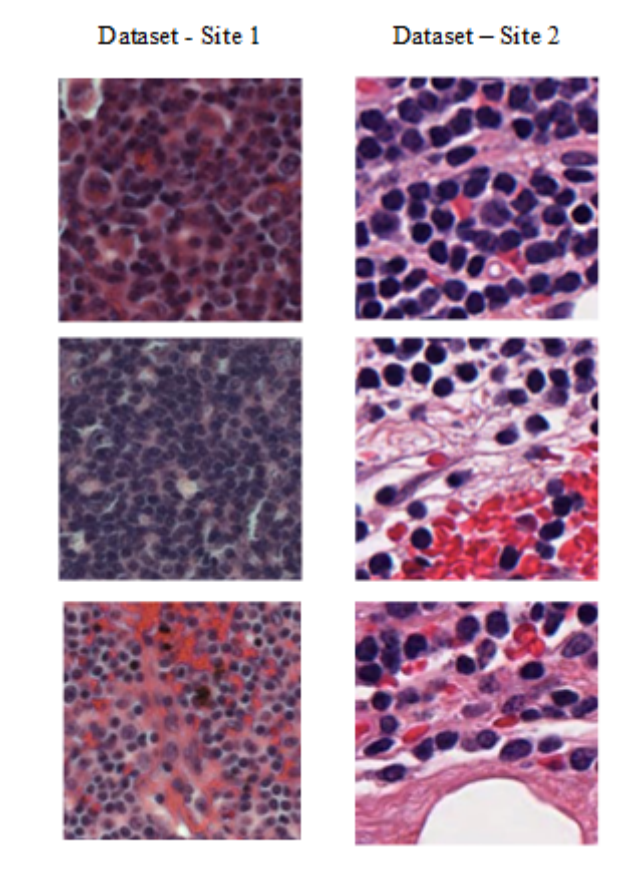
Figure 1.4 - Visual difference of WSI image patches for follicular lymphoma from
different sites. Dataset – Site 1 is publicly available data and Dataset – Site 2 is from a
private dataset.

complex due to changes in the structure of the content in the image. Furthermore, the registration step is only one sub-step towards diagnosis and should have a dynamic approach addressing non-rigid transformations which is not time consuming and requiring very high computational memory. The efficiency of registration as a sub-step is important to be translated to the digital pathology workflow. Requiring high computational power and longer processing times challenges the resources of already challenged health sector [32].

2. High variance in data due to inter-site differences

The pervasive variability complicates the labelling process due to the several tissue types present in a sample [15, 24]. Due to different procedures and standards, the WSIs from different institutions increases the variations which effects the generalisation of DL models. The slides are stained for visibility of different features of cells. The stain preparation, and the standards from one laboratory to another can be different leading to added variation in WSIs [26]. Therefore, as Figure 1.4 shows, the WSIs from different sites show different data distributions [27, 28, 34]. This causes WSI processing to perform differently though the same stain is applied to the same type of tissue. Therefore, a model trained on different laboratories may not perform similar to trained laboratory WSIs [35]. The WSIs should be trained for different laboratories. The uni-task orientation where methods developed to one site/institution is not directly transferable to another institution with data from different modalities or different equipment's [28]. The large variations result in poor performances in the same methods applied to data from different institutions leading to a gap in translating the methods. Requiring annotations from various institutions for the same tasks are not feasible as the complex annotation process needs to be followed repeatedly.

3. Lack of annotated data

The requirement of image-level, patch-level, and pixel-wise annotations to successfully train and evaluate DL models is a difficult task. Lack of labelled data is one of the main challenges identified by many researchers [30]. The need of expert annotated data is limited by time constraint and financial bottleneck for app

development [24]. Crowdsourcing, active learning and generating synthetic data are few options that's considered to increase labelled data. Use of annotated WSIs from each institution to train generalised DL models is not feasible due to the large dimensions and training using patches or segmented regions of interests. Generating synthetic data using Generative Adversarial Networks (GAN) have shown promising results for natural images as well as for medical images. However, the high complex model architectures and long training times are challenging the capability of translating the methods to the digital pathology workflow. Methods which can use the limited annotated data without adding large costs to health systems are required [32]. Therefore, using GAN as an approach to generate synthetic data based on the available annotated data from a new site can reduce the translation gap in digital pathology.

1.4   OBJECTIVES

This section focuses on discussing the main factors affecting translation of AI methods in digital pathology. AI approaches that are developed for natural images and publicly available WSI datasets are not directly transferable. Therefore, enhancements are required to implement AI approaches for digital pathology workflow. Comparisons between models on limited publicly available data and real-world data are presented with the challenges in transferring knowledge between models. Solutions are proposed to overcome challenges identified to improve clinical digital pathology workflow by dealing with scarce data and high variation by minimising complexities in data and models.

Main Objective

The main objective of the thesis is to design and develop deep learning approaches to learn from limited real-world WSI data.
In order to achieve the main objective, following sub objectives are created.

Sub-Objectives

1. Improving multi-stained WSI data registration with fast converging deep learning model

Registration is an important sub-step towards analysing WSIs. WSIs can be stained with multiple stains. Various stains reveal various features in WSIs and provide different information and perspectives to pathologists. The combination of stains would give a further insight into WSIs and help pathologists in their diagnosis. Registration of WSIs places the same WSI with different stains to overlap each other and match the exact points to give a clear insight to the pathologists to co-analyse and provide diagnosis without manually matching WSIs. Due to the high non-rigid deformities in the WSI samples, use of DL to learn and extract hidden deep features for registration has shown progress. However, the high computational memory required to train the complex models is a challenge to adopt deep feature-based registration to digital pathology workflow. Therefore, methods which learn and converge faster by using minimal computational memory are important.

2. Addressing inter-site differences in WSIs using transfer learning

WSIs are giga-pixel images and labelling them are time consuming and requires a large amount of expert knowledge. Therefore, creating datasets for DL models for real-world WSI is a complex task. Pre-trained models are trained on public data and the weights of the models are saved. These models and saved weights are used for similar data which do not have large amounts of training data (Transfer Learning Models). Transfer learning (TL) is highly suitable for WSI processing because of the limited annotated WSI data. Furthermore, training models with data from multiple sites have shown to result in better model performances. Therefore, limitation of acquiring sufficient data from different sites can be handled by only using a smaller dataset from a new site.

3. Addressing lack of labelled data WSIs using synthetic data

Synthetic data is created to support DL models which have minimal number of training data. DL models require large amounts of training data. Therefore, when minimal

number of training data is available, creating synthetic data supports training machine learning models. Medical images with annotations are minimal because medical images and annotation are complex, time consuming and costly. Therefore, creating synthetic data is an efficient method to create images for machine learning models to train. Investigating simpler solution to be able to reduce the gap of translating AI to digital pathology is another aspect of this objective.

1.5 CONTRIBUTIONS

The contributions of this thesis are three-fold.

1. **To simplify and improve performance of unsupervised non-rigid registration by using deep learning features and faster convergence.**

Assisting to identify corresponding tissue types and spatial arrangements in the same tissue sample by using multiple stains is WSI registration. De-staining and re-staining, the tissue specimen changes by adding artifacts and loss of information. Therefore, identifying corresponding locations and conducting co-analysis for diagnosis is a complex task and time-consuming task for expert pathologists. Conventional registration methods require intense parameter tuning. Therefore, use of DL for automated feature extraction demonstrates the benefit allowing faster parameter tuning for registration. However, compared to supervised methods, unsupervised methods are becoming popular due to the limitations in large, labelled datasets for training. The available unsupervised methods are complex, requiring longer training times and using complex architectures with large number of deep features. Furthermore, the unsupervised approaches can be unpredictable. Therefore, a novel method combining supervised and unsupervised approaches which learns and converges faster using minimal amount of computational memory is developed. The method handles the unsupervised registration technique's complexity with a simpler model architecture and faster convergence.

2. **To improve transfer learning-based method by handling inter-site differences with minimal labelled data.**

This explores the possibility of using a small dataset from a class of interest to train deep learning algorithms for FL detection from WSI images. FL is a sub-type of lymphoma which grows silently, without symptoms and therefore harder to identify early. The proposed method reduces the need for a large amount of labelled data by training a classifier for FL in a target environment. In this approach, a pre-trained AlexNet model fine-tuned using a publicly available Lymphoma dataset. The proposed approach combines the Lymphoma dataset to a smaller sample of a private dataset of the FL class obtained from a new site. This approach requires only a small amount of labelled training data from the class of interest to improve the performance on data from a new site. The proposed solution reduces the time and effort incurred to train new models to new data, which helps to minimise translation gap of AI to digital pathology workflow.

3. **To improve classification performance by using synthetic data to handle lack of labelled WSI data for a new site.**

Published research information shows that the lack of labelled data and the problem of generalisation to data from different sites affect the performance of DL models. After studying the availability of a small amount of annotated data for the class of interest from different sites, the chapter focuses on improving generalisation of models to new sites using the one-class data from new sites. The capability of GANs to create synthetic data has been a promising approach to handle limitations in data. Therefore, this contributes and investigates the GANs influence in the one class classification tasks by creating synthetic data for the labelled one class data in the private dataset. In order to classify one-class private data, the public dataset's, negative class is passed down as non-target data for classification. The use of limited one class data from new sites with GAN significantly contributes to reducing the differences in data distributions of different sites which caused a generalisation problem. The technique showed possibility of generalisation by using one-class data without needing to retrain models for the new site, thus contributing to minimise challenges in translate AI to digital pathology workflow.

1.6 THESIS OUTLINE

Chapter 1 provides a description of the recent advancements and challenges in deep learning and digital pathology. The motivation for the proposed approaches to address the problems in applying Deep Learning to WSI are described. The chapter also highlights the contributions in the thesis for DL applications to WSIs.

Chapter 2 presents the prior literature to identify factors translating AI to the digital pathology workflow. Furthermore, a problem-based taxonomy and provides the main types of solutions in the research field. The chapter concludes by providing the gaps in WSI classification.

Chapter 3 contributes to WSI registration with faster convergence to unsupervised non-rigid registration.

Chapter 4 contributes to explore the use of transfer learning to handle differences of data from different sites.

Chapter 5 contributes to study the effects of synthetic data for a new site and handle the challenge of limited number of labelled data.

Chapter 6 summarises the thesis contribution and reflects the opportunities for future work based on the research findings presented here.

Chapter 2:   **Literature Review**

This chapter reviews methods addressing factors of translating AI for digital pathology. Section 2.1 provides a discussion on digital pathology workflow followed by AI applications to WSIs. Section 2.3 presents a taxonomy of factors translating AI in digital pathology and discusses the different methods developed to address the factors.

2.1   DIGITAL PATHOLOGY

Digital pathology is identified as the process of diagnosing using digitised glass specimens for examination and analysis to provide a final diagnosis for a patient. The digitised glass specimens are commonly known as WSIs.

Before introducing WSIs, a manual process of cancer diagnosis was performed for tissue samples prepared on glass slides and using a microscope to examine the tissue samples [12]. The tissue samples needed to be stained to assist the microscopic examination. The staining highlights different tissues and cells with high contrast to help pathologists assess the tissue samples [13]. The process from preparing tissue samples to examination to identify and analyse the cell structures and its features are cumbersome. The microscopic examination requires multiple pathologists' assessment, and the outcome of the examination is subjective, leading to inter and intra observer variability [13, 36]. Furthermore, the physical handling, storage and transportation of the glass slides needs cautious attention [37].

The availability of digitised tissue samples helped to ease the process of tissue sample examination and analysis. The pathologists can examine the digitised samples known as WSIs at different magnification levels, share the WSIs with experts in the field using computer-based systems, which allows viewing WSIs locally or remotely [14, 37]. However, the slide preparation for WSI acquisition is still complicated compared to the radiology image acquisition process [30]. It is a multi-step process that consists of fixation, dehydration, clearing, infiltration, embedding, sectioning, and staining. Figure 2.1 shows the cancer diagnosis process; fixation is required to preserve the tissue in the same state as when they were alive to harden the sample in order to retain

the molecular structure [5]. After fixation, the sample is processed by dehydration, clearing, infiltration and embedding. The sample is sectioned prior to staining, which creates the glass slides. Once the glass slide is prepared, image acquisition can be performed. The process of preparing tissue specimens for acquisition may take 12 hours to 24 hours. As discussed, the complexities of the essential slide preparation process and the growing number of new cases increase the need for automated methods to speed up the whole procedure.
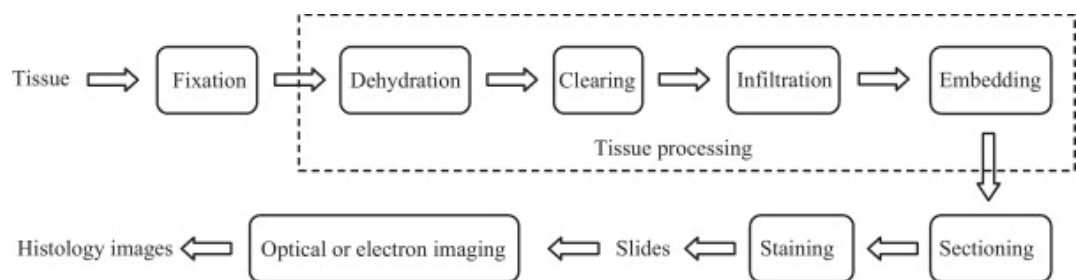
Figure 2.1 - Full diagnosis process of digital pathology imaging in detail

2.2 AI FOR DIGITAL PATHOLOGY

The introduction of WSIs has bridged digital pathology and machine learning (ML) methods to handle different tasks, including classification, segmentation and registration, in order to assist the diagnosis process [7, 17, 18].

2.2.1 Early Medical Image Processing methods

Medical imaging is essential for medical image diagnosis. The complexities in the process from image acquisition to diagnosis show the need for Computer-Aided Design (CAD) systems to assist with image analysis for a faster diagnosis [28]. The use of statistical methods and rule-based methods have aided in early medical image analysis. Various imaging types are available in the medical domain, which includes radiology, histology etc. These image types significantly differ from each other due to its technicalities and capturing angles. Leading to require different image processing techniques [5, 24]. However, due to the computational limitations to handle complex image types and scarcity of digital images, some medical image types such as WSIs challenged to the application of CAD systems. With improvements in technology, advanced medical imaging techniques and computational hardware were built to gain

the capability of processing high dimensional imaging types [14]. The computational methods for WSI diagnosis assistance have to include pre-processing, segmentation, feature extraction, dimension reduction, and detection and classification. The outcome of the detection or classification needs further post-processing and assessment in order to derive the final disease diagnosis and grading results (Figure 2.2) [5]. ML techniques, specifically DL techniques have proven to show high performances in various fields of research [5, 14, 24, 28, 29, 32, 36].
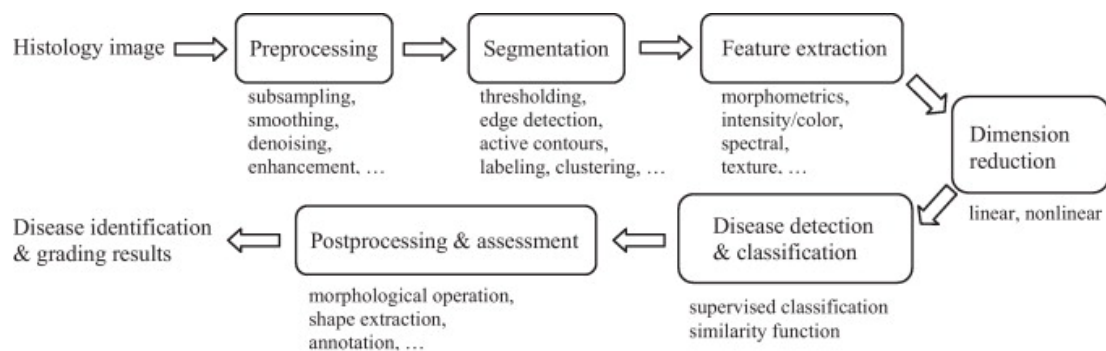
Figure 2.2 - The full process for computational analysis of WSIs

2.2.2 Machine Learning Approaches
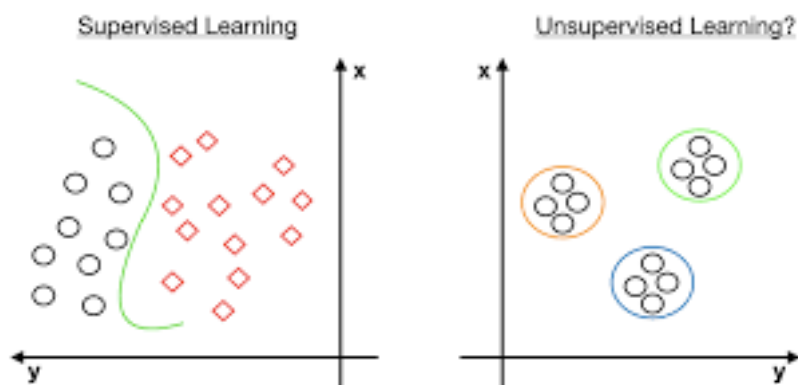
Figure 2.3 - The data distributions for supervised learning and unsupervised learning

Machine learning (ML) provides the capability of learning patterns in data to support a given task. ML approaches have been popular due to the capability of identifying patterns and relationships in complex data based on historical data [13, 15]. Altaf et al., (2019) shows ML for medical image analysis applied under two broad areas of unsupervised and supervised learning [30]:

1. Unsupervised Learning draws inferences from data without any labels.
2. Supervised Learning draws inferences from labelled training data to achieve a regression or classification.

In this section, unsupervised and supervised learning will be further discussed (Figure 2.3). There are different types of function approximators, linear models, decision trees, Gaussian processes, deep learning etc. Unsupervised learning uses unlabelled data to infer and identify patterns to achieve a task. The unsupervised algorithms learn without labels to find patterns in the data, which is fully unguided [15]. These approaches can produce features and clusters which are not visibly obvious, and the inferences can also introduce new knowledge about the data, especially in medical imaging. Hence concluding that, unsupervised algorithms are capable of revealing new knowledge and patterns which were not known before [24]. Surprising results were inferred from unsupervised algorithms in medical imaging. Comparatively there is limited control on the results of unsupervised learning algorithms to supervised learning algorithms.

Supervised learning is mapping and learning from sample inputs to outputs. It is important to have the sample input (x) to $x \in X$. Therefore, supervised learning algorithms match to the actual input (X) and learn to produce the expected output (Y). Supervised learning algorithms are catered into generating the output (Y). Medical imaging with its unique data can be highly influenced by supervised learning because the supervised learning algorithm can be learnt from the medical knowledge and guidance. However, supervised learning approaches require high dimensional feature engineering and the input of medical expertise. Therefore, deep learning, which does not require high dimensional feature engineering, has potential in medical image processing [5, 7, 15, 18, 30].
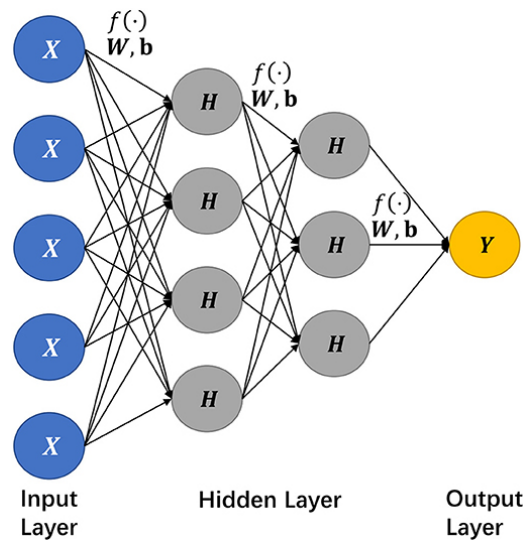
Figure 2.4 - Architecture of a deep neural network

2.2.3 Deep Learning approaches

During the last decade, deep learning has revolutionised image processing, natural language processing and time-series approaches [16, 38]. Deep neural networks can identify hidden patterns in data using its hidden neural network. The deep neural network has multiple layers, which extract various hidden features from data (Figure 2.4). This ability to extract features varies from one layer to another. The abstract features are extracted from the first layers, and the last layers extract more complex and extracts deep features. This ability in deep neural networks has gained momentum in research for image processing.

$$y = f(x; \theta) \qquad (1)$$
$$h = A(W_1 \cdot x + b_1) \qquad (2)$$

General functionality of deep neural network is expressed in Equation 1. The output of the deep neural network is represented by (y). θ represents the learnable weights. x represents the input to the deep learning model. f is the general function that can be the activation function.
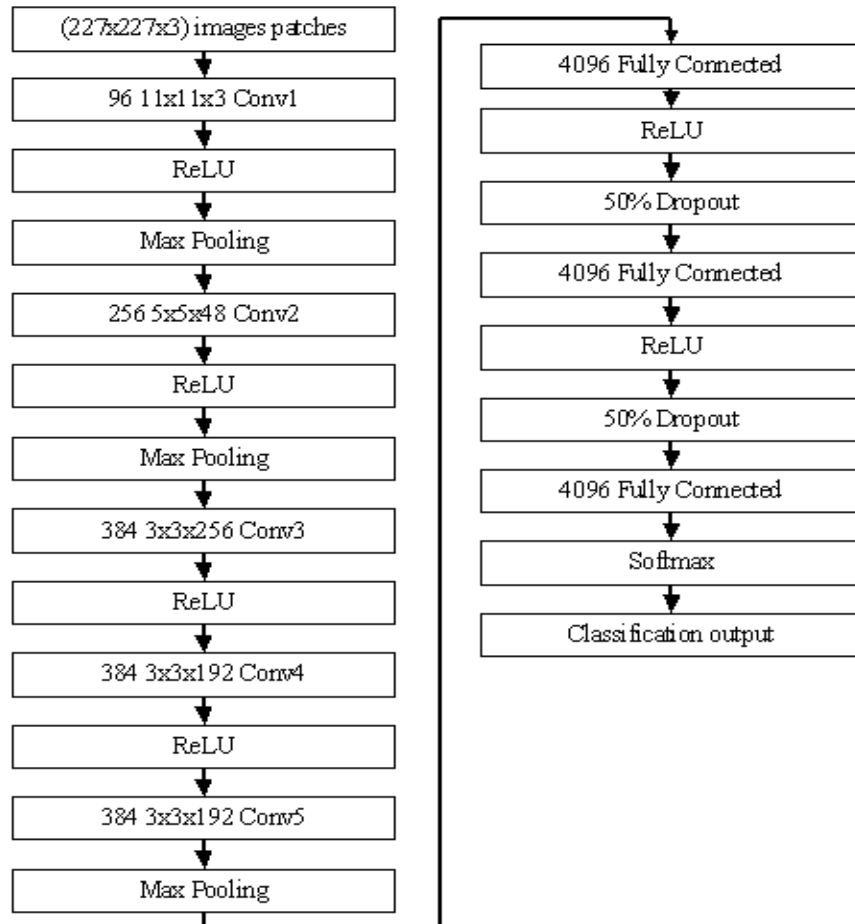
Figure 2.5 - Example architecture of a CNN –
Architecture of AlexNet model.

The deep neural networks have multiple succession processing layers, which comprises non-linear transformations that lead to learning different levels of abstractions. According to equation 2, the deep neural network result (h) is with an input column vector (x) and non-linear activation functionality (A) multiplied by (W_1) and a bias of (b_1).

The non-linear transformation of the deep neural networks can extract the abstract and deep features of the input vectors. Therefore, deep neural networks are capable of learning features from images and supporting image processing tasks. Especially convolutional layers are used for image processing due to the translation invariance property [39]. The Figure 2.5 shows the architecture of a CNN named AlexNet. The layers consist of learnable filters which activate when different features are detected.

The first layers learn to detect the edges, textures and patterns. The following layers detects parts of the objects as well as the whole object. This combination of learning in convolutional layers allows to further elaborate on the analysis of image data. Therefore, Convolutional Neural Networks (CNNs) are adapted heavily into image processing tasks. CNNs are used for medical image processing by utilising the above processing capabilities.

WSIs are large, complex images with complex features. Therefore, CNNs have the capabilities of processing multiple complex features in WSIs. CNNs are capable of extracting features from WSI for analysis [40-42]. Furthermore, WSIs hidden feature can be extracted using CNNs. Therefore, CNNs would have the capability to extract various features and be used for WSI analysis. The complex feature extraction in the convolutional layers has a high potential to support the WSI analysis. However, having multiple convolutional layers is computationally expensive. Furthermore, it's impossible to apply complex CNN models in every situation. Therefore, it is important to come up with the models which are computationally efficient to extract features and analysis of WSIs.

2.3   FACTORS OF TRANSLATING AI

Medical image classification is a popular research domain for applying and developing deep learning techniques. Among the various types of medical imaging, the histology images had a gap in developing machine learning-based techniques due to the unavailability of digitised version of the glass tissue slide samples. However, with the development of whole slide image scanners, the problem of unavailability of data was solved. The WSI domain gained popularity in deep learning-based techniques to assist the diagnosis process. Prior research shows promising results in deep learning-based techniques for WSI. The techniques are developed but not limited to classification, segmentation, object detection and image registration. However, translation of DL methods to digital pathology is challenged by various factors.

The present thesis focuses on factors of translating WSI registration and WSI classification in digital pathology. Registration is the alignment of multi-stained data to assist co-analysis which is a crucial sub-step in the diagnosis process. AI techniques

have benefitted WSI registration by better handling the high dimensionality, large deformations, and faster registration [21, 43, 44]. The registration methods address rigid transformations and non-rigid transformations. Rigid transformations do not change the size or the shape of the images. Rotation, translation, and reflections are considered as rigid transformations [43]. Non-rigid transformations show non-linear changes which occur in the slide preparation process [45]. Recent successful performances have shown that a combination of rigid and non-rigid transformations

Table 2.1 – Main challenges and potential solutions identified in prior research focusing translation of AI in digital pathology

Research Title	Challenges	Potential solutions
Deep learning in histopathology : The path to the clinic [46]	Datasets are not truly representative of clinical data	Use of multi-site data
	Low generalisability	Use of multi-site data still result in poor generalisability. Limited research available.
	Low robustness due to high variation in clinical data	Combination of data with different stains, sites and use of augmentation to artificially generate data. Both data augmentation and image normalisation are necessary
	High dimensionality	Use of GPU, compress WSIs
	Funding	Rural areas can benefit if access full digital pathology infrastructure is possible
	Model evaluation for regulatory approvals	Standards in model evaluation
	Explainability of algorithms	Standards are required due to debatable nature. Different stakeholders expect explanations in different angles.
Closing the translation gap: AI applications in	Pre analytical factors – complexities in glass slides and WSIs. Effects the generalisability of models.	Diversity and representation of clinical data to the model.

digital pathology [25]	Analytical factors – regulatory needs.	Clinical validation, quality control, interaction with pathologists
	Post analytical – Documentation, result verification and communication after initial development and validation.	Validation standards. Including single site to multi-site validation. Careful integration - despite the powerful and accurate models. Model should be able to merge with other systems in place. Collaboration of pathologists, developers, and researchers.
Translational AI and Deep Learning in Diagnostic Pathology [47]	Low ability to generalise - Extremely difficult challenge	Use of multiple variations of data for testing
	Testing with various data is costly and time consuming	
	Legal and regulatory aspects	
Artificial intelligence and digital pathology: Challenges and opportunities [24]	Lack of labelled data, variability, uni-task orientation	Pretraining models, handcraft features, generative models, unsupervised learning, use of features identified by stains
	High dimensionality	Use of image patches and regions of interest. Downsampling (Risks in loss of crucial information)
	Non-Boolean outputs – Sometimes pathologists use descriptive terminology to describe the outcome of the diagnosis.	
	Turing test dilemma –May not know the Turing test explicitly for pathology.	

	Affordability of computational expenses.	
Digital Pathology: The Time Is Now to Bridge the Gap between Medicine and Technological Singularity. [48]	Infrastructure and resource support	
	Integration to existing systems	Communication with pathologists during development process

have a drastic improvement in registration outcome [21, 49, 50]. Furthermore, limited research using unsupervised registration methods address the limitations in annotated data. In recent DL based methods, computational memory requirement is a limitation to translate to clinical workflow. Therefore, computationally efficient methods need further exploration to assist with co-analysing multi-stained data.

Classification assists in categorising the WSIs based on the class of interest. The limitations in labelled WSI data and gathering labelled data from different sites is the main challenge when it comes to classification. The WSI images with image level labels are insufficient due to the rich features present at high resolution of WSIs. Therefore, patch level, and pixel level labelling is essential for training and evaluation which challenge the availability of data. Furthermore, the high variation in data causes limitations in generalised classification methods. The variations in data can be caused by the differences in staining techniques, slide preparation procedures, and different scanning equipment available at each institution. The dataset requires to capture the data distribution to train models with higher generalisability. The collection of data from different institutions is a complicated task. Therefore, methods which use limited labelled data or methods which train in unsupervised manner have been investigated for in the latest research. Identification of such limitations are important to address the challenges in translating AI in digital pathology.

2.4 TAXONOMY FOR FACTORS TRANSLATING AI TO DIGITAL PATHOLOGY
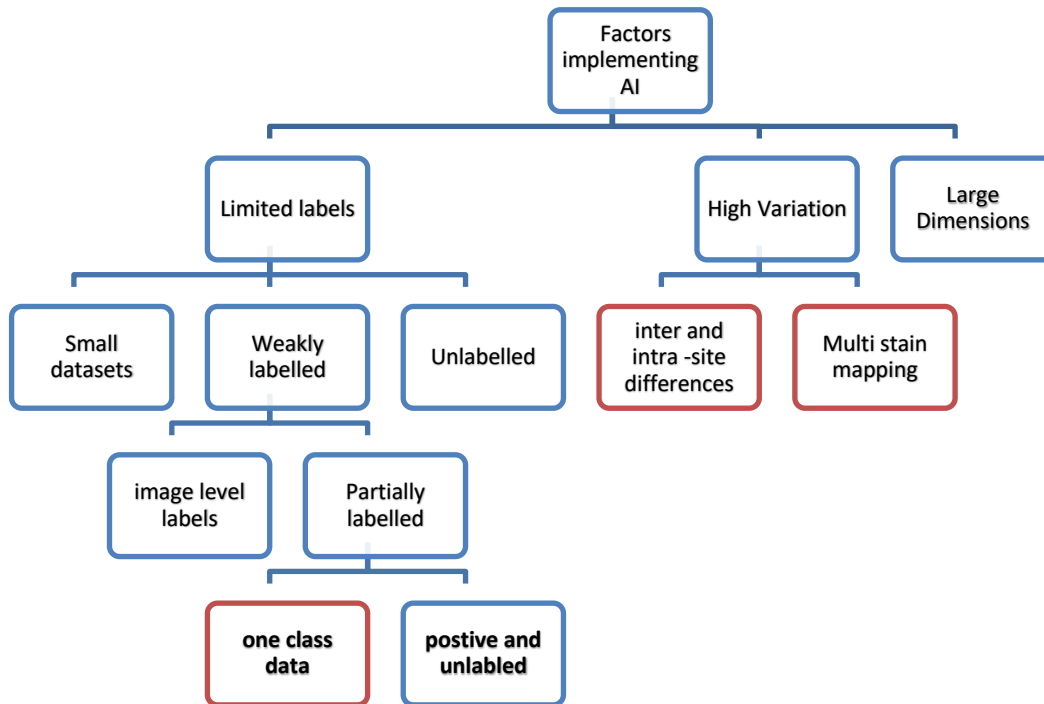
Figure 2.6 - Taxonomy for factors translating AI in digital pathology

Despite the recent success in deep learning for WSI, there are challenges that have been faced by researchers in translating methods to the digital pathology workflow. Prior literature analysis leads to DL technique-oriented solutions and WSI data-oriented solutions [14, 24]. The main challenges that have been identified are scarcity of labels, high variation in histology structures and the large dimensions of WSI images. The Table 2.1 identified main challenges and solutions discussed in prior research focused in exploring translation of AI in digital pathology. The taxonomy shown in Figure 2.6 is based on the main challenges summarised in Table 2.1 and approaches to handle and assist diagnosis. The sections highlighted in red are the main problems focused on this thesis.

The challenge of limited labels is further divided into small datasets, weakly labelled datasets and unlabelled dataset. This categorisation is based on the different availabilities of data. Further categorisations for the weakly labelled data are given as referring to numerous definitions of weakly labelled in the prior literature. We have identified two main categories in weakly labelled data as image-level labelled and

partially labelled data. The one-class problem is identified as one way of handling partially labelled data (highlighted in red in Figure 2.6).

The challenge of high variation in WSIs is caused by inter-site and intra-site differences and co-analysis of multi-stained WSI mapping (highlighted in red in Figure 2.6). The inter-site and intra-site differences have been handled mainly by fusing external data and/or using techniques to minimise differences in the data distributions. The differences in data distribution can be handled by using traditional methods or learning-based methods. The high variation of data due to multi staining of the same sample in different stains is carried out to highlight different cell structures which causes co-analysis of WSIs a challenging task. Registration methods specific to WSIs are required to map the corresponding features in the multiple stains (highlighted in red in Figure 2.6). Another aspect of multi stains have been handled with generating synthetic data. This is used to generate different stains as well as to generate more data from the same stain to generalise models.

Furthermore, the discussion of methods show that different problems focused by various researchers have followed a hybrid/combination of techniques to handle multiple challenges that needs to be addressed. For example, a method may have to address the lack of data, and large dimensionality in the same technique developed [51].

2.4.1 Limited labels

Limitations in labelled data is identified as a main factor challenging the translation of AI techniques to the digital pathology workflow (Figure 2.6). The limitations are due to the specific nature of the labelling procedure for WSIs. The limitation of labelled data has been identified as a subcategory based on prior research. The limitations are present in the form of small datasets, which are fully labelled but are insufficient, or weakly labelled data, which are partially labelled data or as one class dataset in which labelled data are only available for one of the classes. This is mainly due to the unavailability of sufficient data for rare instances. Also, there are instances with no labelled data. The sub-categories of the taxonomy for limited labels (small datasets,

Weakly labelled and unlabelled) and methods used to handle them are discussed below.

Small datasets

DL techniques require a large number of high-quality annotations [52]. However, due to complexities in acquiring images and labelling data, a DL task can be presented with a small dataset in order to develop a solution. Prior research identifies two main incomplete dataset types as scarce datasets, which have limited data, and weak datasets, which have partial labels or image-level labels. Our taxonomy also considered datasets without labels which can be categorised as unlabelled datasets. The two main approaches of handling the datasets are either by increasing the amount of labelled data in the datasets or developing models that can learn with smaller labelled datasets.

Augmentation techniques increase the amount of data in the dataset. There are different methods to augment data. Traditional augmentation by spatial transformation changes, which includes flipping, translation, rotation etc., adds variation to the dataset with additional data as well. Colour augmentation is another aspect in which the data are augmented by changes to different intensity statistic ranges [27, 53]. The brightness, saturation and contrast are the main statistics that researchers have focused to add colour-based augmentation to WSI datasets (Figure 2.7). Generating synthetic data using GANs is another method that is vastly applied in recent research methods. The learning based GANs learns the distribution of the training dataset and generates new data within the learnt distribution [54]. GAN models are also popular for image translation tasks, which can translate images from one style to another. Research has shown success in converting WSI images to the desired image style with style translation GANs. CycleGAN, StyleGAN, StainGAN are examples of image translation methods used for WSI [53-59]. However, GANs are computationally expensive, and the training dataset should consist of a sample that can capture the variation of the WSI domain. Therefore, synthetic data generation with simpler GANs are more feasible for digital pathology.
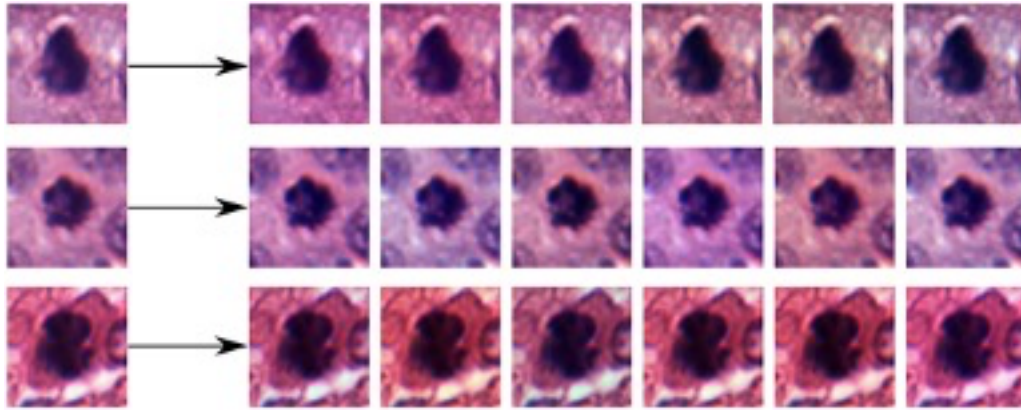
Figure 2.7 - Colour augmentation to WSI patches
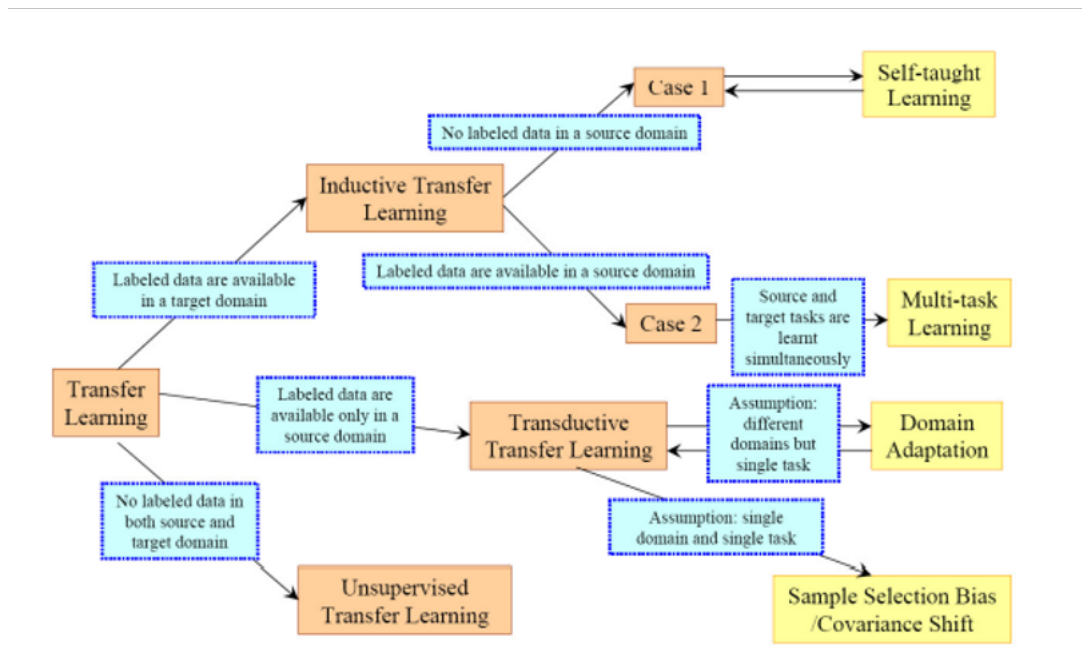
Figure 2.8 - Types of transfer learning methods [60]
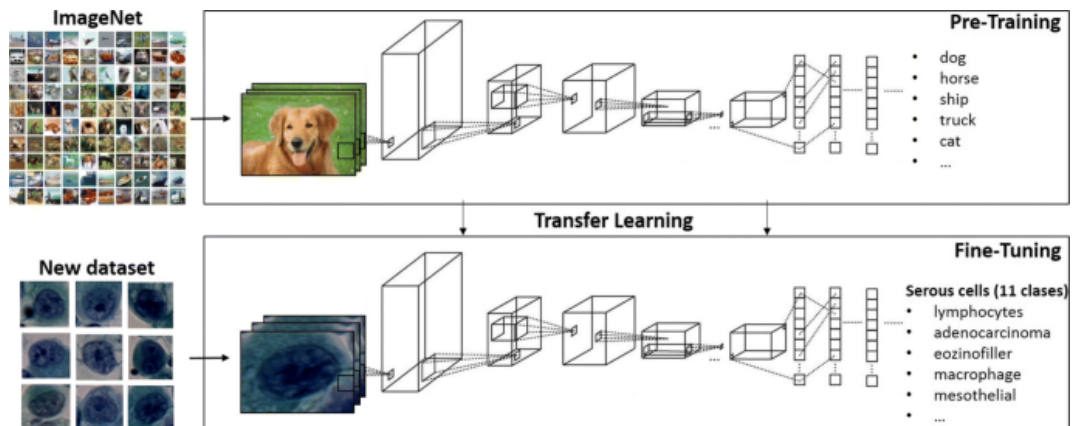
Figure 2.9- Creating a pre-trained model and using it to fine-tune for a medical dataset [62]

Transfer learning (TL) is a popular approach to handle small, labelled datasets. The distribution of transfer learning solutions is based on availability of labelled data. Figure 2.8 shows the types of TL based on the availability of labelled data [60]. If target domain has labelled data its identified as inductive TL, transductive TL will be when only the source domain has labelled data, and unsupervised TL is when labelled data are not available in source or the target. Inductive TL has been further divided into two categories as based on whether labelled data are available (case 1) in the source domain or not (case 2). Case 1 will be identified as self-taught learning, where case 2 is multi-task learning due to source and target being learnt simultaneously. Lastly transductive TL is categorised first as domain adaptation to perform single task over different domains. Second category is sample selection bias with the assumption of single domain and single task.

Recently transfer learning has been applied in two main ways. For fine-tuning or to use as a feature extractor. Transfer learning with fine-tuning uses pre-trained deep learning models and re-trains only the fully-connected layers in order to achieve the specific features of the small dataset [41, 61]. Since the pre-trained model consists of the basic features of data, large amounts of data are not required for training a full DL model (Figure 2.9) [62]. WSI classification with transfer learning methods has achieved better performances compared to models trained from scratch.

Few shot learning methods are popular for instances where a very small amount of data is available. Two networks are trained to learn the models. Chaitanya et al., (2019) proposed a technique using few-shot learning for segmentation based on augmentation [63]. Another technique developed used few-shot learning to classify whether an image patch contains tumour cells [64]. Another method proposed a method for object detection [51].

Weakly labelled

Weakly labelled datasets are defined differently in various research. Weakly labelled data can be identified as datasets with image-level labels and partially labelled datasets [15]. Partially labelled datasets can be a challenge to develop techniques where the

availability of labels is specific. However, new techniques have been introduced to handle unique problems in partially labelled data.

Multiple instance learning is a method that has shown potential in developing methods for datasets with limited labels. The data will consist of a majority of the negative class and a minority of the positive class. Dataset is arranged in two bags, one bag containing only negative data and the second bag consisting of both positive and negative data [65].

One-class labelled data only consist of data from one class, in the instance where few samples of labelled data are available from a rare class commonly known as the class of the interest challenge classification methods due to only having data for one class [66].

Instances of availability of unlabelled data and limited positive class data lead to methods inspired by anomaly detection to be applied to WSIs. Use of positive data to learn the algorithms and boost the performance using the unlabelled data is a semi supervised approach that has been used [67]. Limited research is available in semi supervised methods for WSIs. Therefore, there's scope to improve methods to reduce the translational gap using limited data.

Unlabelled

The final sub category of limited data is unlabelled data that do not have annotations for data. Therefore, clustering methods, auto-encoders are some examples of unsupervised learning methods [15, 24, 68]. Further improvisations and research need to be carried out on unsupervised methods as the knowledge available is limited. The use of DL methods as feature extractors is another popular component in feature learning for the development unsupervised technique.

2.3.2   High Variation

Another major challenge identified in Figure 2.6 is the high variability. The tissue structures are in high variability. Even patient to patient variability can be identified. To handle this, various normalisation techniques have been developed. Furthermore,
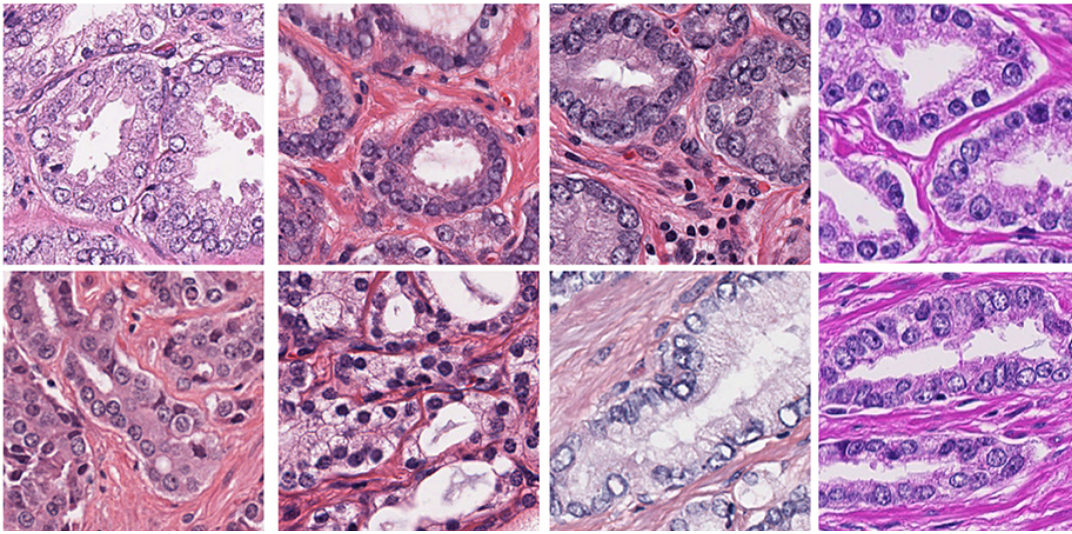
Figure 2.10 - Stain variation in WSI data [53]

due to the differences in the sample preparation procedures and scanning equipment, a variation occurs in the data which leads to poor generalisation in trained models for data from different locations, which were not included in training [26, 69]. Therefore, the performance is negatively impacted during the translation of the developed methods to the digital pathology workflow. Main causes for high variation have been identified as inter-site and intra-site variability and multi-stain mapping.

Inter-site and Intra- site variability

The staining techniques and scanners used at different laboratories cause colour variations in WSIs. Intra-site variability is when the WSI data produced at the same institution has a high variation mainly in the colour distribution (Figure 2.10) [53]. Intra-site variabilities occur due to different timing of staining procedures, and scanning process, and lighting conditions. Inter-stain variation result in high variation in data depending on the institution where the tissue specimens were processed. The inter-site differences mainly occur due to changes in slide preparation techniques and scanning equipment differences.

The stain variations affect the applicability due to performance changes of the automated diagnose techniques. Techniques/solutions to handle the above-mentioned problem take two main categories. Data-oriented solutions and algorithm-oriented problems [18]. Inter-site and intra-site differences in WSI causes poor generalisation

towards trained models [26]. Therefore, it requires methods to mitigate the low performances. The use of external data is one approach to handle the challenge, which results in increased variation in the distribution of data. Minimising differences is another approach to handle the differences in the data distributions [27, 53].

Multi-stain mapping

Use of different stain types on the same sample will highlight different cell types. The staining procedure must be performed for each stain type to be used. After the acquisition of specimen with one type of a stain, it needs to be de-stained which is about a 2-hour procedure [70]. The de-staining process consists of several critical manual steps from washing the current stain and preparing to apply the next stain without damaging the tissue [71]. Thereafter, another type of stain will be applied, and image acquisition is performed. Once the WSIs for the same specimen are prepared with different stains, the examination of conducted by experts. The examination on separate WSIs needs to be consolidated to finalise the examination's outcome. Use of image registration techniques will align the WSIs and support co-analysis for diagnosis. DL based registration methods have shown promising performances. However due to large dimensionality of images and complexity in methods, there's a need for improvement.
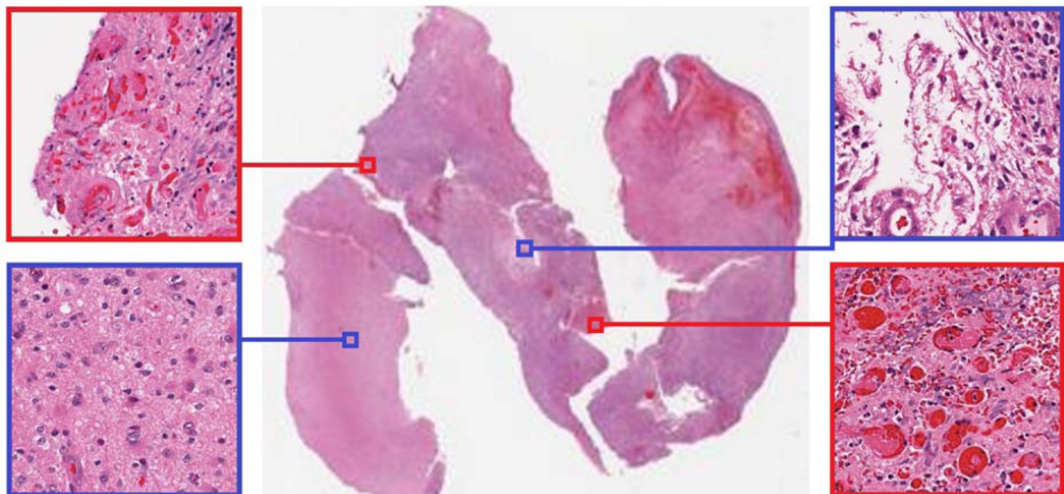
Figure 2.11 - The requirement of ROIs to apply DL [28]

2.3.3   Large Dimensions

The WSIs have captured the specimen prepared in the glass slide at multiple levels. Therefore, the pathologists can navigate the WSI at different zoom levels in order to examine and analyse. The different levels of the WSIs provide different characteristics regarding the tissue specimen's cells, tissue structures and special arrangements. The high-resolution WSIs have captured the entire specimen which enables to visualise and capture the rare formations in cells.

Large dimensions of WSI are a barrier to train models as well as to label images. The large dimensions can be reduced by focusing on the region of interest (ROIs) and processing the ROIs as input to the deep learning techniques (Figure 2.11) [28]. Rich feature extraction is also an important task. If methods are capable of extracting good (capture prominent differentiating features) features, then the model performances can be increased with a smaller amount of data [26, 72]. This type of approaches can support solving the issues with regard to high dimensionality and the need for a large amount of data.

2.5   SUMMARY

This chapter presented the literature for the objectives of the present thesis. The main aim of the thesis is to identify the factors influencing the translation of AI techniques to the digital pathology workflow. The cancer diagnosis process and the application areas of AI are discussed in order to derive the gaps in translating AI. The taxonomy further discusses the limitations in labelled data, high variation and large dimensions as broad areas of factors translating AI to Digital pathology. The taxonomy demonstrates three main approaches to handle the challenges of co-analysing multi-stained data, high variance due to inter-site differences and lack of annotated data.

Chapter 3: Non-Rigid Registration of Multi-Stained WSIs using Deep Convolutional Neural Networks with Super-Convergence

Image registration is a fundamental image processing task that matches the alignment of two or more images taken at different times, different imaging modes or different viewpoints. To register images, the spatial transformation is calculated to align images. A broad range of registration techniques have been developed that can be categorised as rigid registration and non-rigid registration. This chapter aims to enhance unsupervised non-rigid registration of WSIs using super-convergence. Recent AI techniques to WSI tasks have shown the capability of unseen features of high dimensional WSIs. The application of deep features to the registration of multi-stained WSIs have also showed improvements compared to traditional registration methods. Identifying matching features under each stain type is difficult compared to the capability of DL models in learning hidden features from WSIs. The high variations and large deformations in different stains leads to the need of investigating methods to better handle WSI registration with DL to increase potential in improving pathology workflow. The registration can assist an important sub-step of co-analysing multi-stained WSIs for diagnosis. Deep models have shown promising results in registering several types of stains faster, using a combination of rigid and non-rigid methods. These deep models are complex model architectures requiring large amounts of memory to train. Therefore, the focus of this chapter is to investigate the complexity of the deep models and explore the possibility of faster convergence in deep models. The application of super-convergence to a simpler deep model instead of the U-Net like model has shown promising improvement to the model. This chapter consist of sections 3.1 and 3.2 which comprise an introduction and literature review discussing the background and the aims of the research. Section 3.3 presents the proposed improved technique followed by the results and a discussion in sub-section 3.4. The conclusion in the section 3.5 summarises the chapter.

3.1 INTRODUCTION

Cancer diagnosis is a part of the digital pathology workflow conducted by expert pathologists. Identifying and analysing patients' samples requires examination of abnormal cell structures and spatial arrangements [3]. The use of multiple stains to prepare the specimens supports the examining stage by highlighting different categories of cells. There are various stain types which are used including, Hematoxylin and Eosin (H&E) and with immunohistochemistry (IHC) with an antibody against the estrogen receptor (ER), progesterone receptor (PR), and Her2-neu [21, 50]. The Figure 3.1 shows an example of a WSI under multiple stain types.

These stains used in breast cancer diagnosis assist to observe different features in the specimen. The commonly seen H&E staining consists of Eosin which is combined used to identify basic tissue structures. Eosin is a negatively charged acidic colour stain which shows tissue structures in red or pink while hematoxylin dyes the acidic structures in purplish blue showing the nucleus in purple stain. IHC is another mostly used staining type which assists in predicting the stage of cancer according to the presence of types of proteins [73]. ER or PR positivity gives information on important prognostic and predictive features of breast cancer which can influence the therapeutic choice and patient survival [74, 75]. HER2-neu is another stain type in breast cancer diagnosis providing prognostic indicators to guide therapeutic decisions. The HER2/neu positive breast cancer prognosis is indicative of the presence of human epidermal growth factor receptor 2, which promotes growth of cancer cells. The ER, PR and HER2/neu evaluations together are required for management and prognosis of breast cancer. [76, 77].

The different cells using one stain need to be co-analysed with the cells identified using other types of stains. Differential cell characteristics from each stain type provides information for the diagnosis. Due to the complexities in large WSIs, identifying corresponding characteristics for co-analysis is difficult [49]. This is due to the process of re-staining the same specimen for each stain type which results in complex and large deformations. The Figure 3.2 shows the deformation of the same region of the same WSI (rectangle box) in different stains in which rotations and translations are clearly visible. Due to clearing of one stain and applying another stain, the samples may be

deformed. The slight movements, folds and artefacts can occur in subsequently applied stain. During the scanning process, the environment lighting conditions and use of different scanners added further deformations to the WSI [45, 49, 78]. The large deformations need to be resolved to identify corresponding cells in the differently stained WSI samples. The deformations are non -rigid deformations due to the changes to structures and spatial arrangements and requires non-rigid image registration. Non-rigid registration for digital pathology poses unique challenges due to the high resolution, large dimension of images, and the need for fast and accurate performance despite the complex deformations [21, 31].

Expert pathologists are challenged by the need to visually locate corresponding locations of samples which are de-stained and re-stained with various stains [21, 31, 44]. Deformations occur at slide preparation and image acquisition steps. Therefore, consolidating observations on differently stained WSIs of the same sample is complex due to possibilities of being slow, inaccurate and difficulties in applying it to stacks of WSI images of the same sample [49]. Therefore, automated methods to register WSIs are beneficial to reduce the complexity of examining slides with multiple stains [31]. However, automated WSI registration has unique limitations compared to radiology image registration. The high resolution of WSIs, complex and large non-rigid deformations and specimen artefacts due to the re-staining leads to the need of methods addressing the complexities in WSIs [24, 49]. Current computational methods for WSI registration take two main approaches as traditional approaches and ML based approaches. Use of traditional approaches and shallow ML techniques have shown limitations in handling the high dimensionality of the WSIs [16, 21]. The methods require to identify features separately which is complex due to the high variations in multi-stained data. The traditional methods are time consuming to register a pair of WSIs. In comparison, recent methods developed using deep models for registration have shown improvements in identifying hidden features automatically and registering images faster [21, 50]. This is an important outcome to develop AI techniques which could translate to the digital pathology workflow.

The DL methods to solve this complex multi-stain registration uses rigid and non-rigid approaches [21, 50, 79, 80]. Due to the large deformations, the use of rigid methods first aligns the WSIs according to rigid transformations. As the next step applying non-

rigid registrations addresses non-rigid transformations for a better outcome. The combination of these two has shown promising results.

However, this combination is computational costly and cannot be reproduced without a high-power GPU (GTX RTX 2080 minimum). High computational cost is required by the U-Net like autoencoder which is used for non-rigid transformations. Methods requiring high computational costs are limiting the potential opportunities in translating the methods to the pathology workflow. Methods which can successfully adapt to variations in several stains are important to the co-analysis step of the diagnosis process. Therefore, in this chapter a smaller architecture is introduced to learn deep features for the non-rigid transformation which uses super convergence. The proposed architecture can learn the non-rigid transformation and improve the non-rigid registration compared to the U-Net-like autoencoder's outcome.
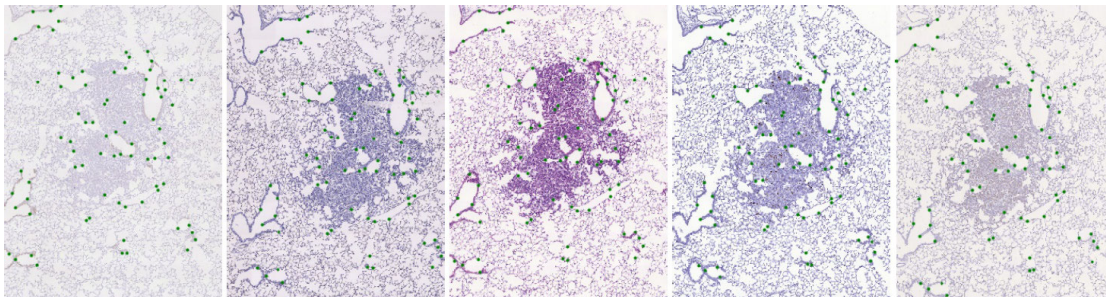
Figure 3.1 - Same WSI in multiple stains from ANHIR challenge data
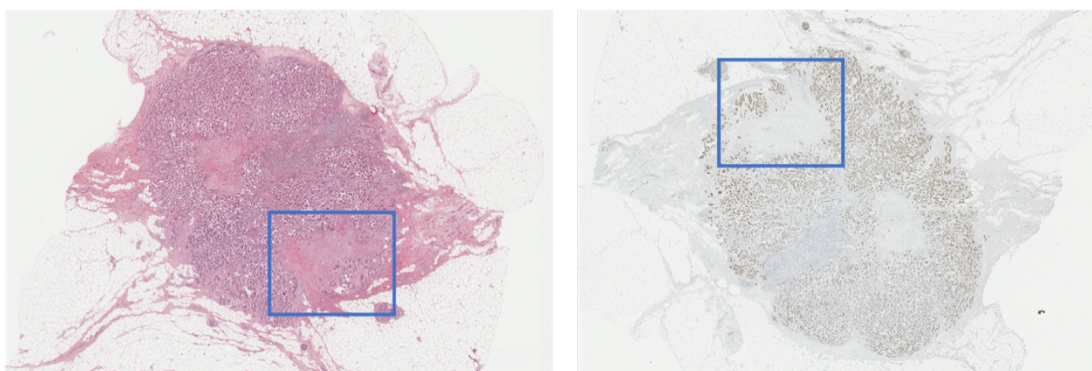
Figure 3.2 - Shows the deformation of the WSI in 2 different stains (H&E and ER) of ANHIR data

3.2   RELATED WORK

Image registration is used in many image processing applications, ranging from medical image processing to 3D image construction [43, 81]. However, image registration has shown a profound effect on medical image processing which can help reduce human error in medical image-based diagnosis [49]. Registration has been applied for radiology imaging types including MRI, CT, X-ray, retinal images and histology images known as WSIs [78, 81].

Prior research for radiology image registration uses segmented images to train CNNs for local and global image registrations [82], an unsupervised method using a convolutional stacked auto-encoder (CAE) to extract features from fixed and moving images showed promising registration outcome [83]. Categories of radiology images consist of different features which provides different types of information. Integration of these information is important for clinical diagnosis [43].

Methods applied to radiology imaging are not directly applicable to WSIs due to the differences in them [28]. The large dimensions of WSIs require them to be processed in smaller regions of interest or patches. Image-level and pixel level annotations are important to differentiate different areas and cells at different magnifications. The large variations in WSIs due to slide preparation, staining process and scanning methods from different institutions challenge developing generalised automated methods for diagnosis. The large deformations in tissue samples caused by re-staining and re-scanning adds another challenge requiring methods to handle misaligned WSIs to support co-analysis. The co-analysis is a crucial step for diagnosis by visual examination is prone to errors and inter and intra variability and therefore efficient AI methods for registration will largely benefit the process [21, 43].

The registration techniques allow to align the WSIs by transforming the different alignments. Image registration can be broadly categorised into two categories, rigid registration and non-rigid registration. Prior literature for WSI registration mostly focuses on traditional methods, with recent research exploring ML based methods.

3.2.1 Traditional methods

The traditional methods focusing on intensity, landmarks and features have been developed. Due to the complexities in WSIs, the traditional methods are not directly translatable and does not perform same when applied to different stains [21]. Prior research showed that different pairs of stains perform differently when same methods are applied [21, 45]. However, combinations of multiple methods have shown better performances for WSI registration [21, 44].

Intensity based methods require mutual information to be presented in the pair of WSIs to be registered [49]. The purpose of using different stains is to identify specific cell types which another stain would not capture. Therefore, the relationship of pixel intensities vary in which intensity based registration shows better outcome with landmark based and feature based registration [49, 84]. The landmark-based methods require accurate corresponding points identified in the WSIs to be registered, which is a complex task. Therefore feature-based methods are developed for WSI registration. Jun Jiang et.al (2019) developed a feature-based method using information-rich low resolution layers for kernel density estimation and regressing against the hierarchical structure [85]. However as the authors mention, this method have not incorporated non-rigid transformation and has been tested only with two types of stains, IHC and H&E [85]. The piecewise approach for WSI registration compares Scale Invariant Feature Transform (SIFT), Multi-Scale Oriented Patches (MOPS) and using a tool developed for manual comparisons. Comparing automated features and manual features across regions of interests or full WSI is a tedious task where different types of stains will show different cell types which means different features to be compared [84].

However, despite the satisfactory performance using multiple methods makes the registration process complex and time consuming. There is a limitation when it comes to translating the developed methods to the digital pathology workflow. It is important to complete the registration in a reasonable time to proceed with steps needed for diagnosis [49]. The need for identifying the suitable features and parameters is a tedious task [21]. It is important to minimise the complexities in the registration process.

3.2.2 Machine learning-based methods

ML methods have shown potential in digital pathology due to the capability of learning features from given data. However, ML methods have limitations of curse of dimensionality. ML models require careful engineering and domain knowledge to develop a feature extractor [16]. On the other hand, DL methods have shown better registration outcome accuracy and time compared to traditional methods [21, 45, 78]. Compared to ML's shallow learning due to the deep nature of model architectures, DL is capable of extracting hidden features automatically from high dimensional WSI data [16]. DL based techniques have shown promising outcome for WSI registration. DL methods extract features and can identify unseen patterns resulting in more features for registration.

- **Supervised learning**

DL methods trained in a supervised manner requires large amounts of labelled data for training. WSI labelling is a complex process in which pathologists examine WSIs of every patient in multiple magnification levels [78]. Therefore methods, which do not use labelled data or only use limited number of data are feasible [82]. Unsupervised methods which do not use labelled data in training is beneficial for registration [43, 86]. The limited research in unsupervised methods using deep features has shown the potential for WSI registration. Improving the efficiency in registration step with simpler methods which is not time-consuming leads to a better outcome in the following steps to diagnosis.

- **Unsupervised learning with Transfer-learning Feature Extraction**

Due to the need of large amount of labelled WSIs, the registration methods have focused on the possible methods which use limited number of labelled WSIs or unsupervised methods which do not use labelled data [83, 87]. The affine registration using unsupervised ResNet features for rigid registration with non-rigid features using U-Net shows comparable promising results [50].

The unsupervised feature extraction has shown promising outcome but is less explored due to model complexities and high computational requirements [44]. WSI registration is only a small step towards the WSI analysis and final diagnosis. Therefore, methods with simpler models using reduced computational memory would be beneficial to the digital pathology workflow [44, 49].

Computationally efficient solutions are an important factor to apply AI methods to real world applications. Previous research has focused on methods which could generate an outcome with a least cost in time and computational memory. Prior research explores the concept of super convergence for deep learning focusing the learning rate, batch size, weight decay and momentum [88]. The performances for Cifar dataset show a comparable result within a smaller number of training epochs [88]. Super convergence for WSI data has improved performance for Pap smear image classification.

3.3   METHODOLOGY

The method presented in this chapter combines a rigid and non- rigid architecture for WSI registration. The rigid method aligns the majority sections of the WSIs. However, fine tuning the registration is important for accurate registration. The method introduced by Wodzinski et al., (2021) uses a U-net like encoder decoder model for non-rigid transformations. This approach is time and memory expensive [50].

The chapter introduces a novel architecture for the non-rigid transformation by using a 2-D Convolutional Neural Network (CNN) with the use of super convergence which achieves comparable results to the Wodzinski et al., (2021) with a large encoder decoder architecture [50].

The proposed architecture uses a 3 layer 2-D CNN with a lower learning rate and a higher weight decay as shown in Figure 3.3. This approach helps the model to achieve faster convergence with limited computational power and reduced time. The proposed non-rigid architecture has a smaller, deep learning architecture and achieves comparable results.

WSI registration uses both rigid transformation and non-rigid transformation. The non-rigid transformation is highly computationally inefficient and requires GPUs and large memory capacities. Furthermore, due to the complexity in the architecture, the model cannot be considered the optimal architecture for non-rigid transformation for WSI registration. The chapter works on the non-rigid transformation for WSI registration based on findings of prior research showed the need of addressing complexities in methods for non-rigid transformations [50]. This chapter introduces a smaller CNN structure for non-rigid transformations. A three-layer 2D CNN architecture is modified for super convergence. The super-convergence has shown to be able to train models faster with comparable results. Reducing the learning rate and increasing the weight decay is a common method of achieving super convergence and by using large learning rates to regularise training and reduce other regularisation methods to balance the regularisation [50, 89].

The simple CNN architecture uses less computational power and memory compared to the prior research outcome. The U-Net based model uses an encoder and decoder architecture which is highly computationally expensive and costly [50]. The 3-layer CNN architecture has an input of a 2D-CNN and output of a 2D-CNN with Relu transformation. The CNN is as shown in Figure 3.3. The proposed architecture using Super Convergence is developed by using PyTorch. The non- rigid registration was implemented in a GPU. The non-rigid CNN architecture follows layers of 2D conv layer, pooling layer, dropouts with learning rate and weight decay. Furthermore, the batch size was reduced to 1. The deformation field from the CNN was extracted for the non-rigid registration. Furthermore, the affine registration step was simplified by reducing the fully connected layers.
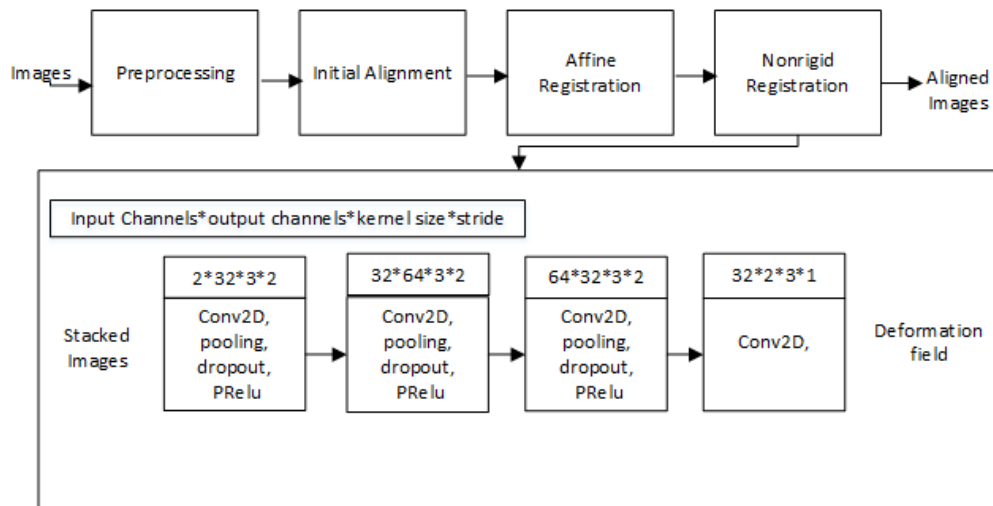
Figure 3.3 - Improved non-rigid registration with 2D CNN which uses super convergence

3.3.1 Dataset

The ANHIR challenge dataset is used in the experiments. The challenge was created due to the complex deformations that occur in tissue samples after staining. The pathologists require to identify corresponding features of same WSI under different stains. Due to deformations, it becomes a tedious task. Therefore, the challenge aimed at developing automated methods which are accurate, fast and robust. The high-resolution dataset consists of data from various organs in different stains. The original size of the WSIs varies from 15,000 × 15,000 to 50,000 × 50,000 pixels.

The data is provided in different scales. The medium dataset consists of data with a scale of 20 – 25. The dataset consists of, breast, COAD, lung lobes, lung lesions, kidney, and mouse kidney. Each WSI is provided in several stains including, clara cell 10 protein, prosurfactant protein C, hematoxylin and eosin, antigen KI-67, platelet endothelial cell adhesion molecule (PECAM-1, also known as CD31), human epidermal growth factor receptor 2 (c-erbB-2/HER-2-neu), estrogen receptor, progesterone receptor, cytokeratin, and podocin. For example, the breast tissue contains unstained adjacent sections stained with H&E and IHC with ER, PR and Her2-neu.

3.3.2 Experiments

The breast cancer dataset consists of WSIs upto 40x magnification which has a average pixel size of 65,000 x 60,000. The experiments consisted of pre-processing, initial alignment and affine registration according to the methods followed by Wodzinski et al., (2021) [50]. Following steps were taken to pre-process the images. Image pairs were padded and converted to uncompressed .mha format. Once data are loaded images are converted to grayscale and down-sampled to 512 pixels to a lower resolution. These images are the input to the segmentation which then is input to initial transformation. The affine registration by a ResNet-like model is conducted before Nonrigid registration. A CNN with four layers is used for nonrigid registration. The Figure 3.3 demonstrates the steps of the method, including the CNNs architecture. The images after the affine registration are the input to the CNN to learn the features and deformations.

The experiments were designed to demonstrate the capabilities of the proposed architecture to improve non-rigid registration mainly to reduce the complex architecture of model DHR_20 shown in Table 3.1 and Table 3.2. The experiments aim to explore the proposed simplistic model using super-convergence compared to the U-Net like architecture.

Optimal model architecture was identified experimentally by comparing the CNN without dropout and CNN with dropout for regularisation. The effect of the concept of super-convergence is a crucial factor for faster convergence. Therefore, we have evaluated the effects of using low learning rates with high weight decay and high learning rates with low weight decay to the study the possibility of faster learning and model convergence.

3.3.3 Evaluation

The evaluation of the registration method was conducted by calculating the Target Registration Error (TRE) values and timing for registration (Equation 3). In TRE landmark coordinates are the x^T and x^W of the target and warped image and Euclidean distance is *de(.)*. All TRE values are normalised by the image diagonal rTRE (Equation 4), where w and h are image weight and height respectively. Additionally, other parameters are, Average robustness, Average median rTRE, Average rank of

median rTRE, Average max rTRE and Average rank max rTRE, Average execution time measured in minutes. The results are calculated using the online submission evaluations of ANHIR challenge [21].

$$TRE = d_e (x_t^T, x_t^W) \qquad (3)$$
$$rTRE = \frac{TRE}{\sqrt{w^2 + h^2}} \qquad (4)$$

3.4   RESULTS

The experiments were conducted to identify the effect of simpler CNN model for non-rigid registration. The effect of dropouts and super-convergence were considered to achieve a comparable result at 10 epochs and exceed the performances at 20 epochs. A GeForce GTX 1080 Ti was used to conduct all experiments. Following are the nomenclature for the experimented models presented in the Tables 3.1 – 3.2.

1. DHR - DeepHistReg – The model presented at ANHIR challenge [add ref]
2. DHR2 - The DHR model trained for 20 epochs to compare the performance.
3. BMC_ND - Proposed base model CNN instead of U-Net model. The proposed CNN is tested without adding Dropouts.
4. BMC - Proposed Base Model CNN tested with dropouts
5. HLLW1 - Proposed CNN model with super-convergence. High Learning Rate, Low Weight Decay trained for 10 epochs.
6. LLHW - Proposed CNN model with super-convergence. Low Learning Rate, High weight Decay for 10 epochs
7. LLHW1N - LLHW without Dropouts.
8. HLLW2 - Proposed CNN model with super-convergence. High Learning Rate, Low Weight Decay (HLLW1) trained for 20 epochs.

Feature extraction using a simpler CNN model compared to the U-Net like model used was one of the aims of the experiments. The benchmark results (DHR2 in red, Table 3.1) at 20 epochs for the breast cancer data of ANHIR data were compared. The CNN without dropout (model BMC-ND of Table 3.1) showed higher TRE for training and testing TRE values and low robustness as Table 1 shows. Therefore, the model was tested with dropouts (model BMC of Table 3.1) due to over-fitting. Compared to the

benchmark results of model DHR2, the model with dropouts resulted in lower performances, but it showed an improvement compared to the CNN without dropout (BMC_ND).

Table 3.1 - Overall performance for breast cancer data. DHR - DeepHistReg, DHR2- DeepHistReg at 20 epochs, BMC_ND – Base Model CNN Without Dropout, BMC – Base Model CNN, HLLW1 – High Learning Rate, Low Weight Decay for 10 epochs, LLHW – Low Learning Rate, High weight Decay for 10 epochs, LLHW1N – LLHW10 No Dropout, HLLW2 – High Learning Rate, Low Weight Decay for 20 epochs.

	DHR	DHR2	BMC_ND	BMC	HLLW1	LLHW1	LLHW1N	HLLW2
Median-max-tre	0.018	0.380	0.406	0.392	0.421	0.396	0.382	0.365
Avg-max-tre	0.021	0.385	0.402	0.386	0.401	0.393	0.374	0.368
Median robustness	1.0	0.924	0.810	0.794	0.632	0.871	0.810	0.897
Avg robustness	1.0	0.560	0.512	0.553	0.516	0.558	0.550	0.567
Median-median-tre	0.002	0.235	0.217	0.195	0.231	0.221	0.216	0.226
Avg-median-tre	0.004	0.215	0.228	0.212	0.239	0.219	0.220	0.210
Median-avg-tre	0.003	0.220	0.231	0.209	0.234	0.221	0.211	0.215
Avg-avg-rtre	0.005	0.214	0.230	0.212	0.235	0.218	0.217	0.209

Table 3.2 - Training and evaluation for Breast cancer data. DHR-DeepHistReg, DHR_20-DeepHistReg at 20 epochs, BMC – Base Model CNN, HLLW1 – High Learning Rate, Low Weight Decay for 10 epochs, LLHW1 – Low Learning Rate, High weight Decay for 10 epochs, LLHW10-ND – LLHW10 No Dropout, HLLW2 – High Learning Rate, Low Weight Decay for 20 epochs.

	DHR	DHR 2	BMC	HLLW1	LLHW1	LLHW1 N	HLLW2
Median-max-tre	0.262	0.380	0.395	0.425	0.408	0.378	0.365
Avg-max-tre-train	0.023	0.377	0.387	0.412	0.406	0.388	0.391
Median-max-rtre eval	0.018	0.394	0.388	0.402	0.393	0.387	0.361
Median-robust train	1.0	0.925	0.898	0.974	0.974	0.936	0.949
Avg-max-tre-eval	0.021	0.387	0.386	0.399	0.389	0.371	0.362
Avg-robust train	1.0	0.811	0.803	0.804	0.811	0.773	0.805
Median-median rtre-train	0.007	0.235	0.175	0.254	0.215	0.216	0.229
Avg-median-rtre-train	0.007	0.234	0.191	0.247	0.229	0.242	0.230
Median-Average-rTRE-train	0.008	0.220	0.208	0.234	0.219	0.209	0.228
Median-Robustness eval	1.0	0.451	0.414	0.241	0.522	0.485	0.534
Average-Average-rTRE-train	0.008	0.231	0.206	0.237	0.226	0.242	0.233
Average robustness eval	1.0	0.497	0.491	0.444	0.495	0.494	0.508
Median-Median-rTRE eval	0.001	0.231	0.217	0.230	0.225	0.216	0.212
Average-Median-rTRE-eval	0.003	0.211	0.217	0.237	0.216	0.214	0.205
Median-Average-rTRE-eval	0.002	0.225	0.221	0.233	0.225	0.215	0.213
Average-Average-rTRE-eval	0.004	0.209	0.213	0.234	0.216	0.211	0.203
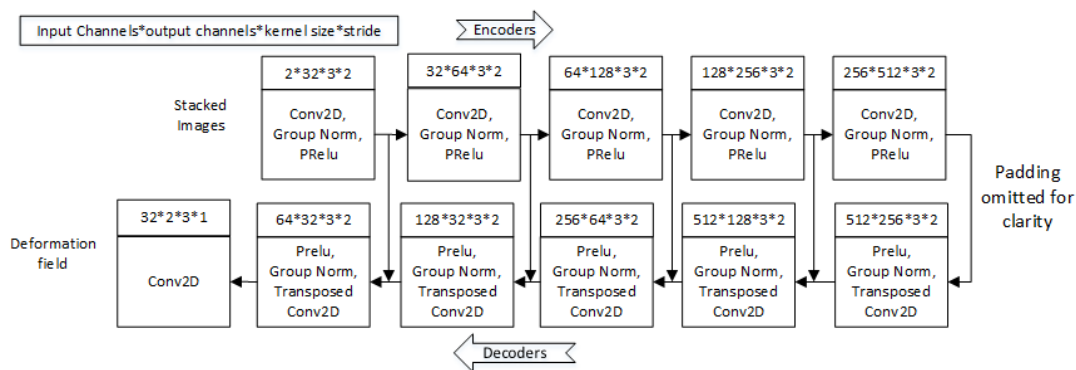
Figure 3.4 - The Benchmark model's U-Net like model for non-rigid WSI registration

Due to the drawback of benchmark method's model complexity and high training time, super-convergence was added to the CNN. We tested low learning rates and high weight decay and high learning rates and low weight decay to study the effect of faster convergence to the model performances. Both models showed an improvement. However, the model trained with low learning rates and high weight decay for 10 epochs resulted in comparable results to DHR2 and at 20 epochs it (HLLW2 model text in red Table 3.1 and 3.2) could exceed the result of DHR2 (shown in red in Table 3.1 and Table 3.2)

3.5   DISCUSSION

Non-rigid transformation is a computationally intensive task which requires a high computational power with multiple GPU cores. Addressing non-rigid transformations in WSIs is a key factor to improve registration performance, due to high deformations in the multi-stained WSIs. Therefore, non-rigid transformation is a key aspect to identify and address the transformations. The method proposed by Wodzinski et al., (2021) used a U-Net like architecture (Figure 3.4) which consists of an encoder and a decoder [50]. The authors specified the complexity of the non-rigid registration method as a limitation of the method. Identifying the non-rigid transformations require learning low-level features from both images to match similarities. Registering non-rigid transformation requires complex models for high dimensional feature recognition. However, the practicality of using complex models in the digital pathology workflow for non-rigid transformations is low. Therefore, finding a solution which is not complex, and suits multiple stain types is an essential requirement.

Importance of bridging the gap in translating DL models to the digital pathology workflow is increasing. DL model architectures to learn specific features of a limited number of stain types limit the model's generalisability to wider range of stain types. Therefore, smaller DL architectures learning abstract features, are sufficient to identify the required features for registration. Complex DL model architectures requiring high computational power to train high dimensional WSIs is another limitation handled by this chapter. Super-convergence is applied in the proposed approach for faster training of DL models.

The proposed model is a small CNN model that is capable of performing similar to a complex model's non-rigid transformation, with a lower number of epochs (Table 3.3 and 3.4) Comparison against a U-Net like model has shown that the proposed architecture outperforms the U-Net model-based approach. Therefore, the smaller CNN architecture with super-convergence has shown to achieve similar results to the encoder decoder architecture.

The improved model's TRE values are a result of the smaller CNN architecture which could extract similar features as the complex U-Net. The Table 3.4 shows the comparison of the two model architectures where the U-Net model has 11 layers and the CNN has 4 layers. These features extracted from the smaller CNN architecture are sufficient for the non-rigid transformation which shows comparable results at 10 epochs and exceed at 20 epochs. Therefore, adding a complex auto-encoder is not required to achieve high result that would consume higher computational memory. Furthermore, using super-convergence, the model is capable of converging faster to an optimal convergence. However, to understand and synthesise the simple model, multiple experiments had to be carried out.

Initially, the CNN size was reduced and given a simple CNN architecture to understand the feature extraction capabilities and reduce the computational costs. Use of the CNN minimised the computational costs to train the non-rigid registration. However, the registration performance was lower than reported which. During the training procedure, overfitting was noticed, and therefore, dropout rates were introduced (Table 3.1).

Table 3.3 - Shows an example of source and target images and the transformed image for benchmark method, CNN outcome before applying dropout and super-convergence, and finally the transformation for the proposed model
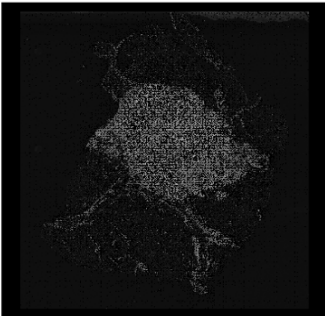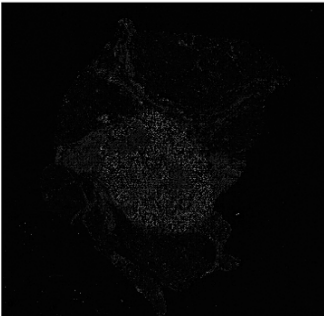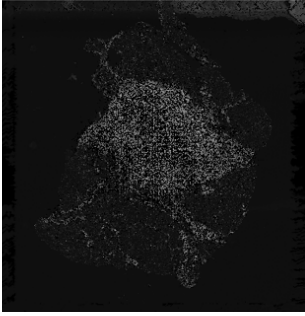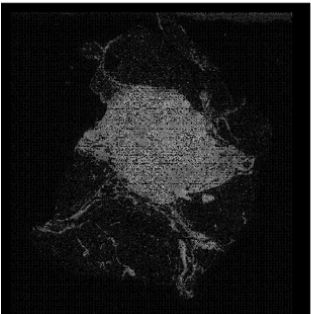
Source	Target

	Benchmark method	
Transformed image	Before applying dropout and super-convergence to the CNN	
	Proposed improved model architecture	

Table 3.4 – Comparison of the base model and proposed model architectures

Non-Rigid Registration Network	Base Model	Proposed Model
Model	U-Net like	2D CNN
No of layers	11	4
Kernel size	3	3
Stride	2	2
Activation Function	PReLU	PReLU
Dropout use	-	√
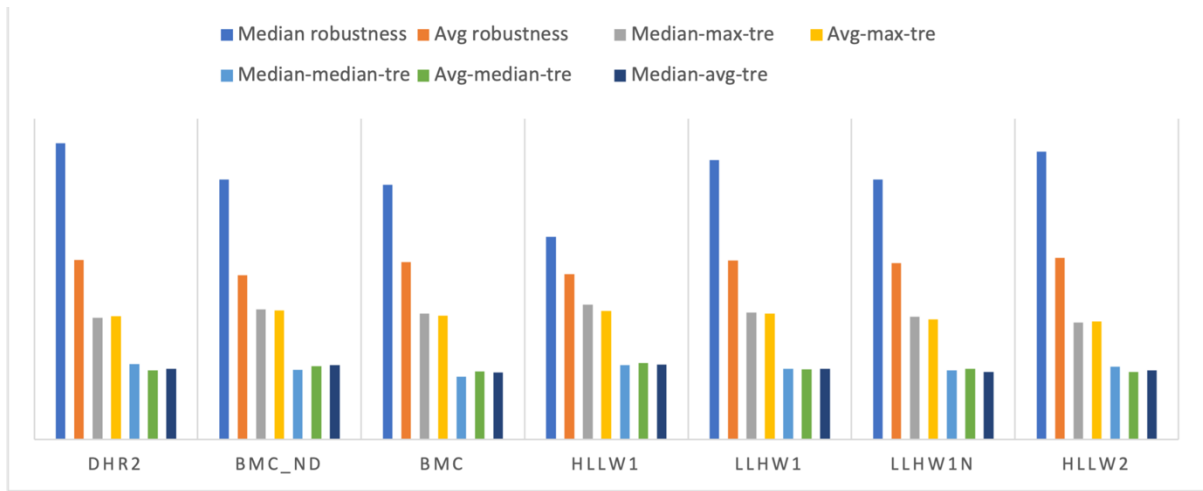
Figure 3.5 - Overall performance for breast cancer data
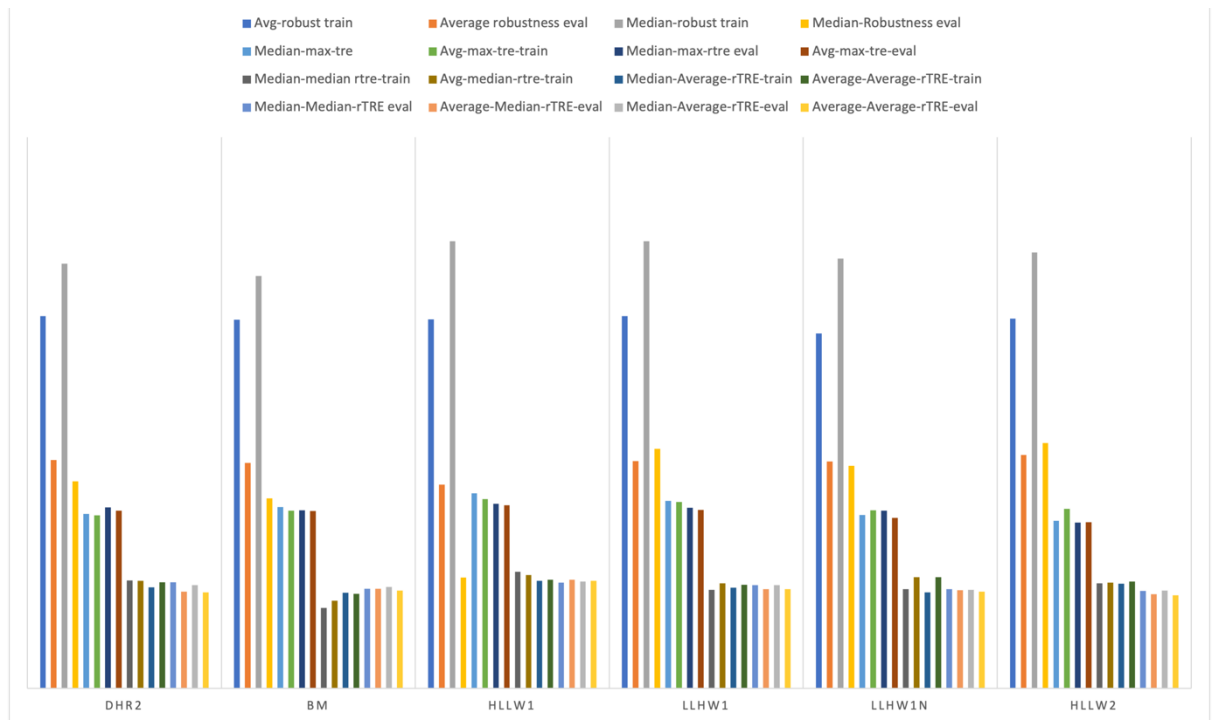
Figure 3.6 - Training and evaluation for breast cancer data
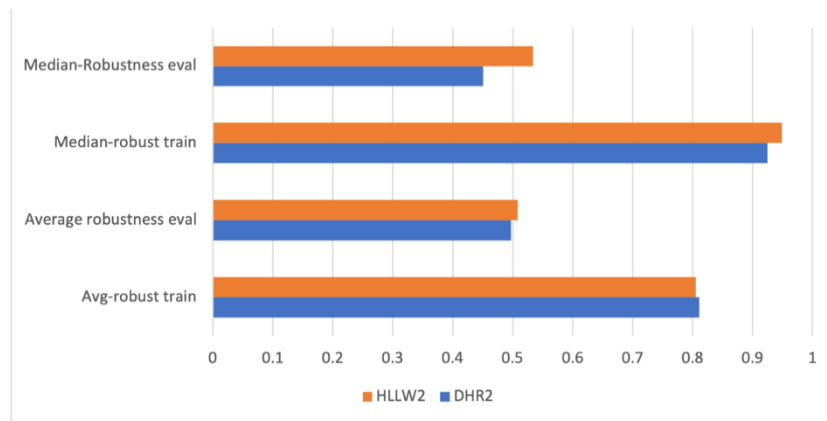
Figure 3.7 - Robustness of the proposed model

Table 3.5 - Comparison of time taken to register 10 pairs of WSIs. Base model and proposed model are compared.

Sample	Time to register (Seconds)	
	Benchmark model	Proposed model
1	0.767	0.746
2	0.759	0.730
3	0.754	0.736
4	0.811	0.762
5	0.770	0.766
6	0.791	0.774
7	0.773	0.766
8	0.784	0.764
9	0.678	0.655
10	0.762	0.747

Use of dropouts handled the model over-fit, however the model did not train sufficiently. Therefore, super-convergence was introduced to let the model learn faster and adapt faster. This approach also, prevented the model from overfitting, but produced a higher result compared to the U-Net based model. Table 3.3 shows an example of the registered image pair of the benchmark method and the registered pair with the improved model.

The Figure 3.5 and Figure 3.6 demonstrates the overall performances and training and evaluation performances of each model. The reduced error rates and improved performance using a simpler model is important as a sub-step in the cancer diagnosis process.

Figure 3.7 further demonstrated the robustness of the models. The median and average values for the evaluation and training stages are comparable and improved with the use of super-convergence in a simpler model for generalised model outcome. Finally, the time for registration was evaluated. Ten sample pairs of the dataset are shown in Table 3.4 with the time spent to register each pair. Comparing the base model and the proposed model timing shows how the proposed model has reduced the time to register. The high robust model with lower error rates leads to time efficient registration of WSIs.

Therefore, it is evident that this approach is a simple and less complex approach compared with the U-Net-like auto-encoder. The introduced model is computationally less complex and dynamic to learn with only 10 epochs.

3.6   CONCLUSION

In this chapter, registration techniques for multi-stained WSIs are explored. The high deformations in WSIs require methods that are able to learn complex features in data to address the transformations. However, the non-rigid methods applied for WSIs are complex models requiring large amounts of data and computational memory which limit the translation of methods to the digital pathology workflow. Therefore, computationally efficient methods that are trained with the least number of labels or allow unsupervised training are needed.

The present approach is a novel, improved, non-rigid registration method for multi-stained WSIs. The results were improved with super-convergence by which the model converged faster as a technique to train a model faster. The improved method achieves improved results with a reduced number of training epochs using a simple CNN. Thus, the model is simplified as well as trains with faster convergence, which helps to translate the method to the digital pathology workflow.

Chapter 4: Improving Follicular Lymphoma Identification using the class of interest for Transfer Learning

Among the leading deaths from cancer, Follicular Lymphoma (FL) is a type of lymphoma that grows silently and is usually diagnosed in its later stages. FL is the most common subtype of non-Hodgkin lymphomas. To increase the patients' survival rates, FL requires a fast diagnosis. While, traditionally, the diagnosis is performed by visual inspection of Whole Slide Images (WSI), recent advances in deep learning techniques provide an opportunity to automate this process. The main challenge, however, is that WSI images often exhibit large variations across different operating environments, hereinafter referred to as sites. As such, deep learning models usually require handling the variation of data from each new site. This is, however, not feasible since the labelling process requires pathologists to visually inspect and label each sample. In this chapter, we propose a deep learning model that uses transfer learning with fine-tuning to improve the identification of Follicular Lymphoma on images from new sites that are different from those used during training. Our results show that the proposed approach improves the prediction accuracy by 12% to 52% compared to the initial prediction of the model for images from a new site in the target environment.

4.1 INTRODUCTION

Follicular Lymphoma (FL) is one of the common B-cell non-Hodgkin lymphomas [90]. It frequently presents at a late clinical stage as painless lymphadenopathy, with symptoms often only appearing after its transformation to more aggressive forms [91]. Accurate and early diagnosis is a critical step as it could improve chances of survival [92]. However, the current diagnostic procedure is time-consuming and requires expert knowledge and multiple reviews before finalising the decision [36]. With the increasing number of new cancer cases being identified [93], the diagnosis process becomes even more challenging.

WSI led ML techniques applied to digital pathology [7] DL, in particular, has achieved high performance in many applications due to its capability to learn the features that are best-suited for the application at hand [7, 8, 18, 94]. However, using DL for the analysis of histology images is challenging due to the lack of labelled training data. This has motivated research towards developing supervised learning techniques that require a minimal number of labelled data [15, 95].

This generalisation problem for data from new sites has been addressed by using pre-processing techniques [7] e.g., by grayscale conversion, colour normalisation, and colour augmentation, or classification algorithms [61], to minimise variation in the data [7]. Grayscale transformation can result in a loss of important features compared to colour representations. Colour normalisation techniques, which match the colour distributions of the source and target domains, are domain-specific [7] and thus require extensive tuning to accommodate variations in data from different sites [98]. These techniques do not perform well on data from a new site as they are developed specifically for the data sites used in training. Therrien et al., (2018) focus on the effect of improving the variation of data used for training by comparing models trained with data from a single site and multiple sites [99]. They showed that a model trained with multiple sites generalises better to data from an unseen new site compared to a model trained with a single-site data.

Since these techniques are specific to the model and datasets on which they have been trained, they do not perform the same on a new test dataset coming from a new site

[26]. Thus, the possibility of re-using a deep learning model trained on one site is limited, especially since some labelled data of past research are not available to the research community [15]. This degradation in performance raises the need to re-train them with a large amount of labelled data from the new site, which is not practical in the pathology context [100]. In fact, to create labelled datasets, pathologists need to visually inspect and manually label WSIs for each class. This is, however, time consuming and can incur a high financial cost.

This chapter explores the possibility of using a small dataset from a class of interest to train deep learning algorithms for FL detection from WSI images. Present research is a method to reduce the need for a large amount of labelled data by training a classifier for FL in a target environment. In this approach, a pre-trained AlexNet model fine-tuned using a publicly available Lymphoma dataset is combined with another smaller sample of a private dataset of the FL class obtained from a new site [98]. This method has shown an increase the accuracy to up to 100% on the private dataset compared to the 58% accuracy achieved in the conventional method. This new approach requires a small amount of labelled training data from the class of interest to improve the performance on data from a new site. As such, the proposed solution reduces time, effort, and costs incurred for accurate diagnosis.

4.2   BACKGROUND

Lymphomas, the most common type of cancers, have more than 38 sub-types. The current cancer diagnosis procedure requires experts to closely study each sample for each cell to identify abnormalities. This process is time consuming and requires expert knowledge [101]. This raises the need for research in lymphoma identification with faster and accurate diagnosis [100]. The use of glass slide samples was a barrier for early research using image processing techniques in the pathology domain. Later, the introduction of WSIs bridged the gap between histology glass sample diagnosis and digital image processing.

Traditionally, image processing techniques have been used to extract hand-crafted, e.g., morphological, and textural, features. These are then combined with machine learning techniques, to support Lymphomas diagnosis [18] and to identify

abnormalities [102]. These techniques, however, require domain-specific knowledge from pathologists [100]. Also, the performance of traditional machine learning techniques heavily depends on the type of features they use [16]. The use of inappropriate features may lead to under-performing models [24].

Deep learning, which has been introduced to avoid the high dependency on feature engineering, is capable of learning the most suitable features from training data [36]. They have been recently applied to digital pathology [24, 103]. and have achieved promising results [103]. For instance, Janowczyk et al., (2016) utilise one model architecture for seven different tasks by only changing the patch selection technique [98]. Another research proposed an automated diagnosis for lymphoma using a deep learning model which achieves an overall of 95% prediction accuracy [104]. Deep learning models, however, are successful when a large amount of labelled data is available. As such, the performance on histology data is not as impressive as for natural image datasets, mainly due to the lack of accurately annotated training data. Consequently, techniques that require limited amount of labelled data are receiving increasing attention from the community [7]. TL is a popular method which requires a limited number of labelled data.

TL is capable of transferring knowledge from a source domain to a target domain and it leads to a decrease in the number of labelled data required from the target domain. In TL models, shallow features are generic features and deep features are specific to the domain [95]. The shallow features are not specific to the source domain or dataset. Thus, these features are applicable to different and new tasks and datasets. Features learned in the deep layers are specific to the task. A base network is first trained with source data and the weights of the network are transferred to train on a target dataset. The transferred features include the shallow features which are common to any type of image dataset [95]. Initialising networks with features of pre-trained networks improve the performance compared to random initialisation [96].

TL has been commonly applied as a feature extractor with fine-tuning pre-trained models. The ImageNet dataset has been commonly used as the source dataset in most TL based models [42]. The differences between the source domain and target domain affects the performance of TL. Therefore, identifying the differences between the

domains is important. Cheplygina et al., (2019) discuss TL in medical domain based on the difference or similarity between the source and target domains and tasks [15]. There are three possible ways that the domains and tasks can be similar or different as follows. First is the same domain and different task option in which multiple tasks for the same dataset is conducted. Second is the different domain and same task option, which resulted by differences in data acquisition procedures and the use of different scanners at different data sources. Third, is the different task and different domain scenario where a source task is used to pre-train a network, which can be as a feature extractor for a classifier, or for fine-tuning to the target task [15].

In the second combination as described before and presented by Cheplygina et al., (2019), different domains and same task resulted from different procedures followed at different sites raise the need to investigate ways of adapting the trained models to new sites [15]. Prior research conducted for medical imaging data have shown the effect to the DL performances for data from new sites. Segmentation for MRI data from different scanners which result in different data distribution due to appearances and populations in images have been studied [15]. In ultrasound images, the absence and presence of blood flow will be assessed based on the type of scan. The following research has focused on techniques that learn features from multiple sources [26]. However, these models generalise poorly to data from a new site that has not been included in the training [97]. Findings by Zech et al., (2018) from a research conducted on generalisability of Convolutional Neural Networks (CNN) shows that CNN performance on test data from sites used to train the model overestimates performance on another site [97]. In pathology, re-training models for new sites are infeasible since labelling data is an obstacle, which requires time and domain expertise. The effect of pre-processing techniques including normalisation have shown an improvement in generalisability [7]. It has been investigated in past research for models with multi-site data. These techniques are heuristic and specific to the domain [105].

Techniques to improve generalisability in models either increase the variability in data or minimise the differences between the source data and target data. Increasing the variability requires adding more data to the dataset which captures the variations in data. Data from different sites can be included to capture a higher variation [97]. In pre-processing approaches, increasing the variation can be achieved through

augmentation or by adding new data from the target domain. In order to minimise the differences between the data, Komura et al., (2018) use pre-processing techniques such as normalisation [7]. This process is specific to the task conducted and may not perform well for new data from new sites. In TL approaches, models aim at achieving optimal learning from a minimal amount of data from the target domain. The sameness or the difference between the source domain and target domain affects the performance. Achieving the capability of handling the variation in data from a minimal amount of data is beneficial to the research.

This chapter addresses the generalisation problem of deep neural networks for Lymphoma diagnosis. In particular, different approaches to improve generalisation were explored by using a small amount of labelled data for one class from a new data site. To minimise the amount of labelled training data required, the chapter also investigates the effects of pre-processing, specifically augmentation and histogram equalisation on the generalisation to new sites.
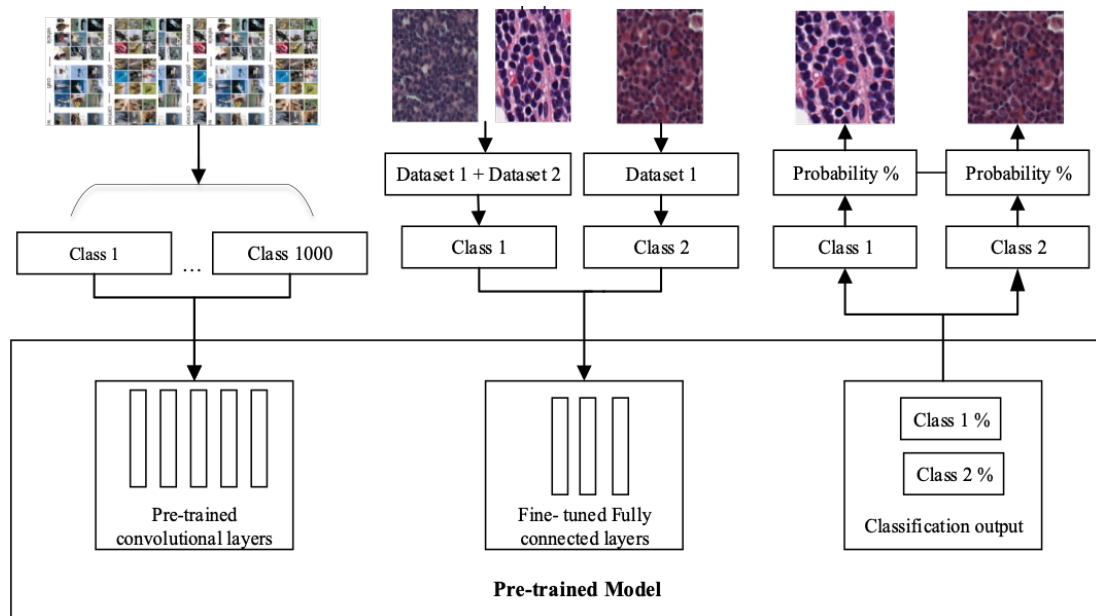
Figure 4.1 - Proposed model. The input to the pre-trained model is created by combining data from the class of interest from site 1 and site 2.

4.3   METHODOLOGY

The method is presented for binary classification using models trained from scratch and pre-trained models.

4.3.1 One Class Joint Classification

This chapter focuses on using a very small amount of data from the class of interest of a new site. The class of interest is usually the positive class in a binary classification problem. The data with cancer cells is the class of interest in a classifier differentiating normal cells and cancer cells. Figure 4.1 illustrates the proposed TL by fine-tuning method, which is to combine a small amount of data from the class of interest of a second site to the first site's class of interest. Five hundred image patches from the class of interest of site two are combined with site one's class of interest.

- *Training Pre-trained model*

The pre-trained model is trained under two options of domain and task combinations presented by Cheplygina et al., (2019) [15].

1. Different domains and different tasks

Different domains and different tasks option is followed when the source domain and the target domain are different, and the source task and the target task are different. The pre-trained model is experimented by using the ImageNet dataset [42], which is composed of natural images, and for the source dataset and it's fine-tuned to a target dataset from histopathology.

2. Different domains and the same task

Second, the different domains and the same task option is followed when the source task and the target task are the same, but the source domain and the target domain are different. According to Cheplygina et al., (2019) two domains are considered different when the data from the source and the target are provided by different sites, which leads to differences in the data distributions due to different slide preparation and scanning procedures [15]. Datasets from two different sites are used for the experiments. The second dataset is from a new site which consists of a small number of images from only the class of interest

- *Training models from scratch*

These models are first trained on data from the first site. Then, the models are re-trained with a dataset created by combining the first site's dataset with one class dataset from the new site.

4.3.2 Data Augmentation and Histogram Equalisation

Augmentation by adding variations to the data or by minimising the difference between the datasets are performed during pre-processing to improve performance. The proposed approach compares the effect of augmentation and histogram equalisation as steps to add variation and minimise the differences between the two sites. Augmentation is conducted by flipping and rotation to add variation. The colour contrast variations in the two datasets were minimised by performing histogram equalisation.

The outcomes of these pre-processing steps are compared with the outcome of models by training models with the combined dataset of the first site and the second site. This was to identify how the data variation from the one-class data of the second site affected the generalisation performance towards the testing data from the second site.

4.4   RESULTS AND DISCUSSION

4.4.1 Datasets

In this chapter, two datasets acquired from different sites were considered.

1. Dataset 1

Experiments were conducted based on a publicly available datasets for Lymphoma subtype classification [98]. The dataset created by National Institute on Aging (NIA) includes three subtypes of lymphoma: FL, Chronic Lymphocytic (CLL), and Mantle Cell Lymphoma (MCL). The images are Hematoxylin and Eosin (H&E) stained, and the data were gathered from multiple sites to add high staining variation to the dataset. The dataset consisted of 374 images of size 1388*1040. The number of images belonging to three classes were 113 for the CLL 139 for the FL and 122 for the MCL.

The images were split into non-overlapping patches of 227*227. This chapter focussed on classifying FL, the MCL and CLL classes were combined to generate the negative class which is defined as Non-FL class.

2. Dataset 2

The second dataset was provided by PathWest Laboratory Medicine WA. It included three H&E stained WSIs, scanned using the Aperio WSI scanner. Experienced pathologists labelled the images for the FL class, which was the class of interest. ROIs were extracted, and non-overlapping patches of 227*227 were created as it is the standard input size for the model (AlexNet) to be applied to. The patches that contained more background were eliminated from the dataset. Patch extraction was performed by using Distinct Block Processing from the blockproc function of Matlab's Image Processing Toolbox. A blind test set containing 213 image patches from this dataset had been created for the experiments. The models trained with dataset 1 were tested with the blind test set from dataset 2.

Since these datasets belonged to different sites, their distributions were significantly different. The variations resulted due to the slide preparation procedures and scanning equipment led to differences in the data. Further, different structural variations were also present in the two datasets (Figure 1.4) The histogram information from the two datasets showed the intensity ranges of the images vary. Figure 4.2 shows the first site's dataset and second site's dataset histogram for the RGB colour space. The first site dataset showed a high frequency compared to the second site. The differences in the datasets were caused by the differences in slide preparation and scanning procedures at the sources. This led to different distributions in the data. As such, deep learning techniques did not perform well for a blind test set from a new site.
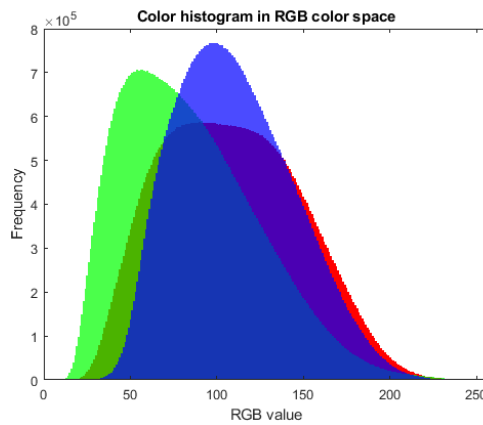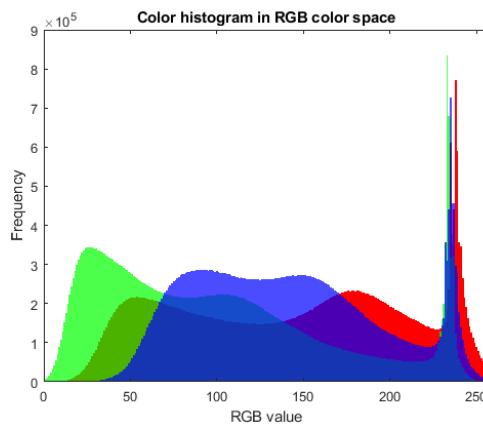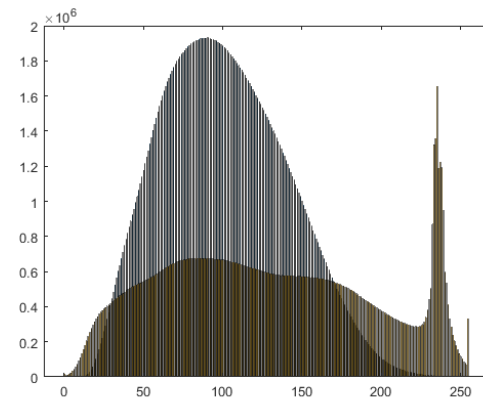
Figure 4.2 - Variations in the histograms for the data from the two sites. TOP: The histogram variation for both datasets. MIDDLE: The colour histogram in RGB colour space for the first site, BOTTOM: The colour histogram in RGB colour space for the second site.

Table 4.1 - The parameters for the trained models

	Model 1	Model 2	Model 3
Input size	227*227	227*227	227*227
Learning rate	$1*10^{-6}$	$1*10^{-6}$	$1*10^{-6}$
Dropout	0.5	0.5	0.5
Epochs	100	100	100
Batch size	128	15	128

4.4.2 Model Architectures

Three models were implemented. Their parameters are summarised in Table 4.1.

1. Model 1 Pre-trained model.

The AlexNet pre-trained model is trained on the ImageNet dataset consisting of 1000 classes of natural images [41]. The model was fine-tuned incrementally to identify the level of fine-tuning required for the dataset as suggested by Tajbakhsh et al., (2016) [41]. Five-fold cross-validation was conducted in all the experiments. Experiments for both different tasks and different domains approach and different domains and same task approach were conducted to understand the effect on generalisability to data from different sites.

2. Model 2 - Trained from scratch.

The model architecture of CIFAR AlexNet was applied to the dataset released in [16]. was used in the experiments. Five-fold cross-validation was used in all the experiments. Table 4.1 shows the model parameters.

3. Model 3 - Trained from scratch.

The performance of fine-tuned AlexNet pre-trained model was compared with the same architecture trained from scratch. Five-fold cross-validation was conducted in all experiments.

4.4.3 Experiments

The following experiments were conducted to study the generalisability of high performing DL models.

1. The pre-trained model was fine-tuned with 6 layers for the dataset from first site and the models to be trained from scratch were trained with data from the first site.

2. The experiments were conducted to identify the effect from data augmentation and histogram equalisation. The dataset from the first site was augmented to study the effect of augmentation.

3. Histogram equalisation was applied to the dataset from site 1.

4. Both augmentation and histogram equalisation were performed to the first-site dataset.

5. The three models were tested with a blind test set from the second site. The accuracies for the blind test were compared.

Further experiments to study the effect of combining data from the second site were conducted. Five hundred image patches from the class of interest from the second site were combined with the first site's class of interest. Then, the above listed experiments were conducted to the models by training with the combined dataset. The effect of joining a small amount of data from the class of interest with the effect of augmentation and histogram equalisation was studied. The study aimed to identify the possibility of limiting the required labelled data only from the class of interest.

Table 4.2 - Incremental fine-tuning with dataset1 to identify the optimal level of fine-tuning.

No of FC layers	Replace 1	Replace 2	Replace 3
Test set - Site 1	97.2	**97.4**	89.8
Blind test set - Site 2	35.0	**58.0**	54.0

Table 4.3 – Accuracy % for test sets from site 1 and blind test set from site 2.

Test set	Model (1) %	Model (2) %	Model (3) %
Test set - Site 1	97.4	**96.58**	92.02
Blind test set - Site 2	58.0	**67.0**	46.0

4.4.4 Fine-tuning level for transfer learning

The experiment was conducted to identify the optimal fine-tuning level required for pre-trained AlexNet to be applied in further experiments (Table 4.2). The experiment was conducted with a training on a dataset from the first site with five-fold cross validation. It showed that the performance increases when fine-tuning up to two fully connected layers but fine-tuning three or more fully connected layers showed a decline. Models learn the low-level features in convolutional layers from the ImageNet data, and high-level features were learned in the fully connected layers from dataset 1. The performances for the first site's data are evaluated using accuracy for a test set from the same site's data and a blind test set created from a new site (Table 4.3) This experiment followed fine-tuning by replacing layer by layer from the last fully connected layer upwards. For this model, the optimal level of fine-tuning was two fully-connected layers.

4.4.5 Comparing models from scratch and transfer learning

The experiments were conducted to identify the generalisability of the models. The models were trained and tested on Dataset 1 to compare the accuracy. The models achieved high accuracy for the dataset 1's test set. Compared to performances on Dataset 1, all models underperformed on predicting the unseen Dataset 2 from the new site. The dissimilarity in the datasets created a high generalisation gap to Dataset 2. Though the domain and task of both datasets were similar, differences in image acquisition tools and techniques affected the features, intensity, and proportion of data.

4.4.6 Effect of augmentation and histogram equalisation.

The effect of augmentation and histogram equalisation in the dataset from a new site were explored (Table 4.4 - Predictions for Dataset 2, Trained on Dataset 1). Predictions for the default state of dataset 1 had decreased with augmentation in some instances. Through augmentation, a model was given an increased number of samples with different views. This specialised the model to learn the training dataset better and results in overfitting, which led to a poor generalisation to the unseen dataset 2. Comparison of the datasets showed different variations in histograms due to differences in the acquisition procedures at the two sites. Thus, the effect of histogram equalisation on generalising the models were tested. The outcome showed a high increase in predictions. Both pre-trained and models from scratch showed an improvement by applying histogram equalisation, which minimised the gap between the datasets. This proved that a main cause for the dissimilarity in the datasets was due to intensity differences in procedures followed at different sources. Applying augmentation and histogram equalisation eliminated the negative effect of augmentation in some instances (models trained from scratch). The pre-trained AlexNet model did not show a negative effect from augmentation due to the transferred weights.

4.4.7 One class joint training

The aim of this experiment was to improve generalisation accuracy for a new site when a small number of images from the class of interest (Dataset 2) was joined with Dataset 1's class of interest (Table 4.4 - Trained on Dataset 1 and Dataset 2). The performance was studied comparing the effect of augmentation and histogram equalisation to Dataset 2, which was jointly trained with Dataset 1. Significant performance improvement was seen in adapting a model to data from a new site. The models resulted in high improvement in predictions recovering the initial generalisation gap in the range of 12% to 52%. Histogram equalisation on dataset 2 did not add a benefit compared to the improvement achieved by augmenting data. Dataset 2 benefitted by

Table 4.4 - The pre-trained AlexNet improved performance for dataset 2 after combining a small amount of data of the class of interest. The Table shows the improved accuracy for the blind test set of dataset 2

Predictions for site 2 (blind test set). Trained with site 1

Dataset1	Model (1)	Model (2)	Model (3)
Default	58.0	67.0	46.0
Augmented	48.0	62.0	48.0
Histogram equalise	85.0	96.0	99.5
Augmented and histogram equalise	88.0	93.0	85.0

Predictions for site 2. Combined Training with Site 1 and site 2

Dataset2	Model (1)	Model (2)	Model (3)
Default	100.0	97.0	100.0
Augmented	100.0	100.0	100.0
Histogram equalise	95.0	100.0	98.0
Augmented and histogram equalise	97.0	97.7	100.0

Table 4.5 - Effect of augmentation and histogram equalisation to predict blind test set from site two

98.1%	99.9%	100.0%
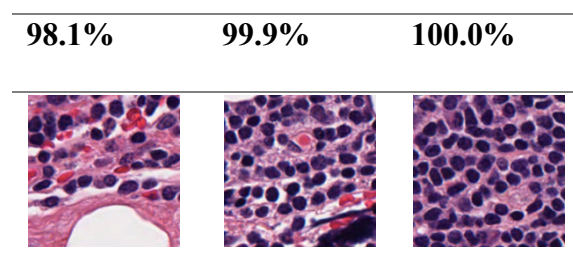
Table 4.6 - The table shows the time taken to fine tune pre-trained AlexNet model and AlexNet trained from scratch

	Pre- AlexNet	AlexNet trained from scratch
Time taken for training	51 min 11 sec	196 min 30 sec
Epochs	100	100
Hardware	GeForce GTX 1080 Ti	GeForce GTX 1080 Ti

augmenting small amount of FL data by adding a higher variance to the training. Both pre-trained and models trained from scratch could predict test images by only joining five hundred images of the class of interest, which minimised the differences between the datasets.

TL by fine-tuning was capable of training in less than half the time compared to the models from scratch, which would be beneficial when adapting a model. TL by fine-tuning had been trained by fine-tuning the last two fully connected layers with the joint dataset. Transferred weights from the ImageNet data and minimal training cost for fine-tuning results would be the best option for the generalised model for the new site. Table 4.6 shows examples of identification from Dataset 2. The proposed method proved that a small amount of labelled data of the class of interest would be sufficient from a new site for generalisation, which helps to improve the use of artificial intelligence in pathology.

4.5 CONCLUSION

In this chapter, a technique is proposed to utilise a small amount of labelled data from the class of interest at the new site to improve the Follicular Lymphoma identification at the new site. The differences caused by the sample preparation process and data acquisition procedure at new sites would normally cause the prediction accuracy to be affected at the new sites. This leads to the need for a large amount of labelled data from the new site to train better models. The complexities in the labelling procedure challenges the establishing an accurate model at the new sites. The proposed model in

this chapter would help to deal with the problem of needs for large amount of labelled data at the target sites. Experiment results showed that promising results could be obtained by combining only five hundred image patches from the class of interest of a new site.

Chapter 5: The use of Generative Adversarial Networks with Multi-site data in One-Class Follicular Lymphoma Classification

Recent advances in digital technologies have lowered the costs and improved the quality of digital pathology WSIs, thus opening the door to apply ML techniques to assist in cancer diagnosis. DL has produced impressive results in diverse image classification tasks in pathology, but the uptake of ML as a diagnostic tool in digital pathology workflow remains limited. A major obstacle is the insufficient labelled data for training neural networks and other classifiers, especially for new sites where models have not been established yet. This has multiple causes including the cost and time required for domain experts to manually segment regions of interest in WSI, in the relatively small archives of WSI available within many pathology institutions. There are other causes such as the limitations of generalising learning from public data sets of WSI that are often significantly larger than local WSI archives, the class imbalance problems that can arise with pathology data sets, and the differences in WSI from different sites due to variations in tissue staining and digital scanning techniques. Until these obstacles are overcome, it is unlikely that pathologists will be able to leverage the full potential of DL in their workflow. Recently, image synthesis from small, labelled datasets using Generative Adversarial Networks (GAN) has been used successfully to create high-performing classification models. Considering the domain shift and complexity in annotating data, in the present study, an approach based on GAN that minimised the differences in WSI between large public data archive sites and a much smaller data archive at the new sites was investigated. This approach allowed the tuning of a deep learning classification model for the class of interest to be improved using a small training set available at the new sites. The approach utilised GAN with the one-class classification concept to model the interest data, which minimised the need for large amounts of labelled data from the new site needed to train the network.

The GAN generated and synthesised one class WSI images were used to jointly train the classifier with WSIs available from the new sites. In this chapter, the proposed approach for follicular lymphoma data of a new site were tested by utilising the data archives from different sites. The synthetic images for the one class data generated from the data obtained from different sites with minimum amount of data from the new site resulted in a significant improvement of 15% for the area under the curve (AUC) for the new site that was needed to establish a new follicular lymphoma classifier. The test results have shown that the test site classifier could perform well by utilising GAN to generate synthetic data from existing data of test sites' archives without obtaining large training datasets.

5.1 INTRODUCTION

Cancer is the second leading cause of death worldwide. The World Health Organisation (WHO) reported 18 million worldwide new cancer cases in 2018 [8]. The reports for commonwealth nations stated a 35% increase of new cases between 2008 and 2018, with nearly 1.7 million deaths in 2018 [106]. Follicular lymphoma (FL) is the most common subtype taking up to 20% - 25% of non-Hodgkin lymphomas. It is crucial to diagnose FL early due to slow growth symptoms shown in later stages [107]. Medical imaging is an essential tool for diagnosis of cancer and cancer research [28, 108, 109].

DL has shown promising performance in pathology, for classification, segmentation, object detection and registration tasks using WSIs and offering opportunities to improve the efficiency and accuracy of pathology diagnosis [18, 108, 111, 112]. However, there are key differences between radiology and pathology images which challenge the translation of DL techniques from one domain to the other [14, 113]. These differences include but are not restricted to the following.

1. The large dimensions of WSI that requires them to be partitioned into large number of patches to use in classification, with partitioning required at both low and high magnifications due to the different features presented at different levels; this presents significant computational challenges.

2. The requirement for image-level as well as pixel-level labelling for grading and localisation in WSI data and for cell-wise classification via cell-level labelling to train DL models, which requires time-consuming and expensive involvement of domain experts [114]. This reflects the pyramidal nature of WSIs which comprise multiple levels with different dimensions containing different types of important features for diagnosis.

3. The significant variation in WSIs between sites are due to differences in tissue preparation steps such as sectioning, staining scanners as well as different scanning procedures [108, 113, 115, 116]. These variations cause many trained models to generalise poorly to data from new sites. Additionally establishing new models for different new sites which required a large amount of ground truth data from the sites [28, 108, 117].

4. The time and specialised domain knowledge required to label data in pathology, are more complex when compared to other medical image types [118]. Images often portray different types of diseases of non-standardised appearance representing large number of pathological abnormalities that require highly specialised domain knowledge, which is a challenge that pathologists can only deal with after years of specialised training [28, 53].

5. Requirement of methods to handle the paucity of large labelled pathology data sets and small data sets from new sites have not captured the wide variance in clinical samples [15, 22, 118].

To minimise the adverse effects of small data sets on network performance in pathology, transfer-learning (TL), weakly supervised classification and the use of synthetic data using Generative Adversarial Networks (GAN) have been used [15, 119]. These techniques address the numerically small size of datasets but still they must deal with variation in pathology data from different sites or hospitals due to the differences in imaging technologies and staining processes discussed above. This problem could be addressed by normalising data to minimise the difference in data distribution and/or increasing variation in local datasets [26, 69]. The combination of these approaches has shown to better improve classification [27, 120, 121].

In the normalisation approach, all the data are normalised into one style for training. Normalisation techniques differ between datasets, and suitable normalisation

techniques must be identified for each different dataset depending on the desired applications [7]. Three main types of methods have been used to handle stain differences; colour matching techniques that match colour to a reference template image, stain-separation methods that normalise each channel independently, and pure learning-based techniques including GAN that handle the problem as a stain transfer method. Learning-based methods reduce the drawbacks of colour matching, which can lead to improper colour mapping due to the use of the same transformation across the datasets, and are superior to stain-separation techniques, which do not consider spatial features of tissue [59]. However, learning as a style transfer technique is computationally expensive and this barrier has prompted a search for computationally simpler solutions to handle the stain differences in WSIs especially for different sites. The variance approach to small datasets attempt to increase the variance captured by the data set and modify the data distribution [26]. This could be achieved by adding more labelled data if sufficient domain experts are available to assign labels. If not, techniques such as TL and GAN that use minimal labelled data can be explored.

TL has shown promising results in pathology [122, 123]. In TL, models trained on a source dataset are adapted to a target dataset either by using the pre-trained model as a feature extractor or by fine-tuning the pre-trained model to the target dataset [15]. TL based methods for WSIs have proven to improve performance using a smaller dataset [22]. However, the investigations of the impact of TL used for new sites with limited labelled data and the impact of using data from multiple sites are not well reported. Prior research presents an approach to fine-tune a pre-trained model using data from the class of interest of a new site trained with a dataset from another site. This reduced the need for labelled data from the new site [122]. However, the approach reported in the paper [31] would lead to overfitting at if insufficient data are available for training. If limited labelled data are available and the training models consist of millions of parameters, steps should be taken to perform thorough evaluations with testing data which captures the data distribution. Therefore, this chapter investigates the possible use of GAN to generate more data for the new site.

Recent research on using GAN for WSI has shown the value of generating realistic synthetic data to increase the labelled data for classification [118, 119]. Where a class imbalance exists, the one-class classification approach can be used [124], and this

focuses on developing non-target data in order to perform binary classification. One-class classification has been discussed in medical image classification and has shown promising results in multiple domain areas, although relatively limited research has focussed on histology image processing [124, 125].

Apparently, GAN for one-class image classification for WSIs has limited investigations. Therefore, the chapter investigates the GANs influence in the one-class classification tasks by creating synthetic data for the labelled one-class data in the private dataset. To classify one-class new site data, the multi-site dataset's negative class is passed down as non-target data for classification. The use of limited one-class data from new sites with GAN significantly contributes to reducing the differences in data distributions of different sites and the resulting generalisation problem. The performance of the classification showed promising improvement in the generalisation by using one-class data without needing to retrain models for the new site.

5.2 GANS FOR DIGITAL PATHOLOGY

GANs have attracted much attention recently and have been used in the medical imaging domain [118]. GANs have also been applied to be used for WSIs, in the areas of augmentation, segmentation, virtual staining, stain normalisation and stain style transfer [54, 59, 119].

The most common application in pathology is to eliminate the stain differences in WSI data, and there is a growing interest in using GAN to produce synthetic images to increase the amount of data to train the DL models [56, 118, 119]. This application addresses a specific issue in pathology, which is that images with small and large amounts of positive features will both be classified as positive by a pathologist, in contrast to general domains which have distinct classes. A method based on CycleGAN has been used to augment positive samples by translating easy-to-classify samples into hard-to-classify samples [26, 69]. GAN as an image translation method has been proposed with a Conditional Generative Adversarial Networks (CGANs) for histopathological to immunofluorescent image translation [126]. Preliminary investigations show that GANs could be used to handle inter-site differences in WSIs was developed in which discriminative knowledge from a source domain was

effectively transferred using a Siamese network [127]. The study investigated colour normalisation and adversarial training to adapt knowledge from the source domain to the target domain, with significant improvement. However, the authors also mention the drawback of two-step training, the effects of using higher number of samples and different complexities of models. The method also depends on the reference images used and the normalisation techniques applied to the images. Improving generalisation through staining invariant features is another approach to improve classification using CNNs [53]. Colour normalisation and colour augmentation have been investigated to address the inter-site differences in WSIs. Furthermore, in the instance when there is only one-class WSIs available, which are from the class of interest, the classification is challenged [66].

An additional problem experienced across imaging domains is the class imbalance that can arise due to numerical imbalance between the positive class (e.g., cancer) and negative class (not cancer). Cancer in medical data is often the minority class due to various factors including the relative paucity of cancer tissue compared to normal background tissue, the complexity of labelling small regions of cancer (often single cells) and the lack of openness of medical domains. Auto-labelling techniques are preferred as a method to handle the cost of the labelling problem. However, in unsupervised techniques, there is no constraint on the boundaries of the clusters, which may fail to provide the accurate segmentation of regions of interest at the pixel level required to develop models [128].

One-class classification has been applied as a learning-based technique using positive and unlabelled data, a novelty and outlier detection technique and a one class support vector machine (SVM) based technique [66, 124]. The limitations of a one class approach has been studied [129]. This includes the tendency of pathologists to label images at the whole-image level regardless of how much cancer is present in the image, whereas in natural image domains the images usually have a distinct label [26, 69]. Therefore, annotations in the images of class of interest are important to train models to assist diagnosis. Much research of deep learning-based techniques and many shallow learning techniques explores the novelty/outlier detection technique. This method focuses on artificially created outliers for binary classification along with the

labelled positive data. The target dataset given to the models should capture the high variability in the distribution to support classification with the artificially created class. Inspired by the prior research for the one-class data problem in other domains, this chapter uses the one-class data as a solution to minimise the required amount of labelled data. Models trained only with a small one-class dataset from a new site decrease the need for a large number of labelled data from the new site [122]. In WSIs, the class of interest has minimal labelled data. Due to lack of generalisation, it is not feasible to directly transfer a model trained with one site's data to another site's data. Therefore, to improve the performance of models, we suggest that the one-class dataset can consist of small amount of data from the class of interest. Limited research focuses on using one-class data to handle the lack of labelled data while handling WSIs from different sites. The chapter introduces a solution to address the limitations in labels and differences in data from different sites in WSIs.

5.3   METHODOLOGY

5.3.1 Overview of the Proposed Structure

Figure 5.1 presents the overview of the proposed architecture to learn from the new-site's WSIs. Figure 5.2 shows the combination of one-class data from a new site and multi-site data. The proposed architecture supports:

- The use of one-class data and GAN to effectively minimise the need for labelled data from new sites.
- Minimising the distribution difference on the WSIs from multiple sites and the new sites.

In the proposed approach, a GAN is used to create new data points for synthetic one class WSIs patches. These synthetic patches are generated for the new Site (S_2) which has minimal number of labelled WSIs. The Classifier combines WSIs' patches from n sites (MS_1) and the new site S_2 which improves the differences of the distribution of WSIs. The GAN generates synthetic data for the new site. Therefore, the classifier is more generalised to classify WSIs from different sites.

5.3.2 Using GAN to Increase the Number of Patches for New Site

GANs are based on two CNNs trained as a generator and a discriminator. The network for the generator learns the distribution of the real images to generate new data points, on which synthetic images are belonging to the real data distribution. The generator's aim is to maximise the capability of creating realistic images to trick the discriminator while the discriminator aims to maximise the capability of differentiating the synthetic data from the real data.

In the proposed approach a GAN is used to create synthetic WSI patches for the new site, S_2 which has a limited number of labelled WSIs for the class of interest. The synthetic WSI patches create additional training data for the new site, S_2. This minimises the differences in sites and improves the classification model. The chapter examines the effects of combining WSIs from a new site, with the use of synthetic data generated based on the Deep Convolutional Generative Adversarial Networks (DCGAN) for the small one-class data from the new site, S_2. The DCGAN consist of a Convolutional Neural Network (CNN) combined with the traditional GAN to achieve a deep feature-based representation of data. The architecture of the DCGAN is capable of generating better quality images with stable training compared to the traditional GAN [130].

The DCGAN is also a model which does not require a high computational power [41]. The DCGAN consists of a generator G, and a discriminator D. The input to G is a vector of 100 elements with a random normal distribution. The generator will learn and create data for the target dataset which is the one-class dataset (G(x) =x). The discriminator's aim is to differentiate the generated fake image (x_f) and real target image (x_t) which were the inputs to the discriminator. The generator and the discriminator are learned adversarial in a minmax game, in which the discriminator's objective is to maximise the ability to differentiate between fake *(G(x)= x_f)* and *(x_t)*, while the generators objective is to create target like synthetic images *(x_s)*. The Equation 5 is used by the generator and discriminator. The DCGAN constructed is used to create *S_2* patches and passed on to the classifier. The model based on the DCGAN architecture is illustrated in Figure 5.1 which is based on [131]. The generator of the DCGAN with the input of 100 element vector outputs a synthetic image after propagating through the model. The output generated is a 64x64x3 image. The network

consists of a fully connected layer and four deep convolutional layers. Batch normalisation layers and Relu layers are applied to the convolutional layers of the network.

The discriminator network of the DCGAN has a CNN architecture with an input of 64x64x3. The synthetic images from the generator and the target data which are real images are learned in order to differentiate the real and fake images. The network of the discriminator consists of four deep convolutional layers and a fully connected layer. The network consists of batch normalisation layers and Relu layers in the convolutional layers. The GAN was trained for 100 epochs.

$$[min_G max_D V(D,G) = E_{xp_{data(x)}}[logD(x)] + \quad\quad 5$$
$$E_{z\,p(z)})]]$$

5.4   ONE-CLASS CLASSIFICATION

As shown in Figure 5.2 the new site, S_2 with its limited WSI patches contain only one class and therefore, a one-class classifier is applied for classification. However, the one-class classifier receives WSI patches from MS_1 (contains patches from the target class and non-target class), S_2 (contains only target class patches) and S_2 synthetic data generated from the DCGAN Figure 5.1. In order to handle the one class problem, the non-target class is taken from MS_1 non-target class. Therefore, the final classifications target class comprises of WSI patches from MS_1, S2 and S2 synthetic and the non-target class comprises only of S1's non-target class. The target and non-target WSI patches are classified using as CNN. The CNN comprises of three layers with sigmoid activation function and the final layer comprises of a softmax activation. The CNN takes an input of 64x64x3. A dropout of 0.5 is applied to all the layers. The CNN optimiser is RMS-prop, and the loss is binary classifier.
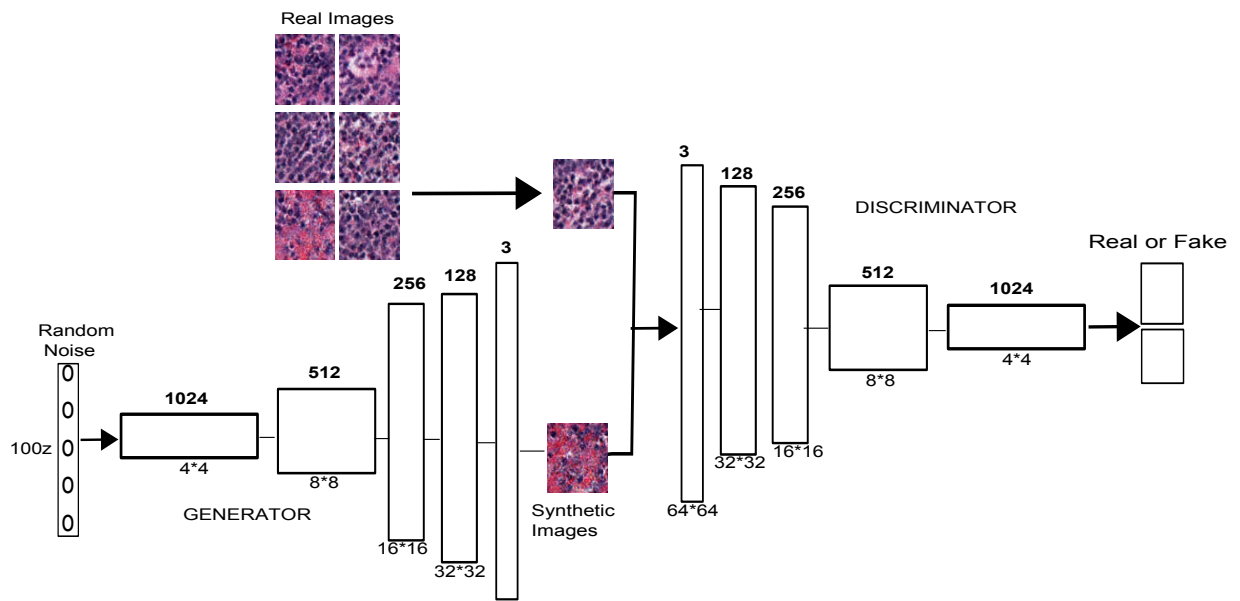
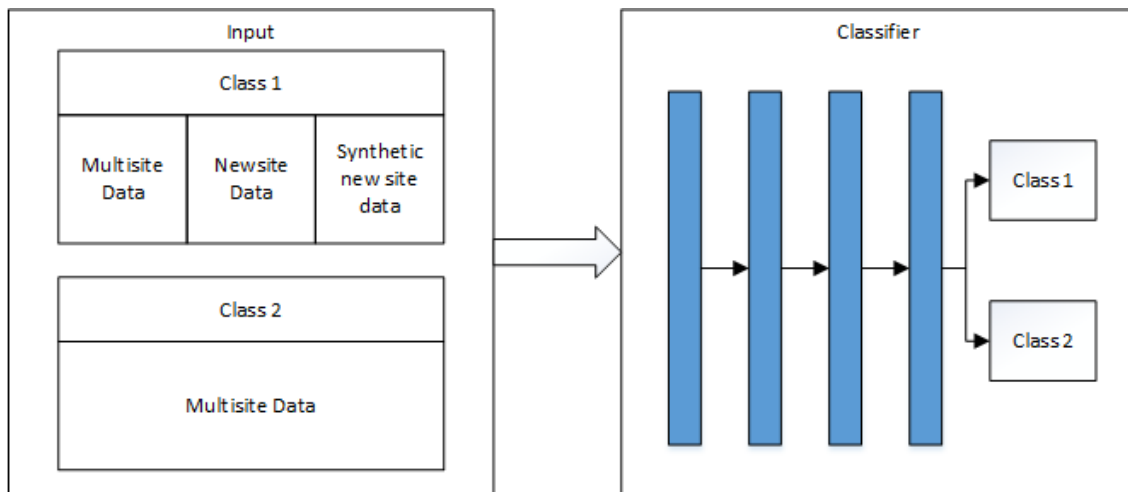Figure 5.1 - The block diagram architecture for the overall proposed model.

Figure 5.2 - The block diagram for input data

5.5 EXPERIMENTS

5.5.1 Datasets

The experiments were conducted based on a publicly available multi-site dataset for Lymphoma subtype classification and a new site class dataset which contain data for the class of interest. The publicly available dataset has been created by National Institute on Aging (NIA). This dataset includes three subtypes of lymphoma (Follicular Lymphoma (FL), Chronic Lymphocytic (CLL), and Mantle Cell Lymphoma (MCL)). The Hematoxylin and Eosin (H&E) stained images were gathered from multiple sites to add high staining variation. Furthermore, the dataset consists of 374 images of 1388x1040 dimensions. Each class has images as follows, 113 for the CLL 139 for the FL and 122 for the MCL. In order to conduct the experiments, the images were split into non-overlapping patches of 64x64. The chapter focuses on binary classification and therefore the CLL and MCL classes were considered as the non-FL non-target class, while the FL class was the class of interest.

The second dataset was provided by PathWest Laboratory Medicine WA. This private dataset is considered as the dataset from the second site. It includes three H\&E stained WSIs, scanned using the Aperio WSI scanner. Experienced pathologists have labelled the images for the FL class, which is the class of interest. The Regions of Interest (ROI) were extracted based on the coordinates of the annotations, and non-overlapping patches of 64x64 were created. The patches that contained more background were eliminated from the dataset. Patch extraction was performed by using Distinct Block Processing from the blockproc function of Matlab's Image Processing Toolbox. A blind test set from this dataset has been created for the experiments. The models trained with the public dataset (multi-site) were tested with this blind test set from the private dataset (new site).

5.5.2 Experiments

The differences in the data from the public multi-site dataset and the private new site dataset were explored. A visual comparison of the differences in data could be identified based on Figure 5.3. Additionally, t-Distributed Stochastic Neighbour Embedding (t-SNE) plots the two datasets which were created to understand the data distribution and differences in the datasets. Identifying the distribution and the

differences is important to address the possible generalisation problems using the proposed approach. The t-SNE plots are suited for visualisation of high-dimensional datasets. Based on the generated t-SNE plot, for the two datasets it was possible to identify a significant difference in the data distributions of the class of interest from different sites. Based on the findings of the differences in the data distributions, detailed experiments were conducted in order to validate the proposed method.
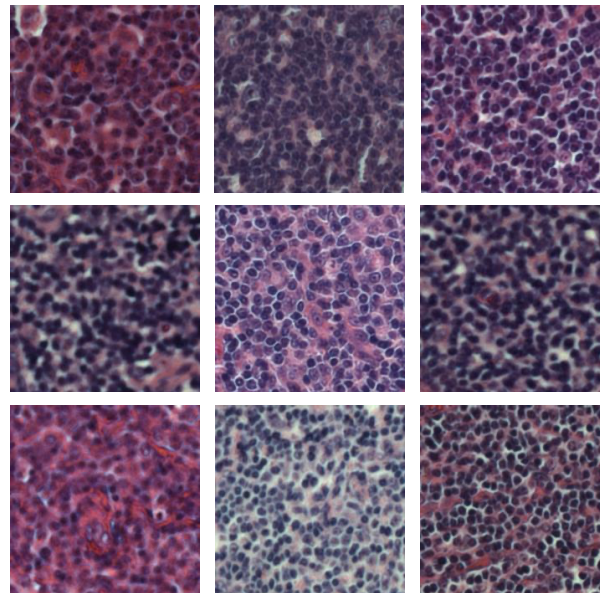
Data from the public source
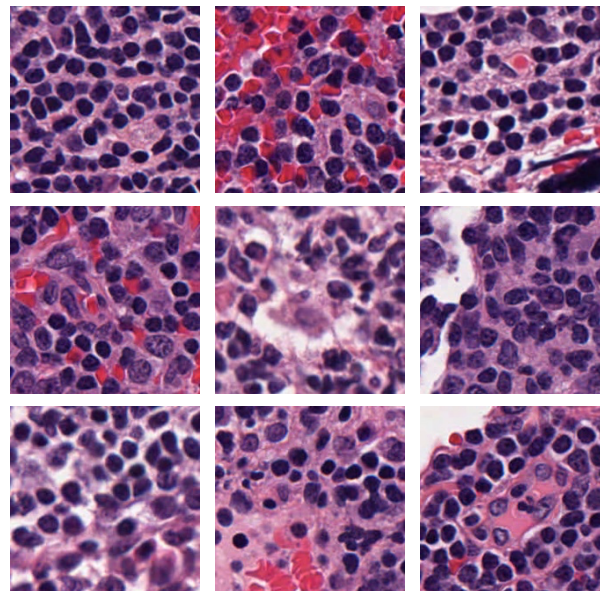
Data from the multi-site data set.

Figure 5.3 - Examples of data from different sites. Figures show the contrast difference and structural variability in the datasets from different sites
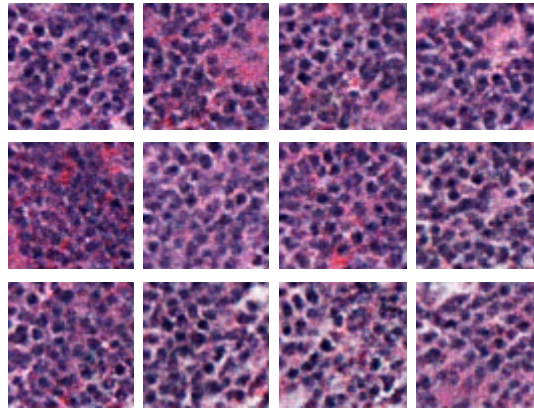
Figure 5.4 - Examples of synthetic data generated using DCGAN for the new one class site

The approach of using synthetic data from GAN to validate the effectiveness of the proposed method was necessary. The proposed method was compared with synthetic data and without synthetic data for a fair comparison.

The experiment without synthetic data was conducted by creating the classifier with data from the first sites. The classifiers were trained, and performance of the classification was obtained. The experiment with synthetic data has two phases. The first phase was to create synthetic data for the one class data. Figure 5.4 provides examples of generated synthetic data using the DCGAN for the new site. The synthetic data were combined with the dataset of first site as the input to the classifier. The proposed approach develops GAN as a method to handle differences in the data from different sites. In order to build the negative class of the multi-site data are considered as the negative class for the one-class dataset. The classifiers were trained with the joint dataset and performance was evaluated comparison with the experiment without any synthetic data.
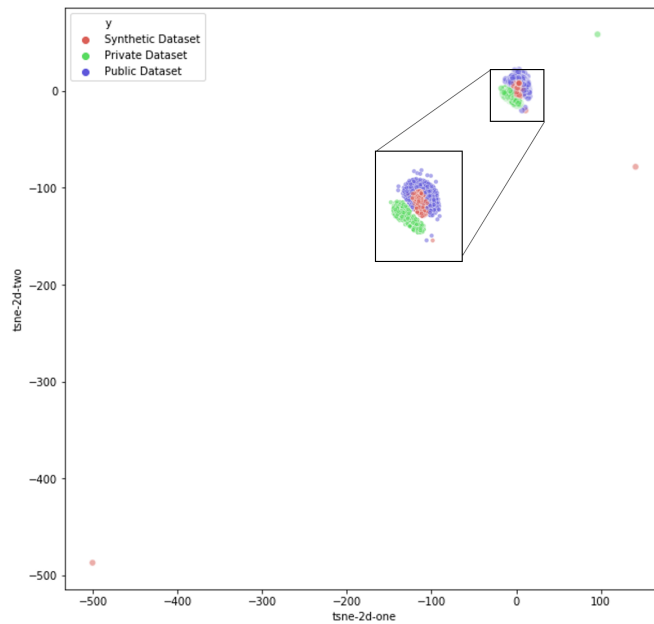
Figure 5.5 - The 2-dimensional t-distributed stochastic neighbour embedded (t-SNE) plot for the class of interest from different sites and synthetic data. Each point represents values of each image patch of WSIs

The experiments were conducted to small scale images though there were successful prior research for images with higher dimensions (256*256, 128*128). The models at relatively small scales for computational ease and prior research showed that still low-resolution images with large number of cells needed improvement compared to cell images [42]. After generating synthetic data for the private dataset, a t-SNE plot was generated in order to identify the changes in the data distributions. Figure 5.5 demonstrates that the synthetic data generated by the GAN had contributed to merge the gap between the distributions of the multi-site dataset and the new site dataset.

5.6   RESULTS

The experiments with synthetic data to classify one class data from the new site were conducted, and the performance was calculated for evaluation. The multi-site dataset and the synthetic data for the one class private dataset were used as the input to the classifier. The classifier was trained with the synthetic data to align the differences between the datasets.

The results showed better performance by using synthetic data to minimise the differences and improve performance for the data from new site. The classifier acts as a method to handle the one class data problem by using the multi-site dataset's negative class. The results showed an improvement to both the new site dataset's and multi-site dataset's performance. Table 5.1 shows the performance without synthetic data and with synthetic data for the validation data of the multi-site dataset. It indicated an 8% increase in Accuracy and Area Under Curve (AUC) values. Table 5.1 also presents the improvements to the F1-score, Precision and Recall performing a fair comparison of the performance for the one-class dataset.

5.6.1 Comparison of Classification

Differences in the new site data caused poor generalisation in trained models which could be seen in Table 5.2 (1st column). The AUC 89.75 for the same site validation set but AUC was 60.66 for the new site test set. In order to conduct a fair comparison, the evaluation metrics for the F1-score, precision and recall were derived, which reflected the performance for the two classes.

The proposed approach utilised a small amount of one-class data from a new site to improve the learned features. Table 5.1 shows that the validation set from the same site also had a performance improvement. Table 5.2 shows the improvement of the test set from the unseen new site.

5.6.2 Using GAN to Handle Differences

In order to handle the differences in the data from new sites, the proposed method takes a different approach using GAN. The synthetic data generated by the GAN was used to increase the amount of one class target data while increasing the variation in the data distribution with a limited amount of labelled data. The enriched data distribution aligns the features in the multiple datasets, which leads to minimising the gap between the datasets. Instead of data normalisation or colour matching, this learning approach was much more beneficial [59]. The other learning-based approaches' model

complexity is a limitation and applying synthetic data as an approach to align the data was much more efficient.

Table 5.1 - Validation set - The model evaluation for one class classification.

	Without Synthetic	With Synthetic
Precision FL	86	**93**
Precision Non-FL	79	**87**
Recall FL	76	**86**
Recall Non-FL	88	**94**
F1-Score FL	81	**89**
F1-Score Non-FL	83	**90**
Accuracy	82	**90**
AUC	82	**89.75**

Table 5.2 - Blind test set - The model evaluation for one class classification

	Without Synthetic	With Synthetic
Precision FL	57	**73**
Precision Non-FL	64	**78**
Recall FL	58	**74**
Recall Non-FL	63	**77**
F1-Score FL	57	**73**
F1-Score Non-FL	64	**77**
Accuracy	61	**75**
AUC	60.66	**75.33**

Chapter 5: The use of Generative Adversarial Networks with Multi-site data in One-Class Follicular Lymphoma Classification

5.7   DISCUSSION

The evidence based on the prior research showed the challenges of WSIs to apply DL techniques. Despite the promising performances in DL application to WSIs, the domain shift caused by the WSIs from completely new sites resulted in a lack of generalisation in models. The barrier of differences could be approached by obtaining large amounts of labelled data from new sites. However, it would be an infeasible approach considering the complex tasks of labelling WSIs and developing specific DL models to cater differences of each site. Thus, there is a need to handle the domain shift in WSIs using limited labelled data. Therefore, inspired by the one-class classification techniques, the approach of one-class data classification using GAN is proposed to handle inter-site differences in WSI data. The method could be used for new sites by only using a small one-class data set of the class of interest.

In the present study a larger dataset from a publicly available source and a small amount of one class data from a new site were incorporated. We provided a comparison of performance for the data from the new site with and without the synthetic data. By using a sufficient number of labelled data from a multi-site dataset and a small number of synthetic data, the accuracy of the CNN could reach an appreciable 15% increase in classifying FL and non-FL data. The method could minimise the need for a large amount of labelled data from the new site and handle differences. This was achieved without compensation for the image variations and non-morphological differences in data from different sites. Figure 5.5 demonstrates the differences by separating the data from the public domain and the private domain into two clusters. It demonstrates that the differences were minimised, after generating synthetic data for the new site. It demonstrates the change to the t-sne plot's probability distribution based on the neighbouring points from the synthetic data [132]. Furthermore, the performance evaluation for the experiments significantly presents that the synthetic data for the one-class data from the new site supported to increase the performance for the unseen new dataset and the validation data of the multi-site data.

Based on the findings discussed in relation to Table 5.1 and 5.2, further experiments were conducted to explore the effect of the amount of synthetic data.

Table 5.3 - The classification outcome for validation data explored with different amounts of synthetic data

	600 synthetic images	400 synthetic images	100 synthetic images	50 synthetic images
Precision FL	93	75	80	60
Precision Non-FL	87	90	96	96
Recall FL	86	92	97	98
Recall Non-FL	94	70	76	33
F1-Score FL	89	83	88	74
F1-Score Non-FL	90	79	85	49
Accuracy	90	81	86	66
AUC	89.75	81	86	65
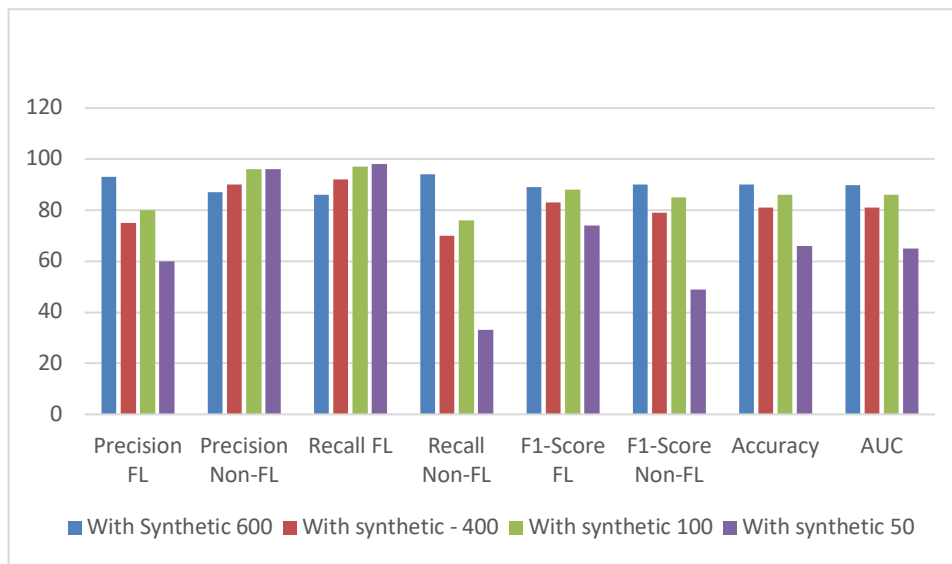
Figure 5.6 - Model performances for validation data using synthetic data ranging from 50 images to 600 images.

Table 5.4 - The classification outcome for blind set data explored with different amounts of synthetic data

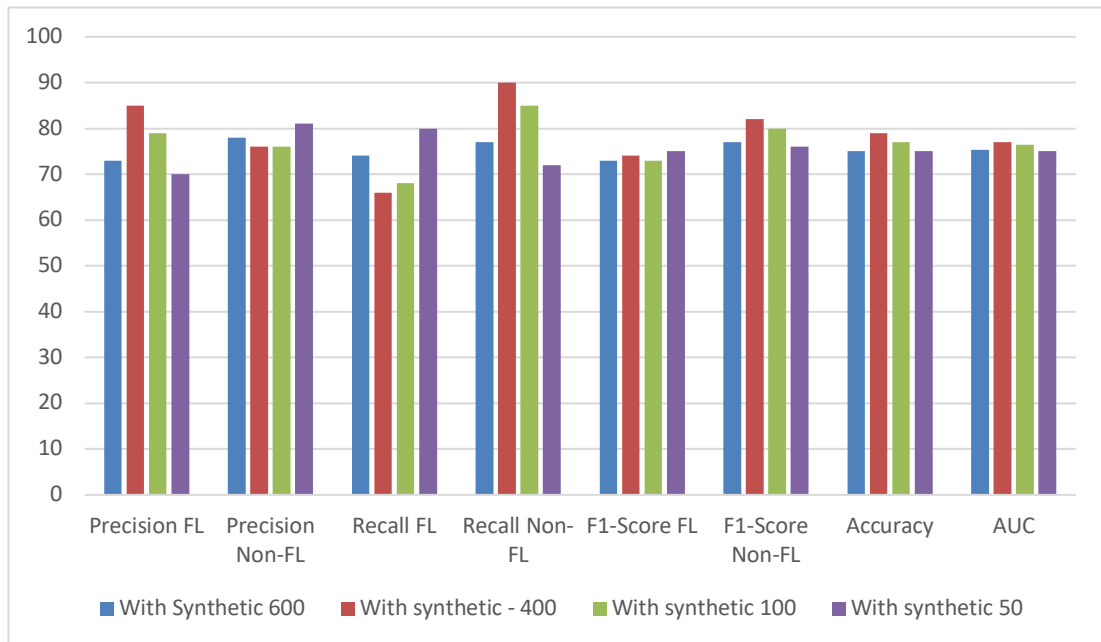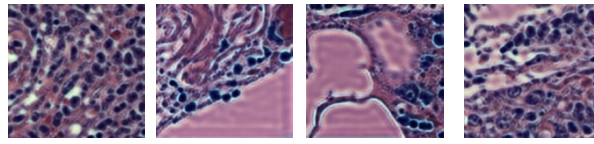|	With Synthetic 600	With synthetic - 400	With synthetic 100	With synthetic 50
Precision FL	73	85	79	70
Precision Non-FL	78	76	76	81
Recall FL	74	66	68	80
Recall Non-FL	77	90	85	72
F1-Score FL	73	74	73	75
F1-Score Non-FL	77	82	80	76
Accuracy	75	79	77	75
AUC	75	77	76	75

Figure 5.7 - Model performances for blind test data using synthetic data ranging from 50 images to 600 images.

Synthetic data for the multisite dataset.
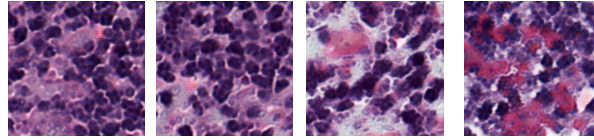
Synthetic data for the new site's dataset.

Figure 5.8 - Examples of synthetic data generated using CycleGAN for multi-site and new site data

The Table 5.3 shows the model evaluations for classification using 600, 400, 100 and 50 synthetic images. The Figure 5.4 shows the table 5.3 in a bar chart where the performance of the classifiers is fluctuating precision for non-FL and Recall for FL are increased compared to decreased amount of synthetic data. However, the Precision for FL, recall for FL, F1 scores, accuracy and AUC are decreasing as the number of synthetic data are decreased. The performance for the blind test shows a different pattern where Recall for non-FL and F1 score for non-FL and accuracy increased at 400 and 100 synthetic images. However, recall for FL, Precision for non-FL have shown a decrease. Therefore, the classifiers show unstable performances as synthetic data numbers are reduced.

Furthermore, the proposed approach is a computationally, less demanding approach than the commonly used GAN architectures. The proposed approach was tested on a NVIDIA 1080i 1 GPU. The most common GAN types, StarGAN, StyleGAN and CycleGANs require a higher GPU power. In general the starGAN requires a NVIDIA Titan Xp GPU with 4000 training images [133]. StyleGAN recommends using NVIDIA DGX-1 with 8 Tesla V100 GPUs and training for a week [58]. Although CycleGAN could use the NVIDIA 1080i GPU, it takes approximately 72 hours to run 100 epochs, whereas the proposed approach generates images by running 100 epochs in less than one hour. Figure 5.8 shows the images generated by CycleGAN after 10 epochs running for 3 hours. It shows image translation converting multisite images to new site images which do not capture the heterogeneous tissue structures, and it fails

in translating private images to public images which have heterogeneous features in the tissue structures. Previous research on CycleGAN also showed that complex texture and shape structures were not captured by the CycleGAN [134]. Adapting a model to capture real-world data distribution is essential. Therefore, the proposed approach is an efficient method compared to the other methods due to its capability of addressing the unique features from the new site using limited computational power. Therefore, the proposed GAN based approach is a faster method which can generate synthetic data with a limited number of labelled data from a new site.

Taken together, these results suggest that there is an improvement in the classification performance of deep learning models by using synthetic data generated by a GAN for a small one class dataset from a new class. It is a promising approach to handle differences in data from new sites and align the data to improve classifiers performance for the unseen data from a new site.

5.8 CONCLUSION

Chapter 5 presents a different approach to handle inter-site differences while minimising the need for labelled data. The chapter presents a GAN-based technique to approach the one-class data problem as a solution to minimise the number of required labels while improving the performance for data from new sites. Based on the empirical evaluation, forwarding data for the non-target class from a different site to be jointly trained with synthetic data and data from the new site for the class of interest shows significant improvement in performance for the new site's data. The proposed technique could be applied to any new site with a minimal number of labelled data for the class of interest. The research discovered a different approach of applying GAN to handle the differences in data from new sites.

Chapter 6: **Conclusions**

Efforts were made to identify factors influencing the translation of AI techniques to the digital pathology workflow. The thesis focused on co-analysing multi-stained WSIs, high variation of inter-site WSIs and labelling for more generalised outcomes from AI. The prior research has shown the limitations in AI techniques which affect the successful translation of them into the pathology workflow. These issues affected the performance of AI techniques especially in machine learning and deep learning techniques. The complexity of WSI data itself challenge the application of AI techniques to pathology. Therefore, highlighted the need to investigate the factors influencing translating AI. Prior research has attempted to develop techniques to address the factors influencing translating AI in digital pathology. However, research is limited in investigating challenges and methods specifically in translating AI to pathology.

Based on findings in the literature, DL based approaches for improving unsupervised non-rigid WSI registration by super-convergence, improving inter-site data predictions via transfer learning using class of interest, and handling lack of data from different sites using class of interest WSIs to generate synthetic WSI data using GAN for classification were investigated.

Contributions of the Thesis

The findings from this research provide contributions to the field of AI in digital pathology. The thesis focused on identifying factors influencing the translation of AI in digital pathology.

The contributions of the present thesis are as follows,

1. The simplified unsupervised non-rigid registration using deep features and fast convergence.

When multi-stained WSIs are created to co-analyse as a sub step of the diagnosis process, the WSIs result in large non-rigid transformations. The complex annotation process limits the applicability of supervised methods and long training times requiring

high computational memory are a drawback in DL methods which increase the translation gap of the methods. Therefore, the thesis focused on a solution based on training a simpler DL model for feature extraction with super-convergence which resulted in faster training and improved registration performance, thus assisting the co-analysis process and minimising the translation gap of AI in digital pathology.

2. Transfer learning-based method to handle inter-site differences using limited data from a new site

A trained DL model should be able to perform similar when predicting for WSIs of the same cancer type from a different site. However, trained DL models do not perform well for unseen WSIs from different sites. This is due to the pervasive variability in WSIs. Therefore, the second contribution of the thesis investigated the differences of WSIs from different sites to handle the affect from inter-site differences to DL models. The solution presented focused on fine tuning a pre-trained model with data from a new site joined with a publicly available dataset. The results showed improvements to the new site only by combining five-hundred image patches from a new site. Therefore, minimising the performance gap of the DL model.

3. GAN based method to reduce need for labelled data from new sites using limited data

The third contribution of the thesis focused on limitations in acquiring sufficient amount of labelled data to train DL models. Complex annotation process limits the ability to have large amount of data from multiple sites. Therefore, in this thesis the possibility of using synthetic data as an approach to handle lack of data from new sites was investigated. Furthermore, to minimise the computational complexities, a simpler DCGAN method was used to generate synthetic data for a new site. The results showed that training a classifier with publicly available data and synthetic data from the new site showed an improvement in predicting for the new site.

Limitations

Model execution depends on hardware availability including GPUs. The speed of model execution largely depends on the available computational power. The chapter 3, 4, and 5 trains DL and GAN models that require a single GPU.

Proposed methods' requires annotated data to train the models. Therefore, the models rely on labelled training data. The training data allows the models to learn and produces better results. However, the proposed methods require only a small amount of annotated data which allows the models to be versatile in real-world application with limited annotated training data.

Unavailability of large datasets capturing the true data distribution of clinical data is a limitation in DL for digital pathology research. A challenge mentioned in translational research is the importance of model testing under different variations of data. Analysis of the clinical data distribution and testing will provide models proposed in this thesis to further improve to meet the need of clinical workflow.

Future Work

Further investigations of the factors translating AI in digital pathology needs to be conducted. The lack of transparency and interpretability is one such important factor to be addressed. The prior literature had identified it under analytical factors which focus on management of tasks between pathologists and AI techniques. This requires handling tasks allocated to AI and pathologists to work collaboratively. At that stage interpretability of AI models is an important factor to be handled.

Furthermore, the factors which the present thesis contributed could be further enhanced to further minimise the gap of translating AI in digital pathology.

1. The unsupervised non-rigid registration technique to handle multi stain registration has been improved using super-convergence. This method is improved using learning rates and weight decay. However, when using the super-convergence, the learning rate, batch, momentum, and weight decay can be optimised to reach a super-convergence. Therefore, it is important to investigate other optimisation options and extend the findings.

2. The limitations in trained deep learning models' capability to generalise to data from new sites is a barrier and using pre-trained networks for one class data improved the performances for the new site WSIs. The pre-trained models can be further improved with fusion techniques to address new site's WSIs.

3. Use of GAN to handle lack of data for classification is proven to perform well. However, the dimensionality of WSIs and computationally expensive GAN models can be explored to further enhance the efficiency of synthetic data generation.

Bibliography

1.   Griffin, J. and D. Treanor, *Digital pathology in clinical use: where are we now and what is holding us back?* Histopathology, 2017. **70**(1): p. 134-145.
2.   Niazi, M.K.K., A.V. Parwani, and M.N. Gurcan, *Digital pathology and artificial intelligence.* The lancet oncology, 2019. **20**(5): p. e253-e261.
3.   Bera, K., et al., *Artificial intelligence in digital pathology—new tools for diagnosis and precision oncology.* Nature reviews Clinical oncology, 2019. **16**(11): p. 703-715.
4.   Lujan, G., et al., *Dissecting the business case for adoption and implementation of digital pathology: A white paper from the digital pathology association.* Journal of Pathology Informatics, 2021. **12**(1): p. 17.
5.   He, L., et al., *Histology image analysis for carcinoma detection and grading.* Computer methods and programs in biomedicine, 2012. **107**(3): p. 538-556.
6.   Pantanowitz, L., *Digital images and the future of digital pathology.* Journal of pathology informatics, 2010. **1**.
7.   Komura, D. and S. Ishikawa, *Machine learning methods for histopathological image analysis.* Computational and structural biotechnology journal, 2018. **16**: p. 34-42.
8.   Ferlay, J., et al., *Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods.* International journal of cancer, 2019. **144**(8): p. 1941-1953.
9.   Sun, C., et al. *Revisiting unreasonable effectiveness of data in deep learning era.* in *Proceedings of the IEEE international conference on computer vision.* 2017.
10.   Xing, F. and L. Yang, *Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: a comprehensive review.* IEEE reviews in biomedical engineering, 2016. **9**: p. 234-263.
11.   Duncan, J.S. and N. Ayache, *Medical image analysis: Progress over two decades and the challenges ahead.* IEEE transactions on pattern analysis and machine intelligence, 2000. **22**(1): p. 85-106.
12.   Farahani, N., A.V. Parwani, and L. Pantanowitz, *Whole slide imaging in pathology: advantages, limitations, and emerging perspectives.* Pathol Lab Med Int, 2015. **7**: p. 23-33.
13.   de Matos, J., et al., *Histopathologic Image Processing: A Review.* arXiv preprint arXiv:1904.07900, 2019.
14.   Srinidhi, C.L., O. Ciga, and A.L. Martel, *Deep neural network models for computational histopathology: A survey.* Medical Image Analysis, 2020: p. 101813.
15.   Cheplygina, V., M. de Bruijne, and J.P. Pluim, *Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis.* Medical image analysis, 2019. **54**: p. 280-296.
16.   LeCun, Y., Y. Bengio, and G. Hinton, *Deep learning.* nature, 2015. **521**(7553): p. 436-444.
17.   Taghanaki, S.A., et al., *Deep semantic segmentation of natural and medical images: A review.* Artificial Intelligence Review, 2021. **54**(1): p. 137-178.
18.   Litjens, G., et al., *A survey on deep learning in medical image analysis.* Medical image analysis, 2017. **42**: p. 60-88.
19.   Miyoshi, H., et al., *Deep learning shows the capability of high-level computer-aided diagnosis in malignant lymphoma.* Laboratory Investigation, 2020. **100**(10): p. 1300-1310.
20.   Feng, R., et al., *A Deep Learning Approach for Colonoscopy Pathology WSI Analysis: Accurate Segmentation and Classification.* IEEE Journal of Biomedical and Health Informatics, 2020.
21.   Borovec, J., et al., *ANHIR: automatic non-rigid histological image registration challenge.* IEEE transactions on medical imaging, 2020. **39**(10): p. 3042-3052.

22. Celik, Y., et al., *Automated invasive ductal carcinoma detection based using deep transfer learning with whole-slide images.* Pattern Recognition Letters, 2020. **133**: p. 232-239.

23. Balkenhol, M.C., et al., *Deep learning assisted mitotic counting for breast cancer.* Laboratory Investigation, 2019. **99**(11): p. 1596-1606.

24. Tizhoosh, H.R. and L. Pantanowitz, *Artificial intelligence and digital pathology: challenges and opportunities.* Journal of pathology informatics, 2018. **9**.

25. Steiner, D.F., P.-H.C. Chen, and C.H. Mermel, *Closing the translation gap: AI applications in digital pathology.* Biochimica et Biophysica Acta (BBA)-Reviews on Cancer, 2020: p. 188452.

26. Gibson, E., et al. *Inter-site Variability in Prostate Segmentation Accuracy Using Deep Learning.* in *MICCAI.* 2018. Springer Cham.

27. Tellez, D., et al., *Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology.* Medical image analysis, 2019. **58**: p. 101544.

28. Ianni, J.D., et al., *Tailored for real-world: a whole slide image classification system validated on uncurated multi-site data emulating the prospective pathology workload.* Scientific reports, 2020. **10**(1): p. 1-12.

29. Brunyé, T.T., et al., *Accuracy is in the eyes of the pathologist: The visual interpretive process and diagnostic accuracy with digital whole slide images.* Journal of biomedical informatics, 2017. **66**: p. 171-179.

30. Altaf, F., et al., *Going Deep in Medical Image Analysis: Concepts, Methods, Challenges, and Future Directions.* IEEE Access, 2019. **7**: p. 99540-99572.

31. Levy, J.J., et al., *PathFlow-MixMatch for Whole Slide Image Registration: An Investigation of a Segment-Based Scalable Image Registration Method.* bioRxiv, 2020.

32. Tellez, D., et al., *Whole-slide mitosis detection in H&E breast histology using PHH3 as a reference to train distilled stain-invariant convolutional networks.* IEEE transactions on medical imaging, 2018. **37**(9): p. 2126-2136.

33. Leo, P., et al., *Evaluating stability of histomorphometric features across scanner and staining variations: prostate cancer diagnosis from whole slide images.* Journal of medical imaging, 2016. **3**(4): p. 047502.

34. Khan, A., et al. *Generalizing convolution neural networks on stain color heterogeneous data for computational pathology.* in *Medical Imaging 2020: Digital Pathology.* 2020. International Society for Optics and Photonics.

35. Bejnordi, B.E., et al., *Stain specific standardization of whole-slide histopathological images.* IEEE transactions on medical imaging, 2015. **35**(2): p. 404-415.

36. Bejnordi, B.E., et al., *Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer.* Jama, 2017. **318**(22): p. 2199-2210.

37. Evans, A.J., et al., *Implementation of whole slide imaging for clinical purposes: issues to consider from the perspective of early adopters.* Archives of pathology & laboratory medicine, 2017. **141**(7): p. 944-959.

38. Rączkowski, Ł., et al., *ARA: accurate, reliable and active histopathological image classification framework with Bayesian deep learning.* Scientific reports, 2019. **9**(1): p. 1-12.

39. LeCun, Y. and Y. Bengio, *Convolutional networks for images, speech, and time series.* The handbook of brain theory and neural networks, 1995. **3361**(10): p. 1995.

40. Litjens, G., et al., *Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis.* Scientific reports, 2016. **6**: p. 26286.

41. Tajbakhsh, N., et al., *Convolutional neural networks for medical image analysis: Full training or fine tuning?* IEEE transactions on medical imaging, 2016. **35**(5): p. 1299-1312.

42. Krizhevsky, A., I. Sutskever, and G.E. Hinton. *Imagenet classification with deep convolutional neural networks.* in *NIPS.* 2012.

43. Alam, F., S.U. Rahman, and A. Khalil, *An investigation towards issues and challenges in medical image registration.* Journal of Postgraduate Medical Institute (Peshawar-Pakistan), 2017. **31**(3).

44. Wodzinski, M. and A. Skalski, *Multistep, automatic and nonrigid image registration method for histology samples acquired using multiple stains.* Physics in Medicine & Biology, 2021. **66**(2): p. 025006.

45. Lotz, J., et al., *Patch-based nonlinear image registration for gigapixel whole slide images.* IEEE Transactions on Biomedical Engineering, 2015. **63**(9): p. 1812-1819.

46. Van der Laak, J., G. Litjens, and F. Ciompi, *Deep learning in histopathology: the path to the clinic.* Nature medicine, 2021. **27**(5): p. 775-784.

47. Serag, A., et al., *Translational AI and deep learning in diagnostic pathology.* Frontiers in medicine, 2019. **6**: p. 185.

48. Sergi, C.M., *Digital pathology: the time is now to bridge the gap between medicine and technological singularity*, in *Interactive Multimedia-Multimedia Production and Digital Storytelling*. 2019, IntechOpen London, UK.

49. Trahearn, N., et al., *Hyper-stain inspector: a framework for robust registration and localised co-expression analysis of multiple whole-slide images of serial histology sections.* Scientific Reports, 2017. **7**(1): p. 1-13.

50. Wodzinski, M. and H. Müller, *DeepHistReg: Unsupervised deep learning registration framework for differently stained histology samples.* Computer Methods and Programs in Biomedicine, 2021. **198**: p. 105799.

51. van Rijthoven, M., et al. *Few-shot weakly supervised detection and retrieval in histopathology whole-slide images*. in *Medical Imaging 2021: Digital Pathology*. 2021. International Society for Optics and Photonics.

52. Tajbakhsh, N., et al., *Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation.* Medical Image Analysis, 2020. **63**: p. 101693.

53. Otálora, S., et al., *Staining invariant features for improving generalization of deep convolutional neural networks in computational pathology.* Frontiers in bioengineering and biotechnology, 2019. **7**: p. 198.

54. Frid-Adar, M., et al. *Synthetic data augmentation using GAN for improved liver lesion classification*. in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. 2018. IEEE.

55. Armanious, K., et al., *MedGAN: Medical image translation using GANs.* Computerized medical imaging and graphics, 2020. **79**: p. 101684.

56. BenTaieb, A. and G. Hamarneh, *Adversarial stain transfer for histopathology image analysis.* IEEE transactions on medical imaging, 2017. **37**(3): p. 792-802.

57. Creswell, A., et al., *Generative adversarial networks: An overview.* IEEE Signal Processing Magazine, 2018. **35**(1): p. 53-65.

58. Karras, T., S. Laine, and T. Aila. *A style-based generator architecture for generative adversarial networks*. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019.

59. Shaban, M.T., et al. *Staingan: Stain style transfer for digital histological images*. in *2019 Ieee 16th international symposium on biomedical imaging (Isbi 2019)*. 2019. IEEE.

60. Pan, S.J. and Q. Yang, *A survey on transfer learning.* IEEE Transactions on knowledge and data engineering, 2009. **22**(10): p. 1345-1359.

61. Ngiam, J., et al., *Domain adaptive transfer learning with specialist models.* arXiv preprint arXiv:1811.07056, 2018.

62. Baykal, E., et al., *Transfer learning with pre-trained deep convolutional neural networks for serous cell classification.* Multimedia Tools and Applications, 2020. **79**(21): p. 15593-15611.

63. Chaitanya, K., et al. *Semi-supervised and task-driven data augmentation*. in *International conference on information processing in medical imaging*. 2019. Springer.

64. Chen, J., et al., *Few-shot breast cancer metastases classification via unsupervised cell ranking.* IEEE/ACM transactions on computational biology and bioinformatics, 2019.

65. Yao, J., et al., *Whole slide images based cancer survival prediction using attention guided deep multiple instance learning networks.* Medical Image Analysis, 2020. **65**: p. 101789.

66. Yang, Y., et al., *One-class classification using generative adversarial networks.* IEEE Access, 2019. **7**: p. 37970-37979.

67. Koohbanani, N.A., et al., *Self-path: Self-supervision for classification of pathology images with limited annotations.* IEEE Transactions on Medical Imaging, 2021. **40**(10): p. 2845-2856.

68. Ranzato, M.A., G. Hinton, and Y. LeCun, *Guest editorial: Deep learning.* International Journal of Computer Vision, 2015. **113**(1): p. 1-2.

69. Gibson, E., et al. *Inter-site variability in prostate segmentation accuracy using deep learning.* in *International Conference on Medical Image Computing and Computer-Assisted Intervention.* 2018. Springer.

70. Hinton, J.P., et al., *A Method to Reuse Archived H&E Stained Histology Slides for a Multiplex Protein Biomarker Analysis.* Methods and protocols, 2019. **2**(4): p. 86.

71. Rana, A., et al., *Use of deep learning to develop and analyze computational hematoxylin and eosin staining of prostate core biopsy images for tumor diagnosis.* JAMA network open, 2020. **3**(5): p. e205111-e205111.

72. Mercan, E., et al., *Localization of diagnostically relevant regions of interest in whole slide images: A comparative study.* Journal of digital imaging, 2016. **29**(4): p. 496-506.

73. Aswathy, M. and M. Jagannath, *Detection of breast cancer on digital histopathology images: Present status and future possibilities.* Informatics in Medicine Unlocked, 2017. **8**: p. 74-79.

74. Ilić, I.R., et al., *The quantitative ER immunohistochemical analysis in breast cancer: Detecting the 3+ 0, 4+ 0, and 5+ 0 allred score cases.* Medicina, 2019. **55**(8): p. 461.

75. Azizun-Nisa, B.Y., F. Raza, and N. Kayani, *Comparison of ER, PR and HER-2/neu (C-erb B 2) reactivity pattern with histologic grade, tumor size and lymph node status in breast cancer.* Asian Pac J Cancer Prev, 2008. **9**(4): p. 553-6.

76. Adams, A.L., et al., *The effect of neoadjuvant chemotherapy on histologic grade, hormone receptor status, and HER2/neu status in breast carcinoma.* The breast journal, 2008. **14**(2): p. 141-146.

77. Sughayer, M.A., et al., *Prevalence of hormone receptors and HER2/neu in breast cancer cases in Jordan.* Pathology & Oncology Research, 2006. **12**(2): p. 83-86.

78. Obando, D.F.G., et al. *Multi-staining registration of large histology images.* in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017).* 2017. IEEE.

79. Wodzinski, M. and H. Müller. *Learning-based affine registration of histological images.* in *International Workshop on Biomedical Image Registration.* 2020. Springer.

80. Wilbur, D.C., et al., *Using image registration and machine learning to develop a workstation tool for rapid analysis of glomeruli in medical renal biopsies.* Journal of Pathology Informatics, 2020. **11**.

81. Oliveira, F.P. and J.M.R. Tavares, *Medical image registration: a review.* Computer methods in biomechanics and biomedical engineering, 2014. **17**(2): p. 73-93.

82. Hu, Y., et al. *Label-driven weakly-supervised learning for multimodal deformable image registration.* in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018).* 2018. IEEE.

83. Wu, G., et al., *Scalable high-performance image registration framework by unsupervised deep feature representations learning.* IEEE Transactions on Biomedical Engineering, 2015. **63**(7): p. 1505-1516.

84. Solorzano, L., et al., *Whole slide image registration for the study of tumor heterogeneity*, in *Computational pathology and ophthalmic medical image analysis.* 2018, Springer. p. 95-102.

85.   Jiang, J., et al., *Robust hierarchical density estimation and regression for re-stained histological whole slide image co-registration.* Plos one, 2019. **14**(7): p. e0220074.

86.   Anyz, J., et al., *Spatial mapping of metals in tissue-sections using combination of mass-spectrometry and histology through image registration.* Scientific reports, 2017. **7**(1): p. 1-13.

87.   Zhao, S., et al. *Recursive cascaded networks for unsupervised medical image registration.* in *Proceedings of the IEEE/CVF International Conference on Computer Vision.* 2019.

88.   Smith, L.N. and N. Topin. *Super-convergence: Very fast training of neural networks using large learning rates.* in *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications.* 2019. International Society for Optics and Photonics.

89.   Oymak, S., *Super-Convergence with an Unstably Large Learning Rate.*

90.   Pasqualucci, L., et al., *Genetics of follicular lymphoma transformation.* Cell reports, 2014. **6**(1): p. 130-140.

91.   Kenzik, K.M., et al., *Congestive heart failure in older adults diagnosed with follicular lymphoma: A population‐based study.* Cancer, 2018. **124**(21): p. 4221-4230.

92.   Fisher, R.I. and N. Khan, *Is observation dead in follicular lymphoma? No, but the apoptosis pathway has been activated.* Journal of the National Comprehensive Cancer Network, 2015. **13**(3): p. 363-366.

93.   Bray, F., et al., *Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.* CA: a cancer journal for clinicians, 2018. **68**(6): p. 394-424.

94.   De Fauw, J., et al., *Clinically applicable deep learning for diagnosis and referral in retinal disease.* Nature medicine, 2018. **24**(9): p. 1342-1350.

95.   Mormont, R., P. Geurts, and R. Marée. *Comparison of deep transfer learning strategies for digital pathology.* in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops.* 2018.

96.   Yosinski, J., et al., *How transferable are features in deep neural networks?* arXiv preprint arXiv:1411.1792, 2014.

97.   Zech, J.R., et al., *Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study.* PLoS medicine, 2018. **15**(11): p. e1002683.

98.   Janowczyk, A. and A. Madabhushi, *Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases.* Journal of pathology informatics, 2016. **7**.

99.   Therrien, R. and S. Doyle. *Role of training data variability on classifier performance and generalizability.* in *Medical Imaging 2018: Digital Pathology.* 2018. International Society for Optics and Photonics.

100.   do Nascimento, M.Z., et al., *Lymphoma images analysis using morphological and non-morphological descriptors for classification.* Computer methods and programs in biomedicine, 2018. **163**: p. 65-77.

101.   Bueno, G., et al., *New trends of emerging technologies in digital pathology.* Pathobiology, 2016. **83**(2-3): p. 61-69.

102.   Hamidinekoo, A., et al., *Deep learning in mammography and breast histology, an overview and future trends.* Medical image analysis, 2018. **47**: p. 45-67.

103.   Yoon, H., et al., *Tumor identification in colorectal histology images using a convolutional neural network.* Journal of digital imaging, 2019. **32**(1): p. 131-140.

104.   El Achi, H., et al., *Automated diagnosis of lymphoma with digital pathology images using deep learning.* Annals of Clinical & Laboratory Science, 2019. **49**(2): p. 153-160.

105.   Wollmann, T., C. Eijkman, and K. Rohr. *Adversarial domain adaptation to improve automatic breast cancer grading in lymph nodes.* in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018).* 2018. IEEE.

106. Lodge, M., *The role of the Commonwealth in the wider cancer control agenda.* The Lancet Oncology, 2020. **21**(7): p. 879-881.

107. Carbone, A., et al., *Follicular lymphoma.* Nature Reviews Disease Primers, 2019. **5**(1): p. 1-20.

108. Campanella, G., et al., *Clinical-grade computational pathology using weakly supervised deep learning on whole slide images.* Nature medicine, 2019. **25**(8): p. 1301-1309.

109. Khan, S., et al., *A novel deep learning based framework for the detection and classification of breast cancer using transfer learning.* Pattern Recognition Letters, 2019. **125**: p. 1-6.

110. Dou, Q., et al., *PnP-AdaNet: Plug-and-play adversarial domain adaptation network at unpaired cross-modality cardiac segmentation.* IEEE Access, 2019. **7**: p. 99065-99076.

111. Yu, K.-H., et al., *Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features.* Nature communications, 2016. **7**(1): p. 1-10.

112. Levy-Jurgenson, A., et al., *Spatial transcriptomics inferred from pathology whole-slide images links tumor heterogeneity to survival in breast and lung cancer.* Scientific reports, 2020. **10**(1): p. 1-11.

113. Dimitriou, N., O. Arandjelović, and P.D. Caie, *Deep learning for whole slide image analysis: an overview.* Frontiers in medicine, 2019. **6**: p. 264.

114. Rony, J., et al., *Deep weakly-supervised learning methods for classification and localization in histology images: a survey.* arXiv preprint arXiv:1909.03354, 2019.

115. Wang, X., et al., *Weakly supervised deep learning for whole slide lung cancer image analysis.* IEEE transactions on cybernetics, 2019. **50**(9): p. 3950-3962.

116. Tschuchnig, M.E., G.J. Oostingh, and M. Gadermayr, *Generative adversarial networks in digital pathology: a survey on trends and future potential.* Patterns, 2020. **1**(6): p. 100089.

117. Hägele, M., et al., *Resolving challenges in deep learning-based analyses of histopathological images using explanation methods.* Scientific reports, 2020. **10**(1): p. 1-12.

118. Yi, X., E. Walia, and P. Babyn, *Generative adversarial network in medical imaging: A review.* Medical image analysis, 2019. **58**: p. 101552.

119. Hou, L., et al., *Robust Histopathology Image Analysis: to Label or to Synthesize?*

120. Gour, M., S. Jain, and T. Sunil Kumar, *Residual learning based CNN for breast cancer histopathological image classification.* International Journal of Imaging Systems and Technology, 2020. **30**(3): p. 621-635.

121. Kassani, S.H., et al., *Classification of histopathological biopsy images using ensemble of deep learning networks.* arXiv preprint arXiv:1909.11870, 2019.

122. Somaratne, U.V., et al. *Improving follicular lymphoma identification using the class of interest for transfer learning.* in *2019 Digital Image Computing: Techniques and Applications (DICTA).* 2019. IEEE.

123. Kong, B., et al. *Invasive cancer detection utilizing compressed convolutional neural network and transfer learning.* in *International Conference on Medical Image Computing and Computer-Assisted Intervention.* 2018. Springer.

124. Perera, P. and V.M. Patel, *Learning deep features for one-class classification.* IEEE Transactions on Image Processing, 2019. **28**(11): p. 5450-5463.

125. Sabokrou, M., et al. *Adversarially learned one-class classifier for novelty detection.* in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 2018.

126. Tavolara, T.E., et al., *A modular cGAN classification framework: Application to colorectal tumor detection.* Scientific reports, 2019. **9**(1): p. 1-8.

127. Ren, J., et al. *Adversarial domain adaptation for classification of prostate histopathology whole-slide images.* in *International Conference on Medical Image Computing and Computer-Assisted Intervention.* 2018. Springer.

128. Van Engelen, J.E. and H.H. Hoos, *A survey on semi-supervised learning.* Machine Learning, 2020. **109**(2): p. 373-440.

129. Gamper, J., et al., *Meta-SVDD:Probabilistic Meta-Learning for One-Class Classification in Cancer Histology Images.* arXiv preprint arXiv:2003.03109, 2020.

130. Diaz-Pinto, A., et al., *Retinal image synthesis and semi-supervised learning for glaucoma assessment.* IEEE transactions on medical imaging, 2019. **38**(9): p. 2211-2218.

131. Radford, A., L. Metz, and S. Chintala, *Unsupervised representation learning with deep convolutional generative adversarial networks.* arXivpreprintarXiv:1511.06434, 2015.

132. Van der Maaten, L. and G. Hinton, *Visualizing data using t-SNE.* Journal of machine learning research, 2008. **9**(11).

133. Chung, D. and E.J. Delp. *Camera-aware image-to-image translation using similarity preserving StarGAN for person re-identification.* in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 2019.

134. Zhu, J.-Y., et al. *Unpaired image-to-image translation using cycle-consistent adversarial networks*. in *Proceedings of the IEEE international conference on computer vision*. 2017.