University of Windsor

Scholarship at UWindsor

Histopathology Classification of Colorectal Cancer Whole Slide Images Using Color Features with Deep Residual Transfer Learning

Ali Hassan
University of Windsor

Follow this and additional works at: https://scholar.uwindsor.ca/etd

Part of the Computer Sciences Commons

Histopathology Classification of Colorectal Cancer Whole Slide Images Using Color Features with Deep Residual Transfer Learning

By

Ali Hassan

A Thesis

Submitted to the Faculty of Graduate Studies

through the School of Computer Science

in Partial Fulfillment of the Requirements for

the Degree of Master of Science at the

University of Windsor

Windsor, Ontario, Canada

2022

Histopathology Classification of Colorectal Cancer Whole Slide Images Using Color Features with Deep Residual Transfer Learning

by

Ali Hassan

APPROVED BY:

―――――――――――――――――――

F. Gherib

Department of Civil and Environmental Engineering

―――――――――――――――――――

I. Ahmad

School of Computer Science

―――――――――――――――――――

B. Boufama, Advisor

School of Computer Science

August 15, 2022

Declaration of Originality

I hereby certify that I am the sole author of this thesis and that no part of this thesis has been published or submitted for publication.

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis and have included copies of such copyright clearances to my appendix.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution.

Abstract

Colorectal cancer (CRC) is an emerging global health concern. An average of 73 Canadians will be diagnosed with CRC every day and 27 Canadians will lose their life as a result of it. CRC accounts for 12% of all cancer deaths in Canada in the year 2020. Early and accurate diagnosis is vital in saving lives as it significantly influences the length of survival of the patient. Deep learning can be leveraged to aid in the task of identifying cancerous cells within pre-cancerous tissue samples, which are taken from colorectal polyps of patients for CRC screening. In this study, an attempt to improve existing supervised methods of classification of colorectal cancer is made. By revamping/improving the deep learning architecture in ResNet. The network will be trained on a much larger and relevant dataset of colorectal WSI (Whole Slide Image) patches. This study aims to attain better overall accuracy by incorporating color features, which have not been concentrated on in previous studies. All while retaining similar performance as compared to existing state-of-the-art methods of CRC classification. Four network models are applied to a large histopatological dataset. All network models are variations of Residual networks at multiple depths. The best results are attained using a pre-trained ResNet-50 model. The overall results show that the residual network performs similary to much deeper DenseNet-121 model and better than the cell level framework described in a previous study. The ResNet-50 model achieved 88.58%, 92.04%, 81.86%, 86.65% Accuracy, Precision, Recall and F1-Score respectively.

Dedication

I would like to dedicate this thesis to my parents who made me into the person I am today...

Acknowledgements

I want to express my sincere gratitude towards my advisor Dr. B. Boufama for treating me with such patience during COVID, which ended up being one of the toughest time in my life, both mentally and physically. I also would like to thank all those who not only motivated me but helped me during this trying time upon all of us. May Allah help and guide us all... Ameen.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter starts with the introduction to core concepts. It elaborates the various disciplines that fall under Artificial Intelligence including, *Machine Learning, Deep Learning and Computer Vision.* Further it introduces colorectal cancer and its types and causes. Finally, we briefly explain what Digital Pathology and take a look at the types of test used to detect cancerous cells in digital pathology and how it relates to the topic of this research.

1.1 Artificial Intelligence

Natural intelligence is one which is exhibited by humans and animals. In contrast to that Artificial intelligence is the study of intelligent agents that are working towards maximizing its odds of achieving a specific goal in a perceivable environment. Artificial intelligence is an umbrella term that incorporates various sub-disciplines. Some early examples of artificial intelligence include chat bots, optical character recognition and recommendation systems etc.

Joseph Weizenbaum at the *Massachusetts Institute of Technology (MIT)* created a chatbot named Eliza in 1966. It acted upon a set of defined rules. Its capable of having conversations based on a script. The input given to Eliza is scanned for various keywords. this triggered an output based on a set of defined rules. Eliza is considered one of the early applications of artificial intelligence. The method of processing inputs that Eliza utilized,

are still similar to that of modern virtual assistants. It is noted that some of the patients used to forget that they are conversing with a computer program [1]. The most famous script for was the doctor script, which transformed Eliza a psychotherapist [11].

FIGURE 1.1: Conversation with ELIZA [1]

Another first application of Artificial intelligence was in the form of a chess playing computer. The concept was defined by Alan Turing. He described chess as a source of clearly defined yet, challenging problems. At first the computer would play using an exhaustive search of all available moves that can be performed. This approach is impractical as chess has an enormous number of available moves considering all the pieces. In *1949* Turing predicted that in the future computers will be able to play chess at a significant level of difficulty. [12]

Fast forward 50 years, in the year 1997 a computer named *"Deep Blue"* was developed by International Business Machines Corporation (IBM). Because of the advancements made in computer hardware in those 50 years, this computer was able to beat the current world champion "Garry Kasparov" in a match of six individual games. This computer ran an algorithm called "Turbo Champ" that was created by Alan Turing in the late 1940s [13].

1.2 Machine Learning

Machine Learning programs can carry out tasks without explicit instructions. It allows computers to learn from historical data in order to carry out a specific job. It is possible to write algorithms that tell the machine how to accomplish all the steps required to solve a problem. For basic jobs, a programmer is able to write all the instructions manually so that the computer can accomplish a task. However, for complex programs it is a hassle for the programmers to provide all the instructions necessary to complete the task. In practice, assisting the computer in developing its algorithm rather than having human programmers explain each required step can be more productive [14]. in conclusion, machine learning is a way of making computers learn and remember from past experiences.

Machine learning utilizes mathematical models to infer outputs based on incoming samples. The procedure can be further broken down into three steps. The first step in which a suitable data loader is responsible to process data and forward it to the Machine learning algorithm in a proper manner. Next we design a machine learning algorithm to learn features from the data by solving optimization problem minimizing the error rate. The final step is to run the model let it train completely and then evaluate it performance based on suitable metrics. From that we end up with a model that is capable of learning and predicting data similar to the one it is train upon. Machine learning algorithms can be further divided into two types. The type of machine learning algorithm to be applied depends on the type of the dataset.

Supervised Machine Learning : Suitable when the dataset comes labelled into classes.

Unsupervised Machine Learning : Suitable when the dataset is unlabelled.

1.2.1 Supervised Machine Learning

It is defined as a approach to make use of algorithms to train classify and predict labelled datasets. Figure 1.2 shows the basic flow of supervised machine learning. The raw data is labelled into classes. The labelled dataset is divided in to training and testing sets. These sets are provided to the algorithms which uses the training data to learn the classes and predicts the classes of the unseen test set. Common supervised algorithms are Naive

Bayes, Support Vector Machine (SVM), Linear Regression, Logistic Regression and Random Forest.
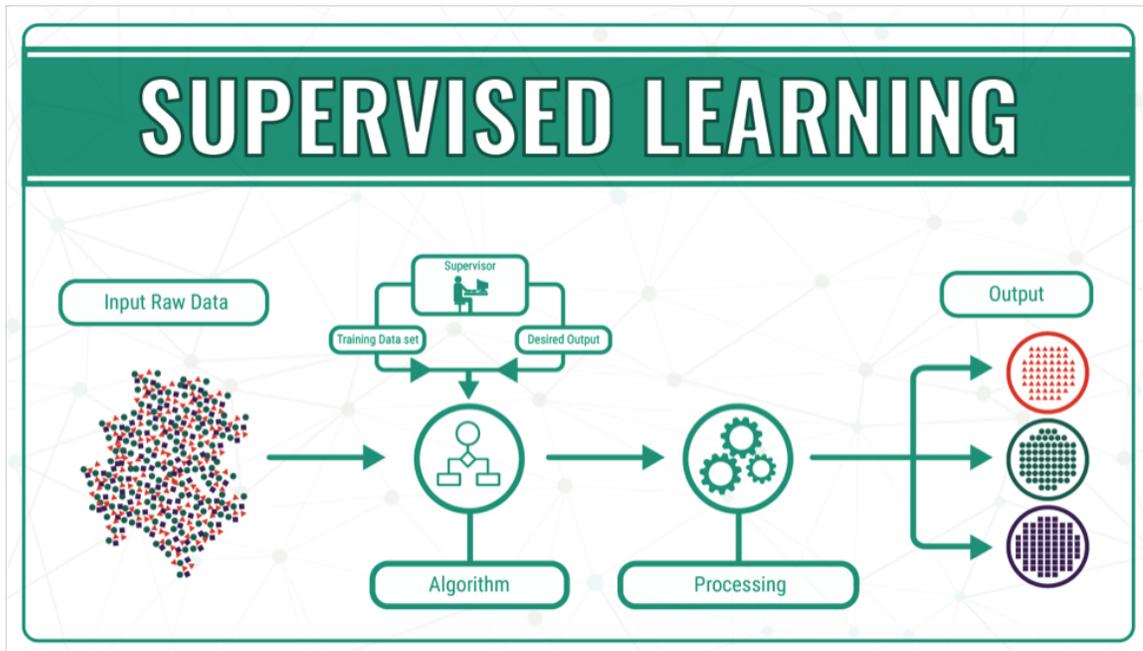
FIGURE 1.2: Supervised Machine Learning [2]

1.2.2 Unsupervised Machine Learning

It is defined as a approach to make use of algorithms to group similar data among unlabelled datasets. Figure 1.3 shows the basic flow of unsupervised machine learning. The raw unlabelled data is interpreted to find features that can be used to group data based on similarity. Multiple algorithms/features can be tested to evaluate the best suited algorithm. The output is in the form of classes in which similar data points are grouped based on. Common unsupervised algorithms are K-means clustering, K Nearest Neighbour, Principle Component Analysis (PCA), Singular Value Decomposition (SVD) and Hierarchical clustering.

1.3 Deep Learning

Deep learning is a branch of machine learning. It deals with deep learning algorithms structured similarly to a human brain. Deep learning is responsible for improving Convolutional
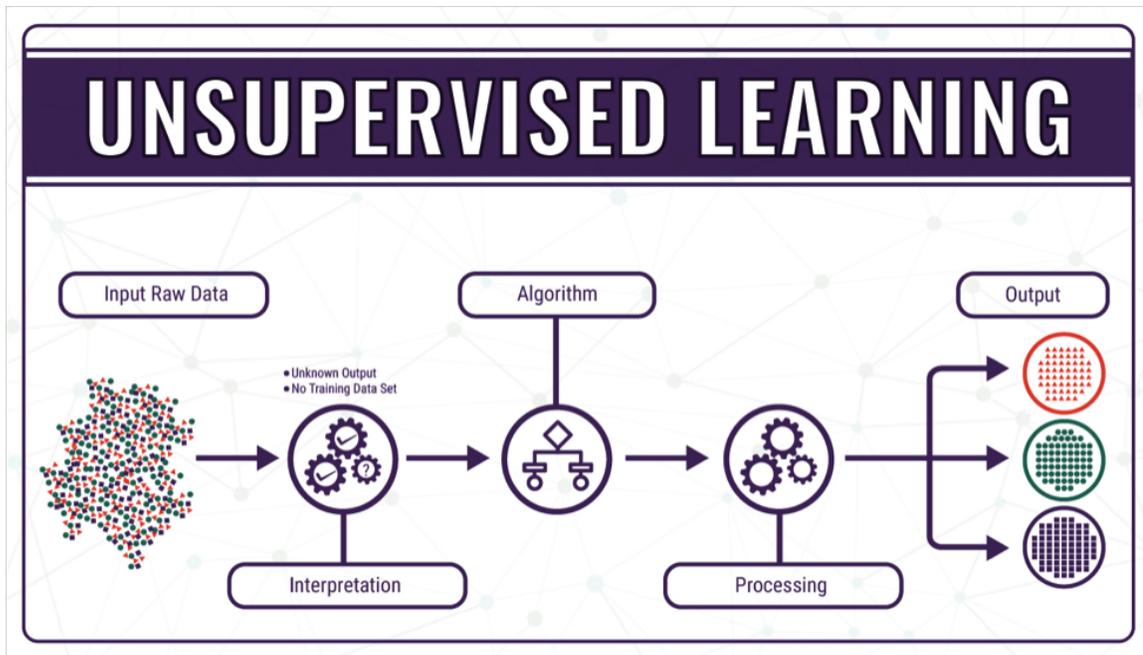
FIGURE 1.3: Unsupervised Machine Learning [2]

Neural Networks substantially. As a result, Deep Convolutional Neural Networks have become an essential part of computer vision. CNNs are great for vision-based tasks. They are able to learn and recognize patterns and objects in relation to the task at hand. Convolutional Neural Networks (CNNs) are a type of Artificial Neural Networks (ANNs). In simple words, an Artificial Neural Network consists of multiple units called neurons. Each neuron is a part of a layer. Subsequent layers have a set of connections. The network must have an input layer, a variable number of middle layers, and an output layer. The input data is transformed as it flows through the network. The output layer generates predictions based on the learned features [15].

Structure of a simple neural network can be observed in the Figure 1.4. A neural network consists of an input layer *(x)*, a variable number of hidden layers and at the end the output layer *(y)* . There are a set of weights *(w)* associated with each connection and a bias *(b)* added to it. To add non-linearity, each layer has an activation function. A loss function monitors how well the learning is being performed by checking if an output is different than the desired output. The adjustments are made to the weights w based on loss function also known as the error rate. There are a number of loss functions that can be used based upon the type of data and the approach. When the network is initialized the weights w are set to

random values. As the network proceeds learning, it makes adjustments to these weights. The adjustments are made during backpropogation step by calculating the gradient of the error rate with respect to the network weights. These adjustments are made repetitively until a network is able to predict an output y with a desired degree of accuracy and low error rate [16].
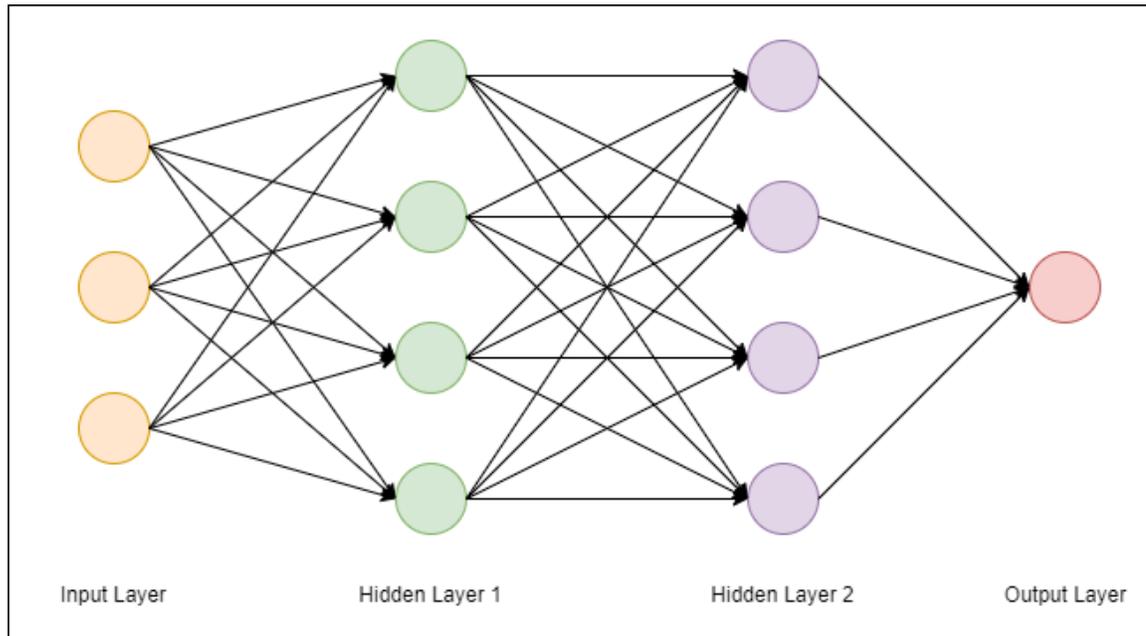
FIGURE 1.4: Structure of a Simple Neural Network

1.4 Computer Vision

Human eyes are capable of perceiving visual information from the surroundings. The visual data, gathered at a very high rate is compressed by the visual cortex and then provided to the brain. Only a subset of this data is perceived by the human brain [17]. The definition of vision is hard to summarize because the information perceived is relative to the task. Computer vision deals with techniques of extracting information from visual data similar to that our eyes perceive from the world around us. Creating computer vision solutions may come with inherent challenges which depend on the type of data available for the specific use case. Computer vision enables computers to achieve a high level of understanding of images and videos. The Figure 1.5 shows that computer vision integrates all sub-types of artificial intelligence.
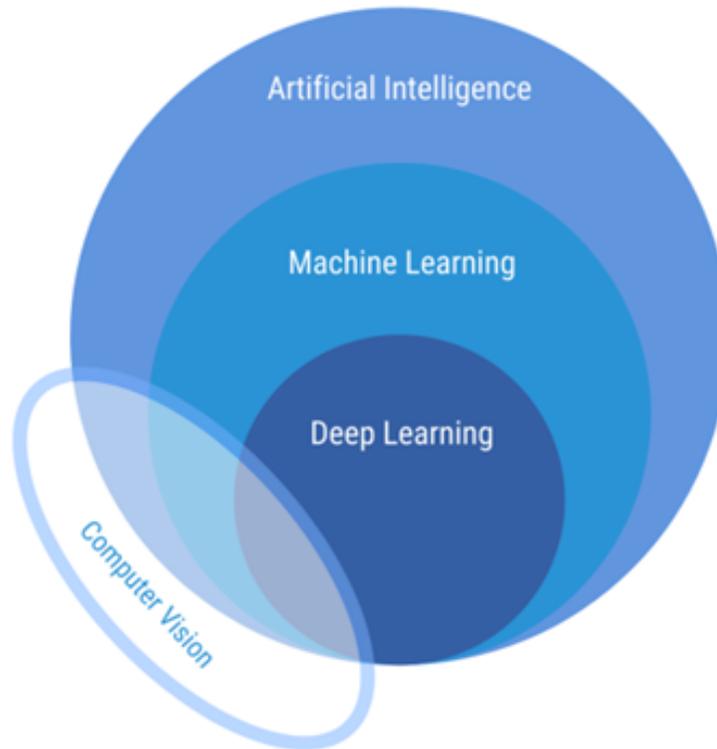
FIGURE 1.5: Computer Vision as a discipline in Artificial Intelligence [3]

Computer vision is an umbrella term, it incorporates multiple topics of research. Some of these topics include scene recognition and scene reconstruction, augmented reality, object tracking and recognition, motion detection and autonomous driving assistance systems. The tasks achieved by computer vision are vast and each application comes with its own set of challenges. For example the input image data may encounter variations, such as lighting, contrast, viewing angle, occlusion and barrel distortion due to a wide angle camera [18]. Face recognition is one of the applications of computer vision. In [4] authors compare state of the art Deep face recognition solutions. The Figure 1.6 shows the complete deep face recognition system workflow. First the faces need to be detected and aligned in the processing stage. The face recognition solution have a anti-spoofing layer to determine if the face is live. After the processing the faces, different deep network architectures are used to extract discriminating features during training phase. Face matching algorithms are used to classify the features extracted during the testing phase.
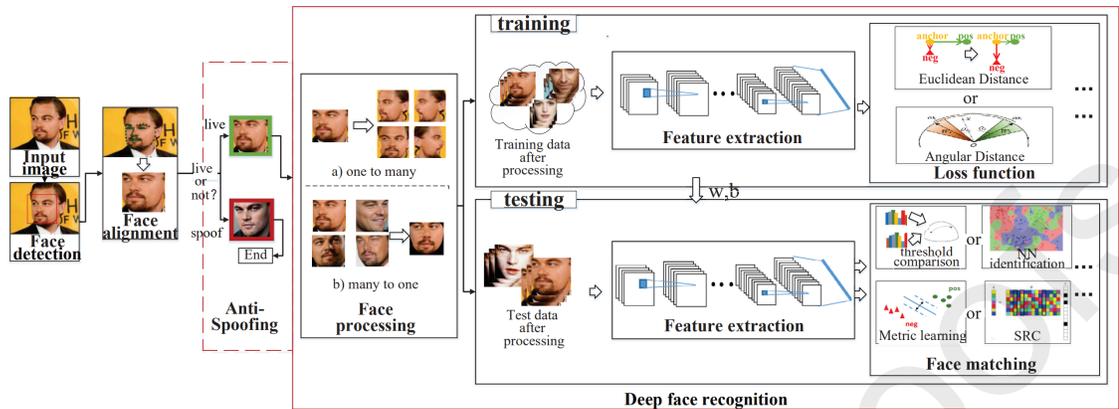
FIGURE 1.6: Face detector with alignment and recognition [4]

1.5   Colorectal Cancer

Most colorectal cancers start as a growth on the inner lining of the colon or rectum. These growths are called polyps. Some of the polyps may become cancerous over time. The chance of a polyp turning into cancer depends on the type of polyp and certain genetic factors. There are several types of polyps, categorized by the shape, size, risk of turning cancerous and causing health issues for the patient. Stages of colorectal polyps can be observed in Figure 1.7. The basic types of polyps are detailed below. [19]

Adenomatous polyps (adenomas): This is the most common type of polyp, around 70% of all detected polyps are adenomas. It may take several years, but these polyps can turn into cancerous. Because of this, adenomas are identified as a pre-cancerous condition and should be kept under observation. [19]

Hyperplastic polyps/Inflammatory polyps: These polyps are smaller, In general they are not designated as a pre-cancerous condition. But these polyps should be cancer-screened if they're larger than 1cm in diameter. [19]

Sessile/Traditional serrated polyps (SSP/TSP): These polyps are harder to detect as they are flat. Once detected, they are treated as a high risk like adenomas because they are rarely benign and have a high risk of causing colorectal cancer. [19]

To determine if a polyp is cancerous pathologists must obtain a sample and perform a text known as H&E or hematoxylin and eosin test. It helps pathologists to get a detailed
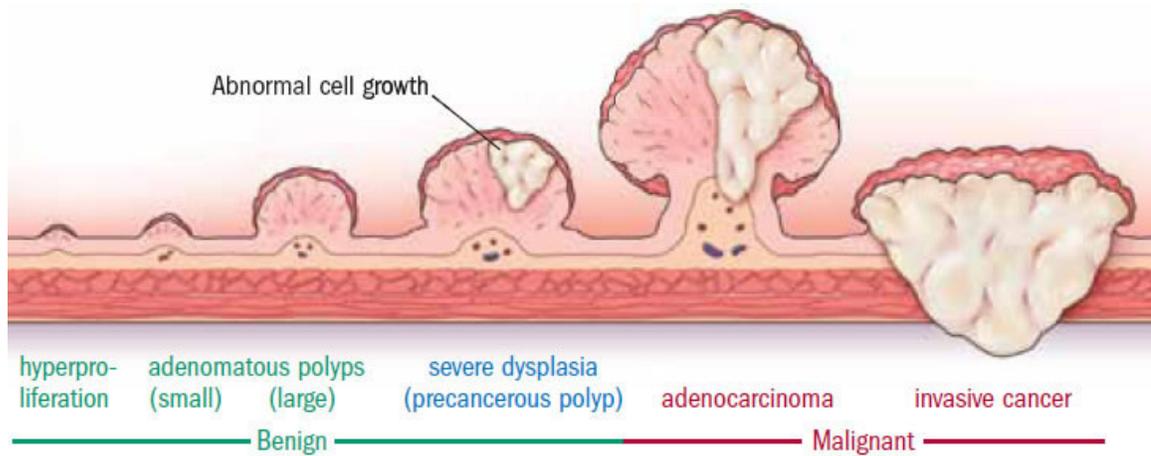
FIGURE 1.7: Stages of colorectal polyps - image acquired from Harvard Health [5]

view of the polyp tissue. The staining results in cancerous tissue and non-cancerous tissue to be color differently so it can easily be distinguished by the pathologist. The results of a H&E test are shown in the Figure 1.8. The hematoxylin colors the cell nuclei purplish blue and the eosin makes the extracellular material pink in color [20].
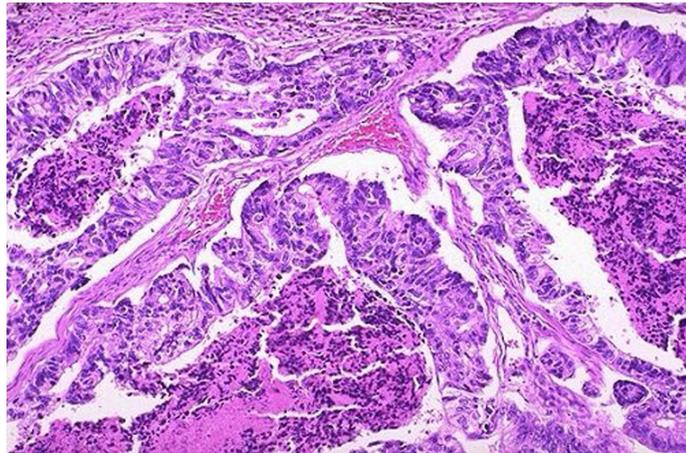
FIGURE 1.8: Examples of H&E Staining Test - image acquired from UNIPA Springer Series
[6]

1.6   Digital Pathology

Digital pathology incorporates the capturing of digital images of the H&E test with the help of a digital microscope. It also includes managing, analyzing and evaluating the digital images, also called WSI or Whole Slide Images, with suitable computer algorithms [21].

Although pathologists' core diagnostic procedure has remained relatively unaltered since the late 1800s, improvements in information technology have created substantial prospects for image-based diagnosis enabling the development of research applications. Histopathology has fallen behind other healthcare disciplines, such as histology, where digitalization is readily applied. As whole-slide imaging equipment becomes more practical and accessible, practices will progressively adopt the technology and, as a result, generate new data, eventually surpassing the current volume of histological imaging data. Many supervised learning techniques capable of segmenting and classifying images are applied to commonly performed pathological tasks [22]. Hence the significance of machine learning in computer vision cannot be understated.

1.7   Research Motivation

Biopsies are a standard procedure used to detect cancerous cells within a tissue sample. Other applications may include detecting tissue rejection in transplanted organs. It is mainly used to detect the presence and spread of cancerous regions in the body. The digitization of these images has enabled the field of digital pathology to emerge as modern digital scanners can take high-resolution images of the tissues. These images are utilized to learn and enhance the diagnostic process.

Digital pathology has gained a lot of recent popularity because of the availability and accessibility of powerful computing hardware. This intern encouraged development and academic research into machine learning applications for digital pathology. The potential of this field of study to act as an aid during the colorectal cancer diagnostic process is relatively higher than other medical diagnostic applications. Researchers can develop a tool to take advantage of the massive image data generated by these biopsies. Learning from historical data, we can detect early signs of cancer development and take action ahead of time to help prevent the rapid spread of the disease.

1.8   Research Contribution

The main contribution of this research is to explore the development of a cancer diagnostic system that would be comparable to the state-of-the-art diagnostic algorithms in terms of running efficiently but can attain greater accuracy.

1. To explore and analyze the application of popular state-of-the-art architectures for colorectal whole-slide image classification.

2. To adapt and optimize a specialized model of a popular state-of-the-art architecture to colorectal cancer classification.

3. To compare and evaluate the specialized model with other advanced research in the same domain.

The results of this project can later be utilized for various applications of whole slide images for detecting the type and presence of cancerous tissue.

1.9   Thesis Outline

The chapters of this thesis is organized in the following sequence:

- Chapter 1 explained background information, motivation, and contributions of the research study.

- Chapter 2 describes the literature survey on previous work done in this field of research.

- Chapter 3 provides a detailed description of the proposed solution, data acquisition, data pre-processing techniques, experimentation environment and evaluation metrics.

- Chapter 4 gives information about the datasets, results and analysis.

- Chapter 5 concludes this thesis project and provides future work for reference.

Chapter 2

Literature Review

This chapter describes some previous works in cancer detection and cancer classification of histopathology slides. It also describes the role of transfer learning in solving classification problems.

2.1   Transfer Learning

This section reviews current research works that use transfer learning with deep neural networks. From the wide variety of existing neural network architectures available, each network was trained on a separate dataset. The weights resulting from the previous experiments can be reused instead of re-initializing it and starting from scratch. The works relating to the advantages of pre-trained networks are explored below.

Transfer learning problems uses prior knowledge to combat a newer task. It adjusts the model weights to adapt to the new data. The domain in which the network was trained before is called the source domain. The domain to which the network is going to adapt is called the target domain. Wan et al. used a transfer learning based approach to study its advantages on existing constraint satisfation problem (CSP) algorithms and deep neural network (DNN) algorithms that are used to study of EEG signal analysis [23]. In [24] the authors created an end to end solution for detection of COVID-19 in chest x-rays using deep convolution networks.the approach augments the existing network to add a Bidirectional

Long Short-Term Memories (BiLSTM) layer, which takes temporal properties of the data also into consideration. Due to the nature of deep learning approach the solution does not require any hand crafted features to evaluate a newly provided sample. It can easily be used as an aid in identifying the effect and seriousness of COVID-19 on the affected patients. [25] presents a deep CNN based approach for detecting COVID-19 induced pneumonia. The dataset consisted of X-ray images. The authors explore existing network architectures including ResNet18, InceptionV3, DenseNet and SqueezeNet. The best classification was achieved using pretrained DenseNet, which came out to be 97.94%. In [26] the authors propose a deep-learning based technique for diagnosing COVID-19 using X-ray images. In this solution transfer learning based VGG-16 and VGG-19 had the best results among the several models that were compared. [27] authors ask a very important question, they ponder whether it is advantageous to use pre-training in existing network architectures in order to classify a histopathology breast cancer dataset or weather it is better to use the pre-trained weights. Regardless of the fact that the pre-trained weights are not from the same domain of study or a similar type of dataset. The results showed that a fine tuned VGG-16 network with transfer learning exhibited a higher accuracy than a non-pretrained model at 95.95% and 92.65% respectively.

2.2   Histopathology Classification

In this section, we examine the previous works in the field of histopathology classification. The applications of histopathology are vast and are not constrained to a specific type of cancer. They relate to different kinds of cancer across several human organs.

The authors of [7] present a review of the categories of operations that are usually used for H&E slides. The main categories are image enhancement, image segmentation, feature extraction and classification. The Fig.2.1 shows the complete sequential process in a block diagram. Image enhancement incorporates the processing of the data for effective applications of algorithms to classify a dataset. This operation is a part of the pre-processing step. Examples for the image enhancement are image scaling, image filtering, contrast enhancement and more. Image segmentation deals with detecting and segmenting the region of interest. The region of interest can simply be described as the part of the whole slide image

that contains the abnormal/cancerous tissue. the segmentation can be performed using an number of techniques. The techniques detailed by the authors are clustering based segmentation, region based segmentation, threshold based segmentation, edge based segmentation and lastly a graph based segmentation. Next category is the feature extraction. This can be done in one of two ways. Manual feature extraction and automatic feature extraction. Manual feature extraction calls for figuring out and describing the features which can be applicable for a given data and implementing a technique to extract the strong relevant features. It is difficult to figure out the relevant features manually. Some of the ways to extract specific type of features manually, are texture based features extraction, intensity based features extraction, shape based features extraction and handcrafted features. Automatic feature extraction is used mostly for deep learning solutions. Deep learning algorithms are capable of extracting highly relevant features automatically. The last category of operations for the H&E slides is classification. Based of the type of feature selection method, a classification algorithm is used to classify the data. The most common type of method used with H&E slides are deep learning algorithms as they extract relevant features automatically. The features extracted automatically may not be evident in the image. Hence using manual feature extraction may become a tedious task. However, this situation may change for different types of image data. But in the case of the histopathological data deep learning networks are preferred. The type of classifiers described in [7] that can be applied for classification of the data are linear classifier, naive Bayes classifier, nearest neighbour classifier, support vector machines, decision trees, boosted trees random forests and neural networks.
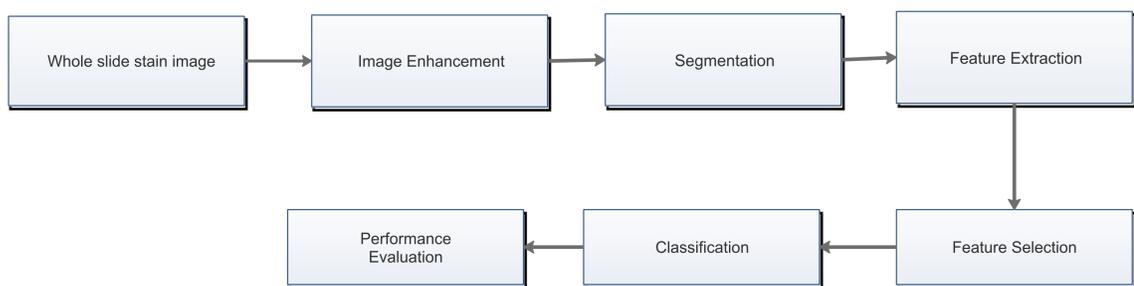
FIGURE 2.1: Block Diagram with sequence. [7]

The author of [28] utilizes several feature extraction techniques to help classify two datasets of colorectal cancer WSIs. First regions of interest (ROI) are identified from

the WSI. Then feature extraction is performed on these regions. Features extracted are dependent on the precision of the ROI segmentation. Color features are extracted among other features including morphological features such as area ,circularity, solidity, perimeter and eccentricity. The color features were largely ignored in previous studies and only the grey channel. The useful feature may have been evident in a color space may be lost when the image is flattened into grey-scale. The color spaces detailed in the study are RGB, HSV and YUV. The proposed solution uses four algorithms including SVM, MLP, Random Forest and Naive Bayes. These algorithms were applied to two datasets. One of the datasets consists of eight classes and the second one contains two classes. The solution achieves a maximum accuracy of 98% on the first dataset and 91% on the second dataset. The recall and precision are 98%, 94% and 88%, 91% for the first and second dataset respectively.

2.3   Deep Residual Networks

In this section, we examine the previous works in the field of Deep residual networks. The main applications of residual networks are mage recognition and classification problems.

In [29] authors discuss the efficiency of residual learning in deep neural networks. The image recognition abilities of residual networks ranks them among the top networks in the world. These networks are usually hundreds and sometimes thousands of layers deep. Most of these layers consist of residual blocks stacked upon one another. For the optimization of these residual networks additional residual network can be added to optimize the original residual mapping. The concept is called Residual Networks of Residual Networks (RoR). The authors argue that simply stacking the residual blocks to create deep networks limits the optimization capacity. Hence the additional residual mapping allows for greater optimization of the original residual network. The application of RoR can be implemented for a variety of residual network architectures such as ResNets, Wide ResNets and Pre-ResNets. The authors claim to have achieved improved results with the CIFAR-100, CIFAR-10 and SVHN datasets with testing errors of 19.73%, 3.77% and 1.59% respectively.

In [8] The authors use two customized Residual Networks models in order to classify a cancer dataset. The dataset consists of two classes Benign and malignant. The authors

convert the images to grey-scale and use Contrast Limited Adaptive Histogram Equalization (CLAHE) in the pre-processing step. The process can be observed in the Fig.2.2.
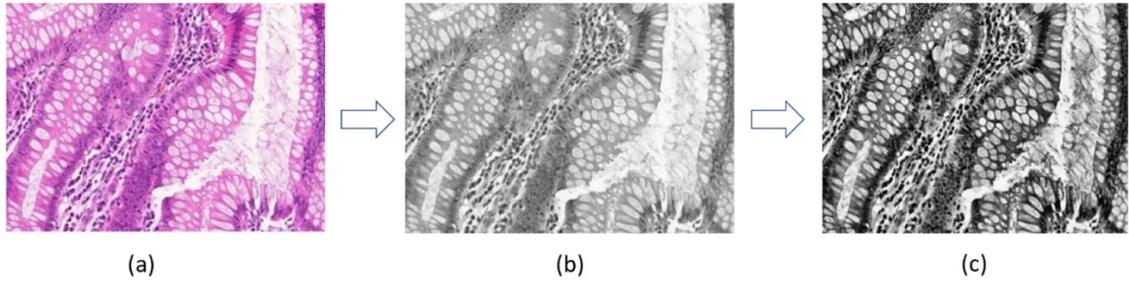
FIGURE 2.2: Preprocessing of images in [8]

Fig.2.2 (a) shows the original image. Fig.2.2 (b) shows the original image after conversion to grey scale. Fig.2.2 (c) shows the grey scale image with CLAHE. The color features were ignored in this case aswell. The dataset was split into 80%-20% training and testing sets. The highest accuracy was 85% with the first model and 88% with the second model. But the second model achieved higher sensitivity percentage at 93% while the first model reached 83%.

The Table 2.1 and 2.2 show some of the most important studies done previously in this field and their respective aim, approach and flaws.

TABLE 2.1: Similar previous work in the field

Ref	Aim	Approach	Problems
[10]	To segment cancerous regions and classify WSI images as cancerous and non cancerous.	A two-step framework combined to accurately classify WSI.	- Cell level framework has low standalone accuracy. - Mis-classification of blood cells as cancerous cells. - Combination framework can benefit from improving the cell level framework.
[30]	To classify diffuse-type ADC in endoscopic biopsy specimen whole slide images (WSIs).	- Identify poorly differentiated ADC. - Using pre-trained networks on WSI of other types of cancer.	- Majorly dependent on parameter tuning of existing networks trained on abdominal WSI. - Only Greyscale features utilized.
[31]	To segment lesions and classify tissues in WSIs as benign or malignant.	- Sliding window based patch selection for training input. - New architecture based on existing VGG-16.	Good for segmentation but less accurate in classification.
[32]	Accurate CRC diagnosis using weakly labeled pathological whole-slide image (WSI) patches.	Used enormous amount of data from over 9000 patients classified using Google Inception V2 model.	- Patches used were only in greyscale. - Missing color features.

TABLE 2.2: Similar previous work in the field contd.

Ref	Aim	Approach	Problems
[28]	To classify colorectal WSIs using colour and morphological features.	• Nuclei segmentation performed to extract useful features from the image.	• Deep learning approach unexplored. • Dependant completely on hand crafted features.
[33]	Segmentation of sparsely occurring glomeruli in high resolution renal WSI.	• Sparse glomeruli are segmented from renal WSI using cascade CNN, based on pre-existing networks.	• Highly imbalanced classes impact results negatively. • More data augmentation could be beneficial.
[34]	Classify normal histopathology images from 24 classes, taken from different sections of the body.	• ResNet-50 architecture used to classify coloured and greyscale images individually.	• Highly accurate classification due to well separated classes in the dataset. • Small dataset with highly distinct texture features.
[8]	Binary classification of colorectal cancer samples from a dataset.	• ResNet architectures used to classify only greyscale images.	• Small dataset of only 165 images. • May perform worse on unseen samples. • Accuracy 77%

Chapter 3

Proposed Architecture

This chapter explains the approach taken in this study to classify digitized histopathological images of colorectal cancer. The deep learning methods involved in the research and the selected deep convolutional neural network architectures are also detailed. The approach also involves transfer learning. All these aspects are further detailed below.

3.1   Motivation

To find an effective solution to the problem of classifying the dataset of digitized histopathological images, Akhtar introduces an approach to include the color features. In previous studies, that include manual extraction of features to classify colorectal histopathological data, the color features were largely ignored and only the grey channel was used to extract the morphological features such as area, circularity, solidity, perimeter and eccentricity. The useful feature may have been evident in a color space may be lost when the image is flattened into grey-scale. The color spaces detailed in the study are RGB, HSV and YUV [28]. In this study we utilize the RGB components of the images to extract features automatically. To achieve this a specialized deep convolutional neural network architecture called a Residual Network Model is used. In order to achieve swift performance and learning ability from the network a concept of transfer learning is also explored. This concept should enable the network to learn better more useful features faster than a standard/ non pre-trained residual neural network model.

3.2   Residual Networks

Deeper networks are computationally expensive. To enable the efficient use of deeper networks, the number of layers is an essential consideration. How deep does a network need to be in order to solve a particular problem ? The answer to this question is that a dynamic network model is needed. This network should be able to determine if the network needs to be deeper to achieve better results. To achieve this first, we need to understand the degradation problem to gain important insight as to why dynamic layers are required. In the Figure 3.1, Three variations of a simple convolution neural network are shown. Let's assume the network 3.1 (a) performs with a 90% accuracy for the given task. Kaiming et al. have shown that a deeper network leads to performance degradation when applied to the same task. Which means the deeper network is unable to achieve higher accuracy. Simultaneously using more resources as the deeper network consists of more parameters as depicted by the Figure 3.1 (b). The Figure 3.1 (c) shows that, if we replace these layers with identity mapping layers, such that whatever input is given is passed to the next layer as it is. The authors noted that adding identity layers does not add complexity while the resource usage stays similar to the network in Figure 3.1 (a).
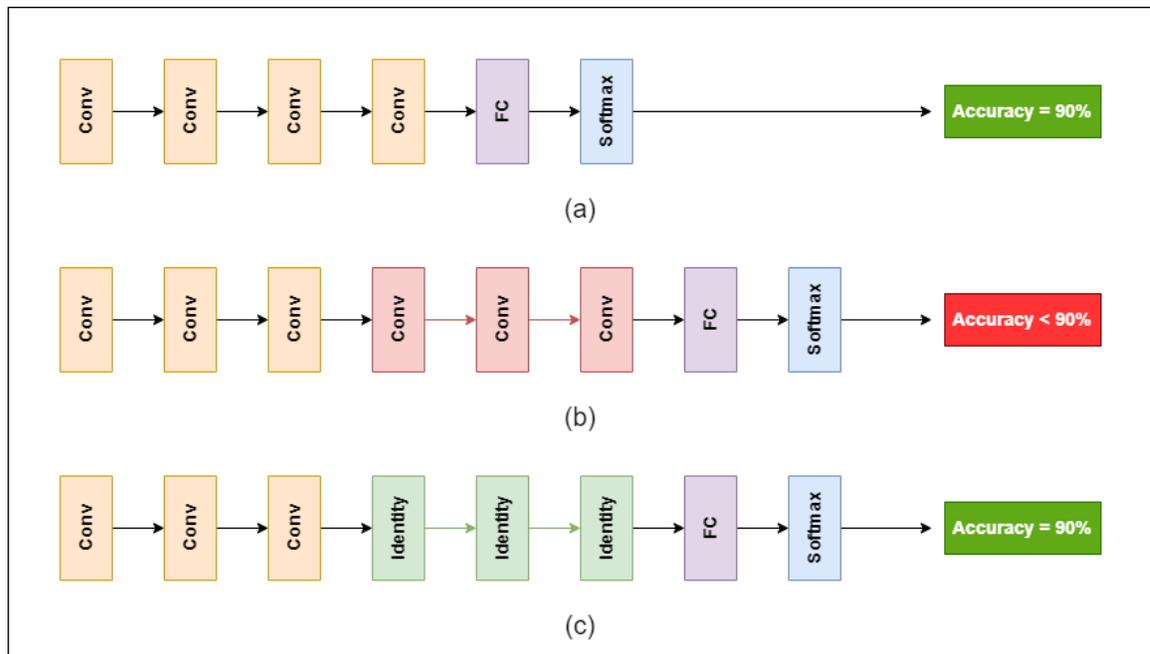
FIGURE 3.1: Degradation Problem in Deep Networks

Kaiming et al. introduced a model called a residual neural network [35]. The main dif-

ference between a residual neural network and a standard neural network is that a standard neural network passes the input x from a function F, written as F(x). The function F(x) then tries to figure out a direct mapping to the output. Hence each layer is involved in the learning process regardless if it has learned anything useful to contribute to the mapping. The structure of a standard neural network block can be observed in the Figure 3.2. The network block consists of two weight layers and a non-linearity between them. In this case were using ReLU activation function to introduce non-linearity in the function. The problem in this situation is that the complexity of a direct mapping increases as the goes deeper. The number of parameters are increased by including every layer of the network. As a result of more resources are consumed in training deeper networks while the deeper network may not be utilized effectively to learn the features.

Compared to the standard neural network, a residual block adds an slight variation to the network block. Figure 3.3 shows a simple residual network block. It has the same architecture as the standard block discussed previously. The authors introduce a extra connection called a residual or skip connection. This connection enables an identity mapping from the input to the output. The residual connection allows the network to enable or disable each residual block based on the information learned. In other words, whether the output of that block is positively contributing to the network. Using this connection, the network can automatically decide how deep it needs to be in order to solve the problem at hand.

Training deeper networks presents another problem of vanishing and exploding gradients. An exploding gradient would mean that the parameters are adjusted by a large value. This impacts the learning of the network significantly. The formula for the weight adjustment in Stochastic Gradient Descent (SGD) is given by:

$$w(t) = w(t\text{-}1) \text{ - } \textbf{\textit{Learning Rate * Gradient}}$$

In the equation above the ***Gradient*** in the derivative of loss ***(dLoss)*** with respect to the weight ***(dw)***. When the gradient is large, this causes the adjustment in the weight to be large. For this reason the network becomes unstable and is unable to learn effectively from the training data. Similarly, the problem of vanishing gradients results in a very

minor adjustment in the weights. This results in the network not being able to learn new information from the training data. This results in slower learning in the network. As a consequence to that, the advantage of having a deep network is lost. Because it will consume a lot of resources while not being able to learn better compared to a less deeper network. Residual networks minimize the issue of exploding and vanishing gradients by keeping the mean of the data steady after each residual block. A residual network achieves that because the last layer inside the residual block is a batch normalization layer, which gives an output normalized between -1 to 1. Having this normalization in each residual block, eliminates this problem which might have impacted the learning of the network negatively.
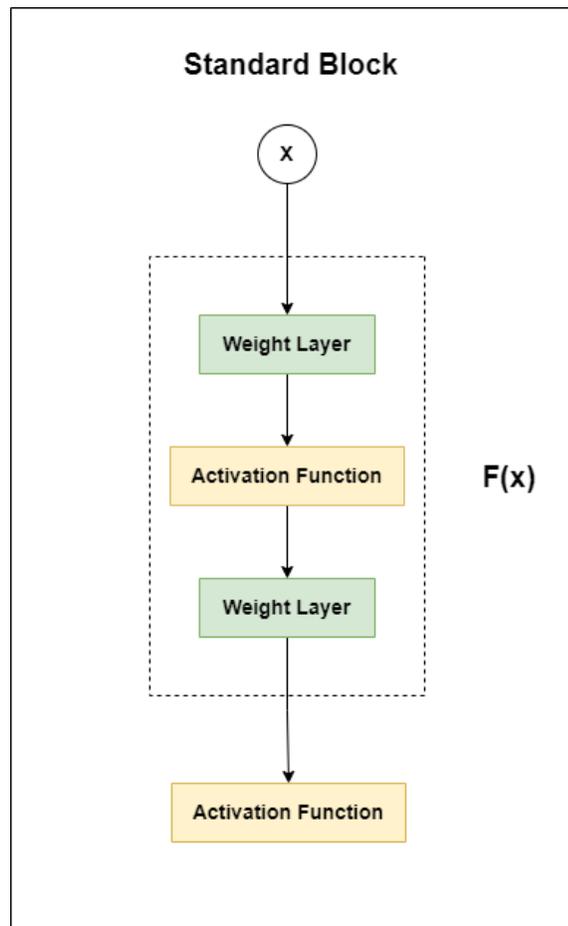
FIGURE 3.2: Standard Neural Network Block

Apart from the simple networks residual networks can also be customized depending on the scenario. In Figure 3.4 we can observe in more detail how each network is structured. Figure 3.4 (a) shows a standard convolutional neural network. It consists of three convolution layers the output of each convolution is batch normalized such that the output
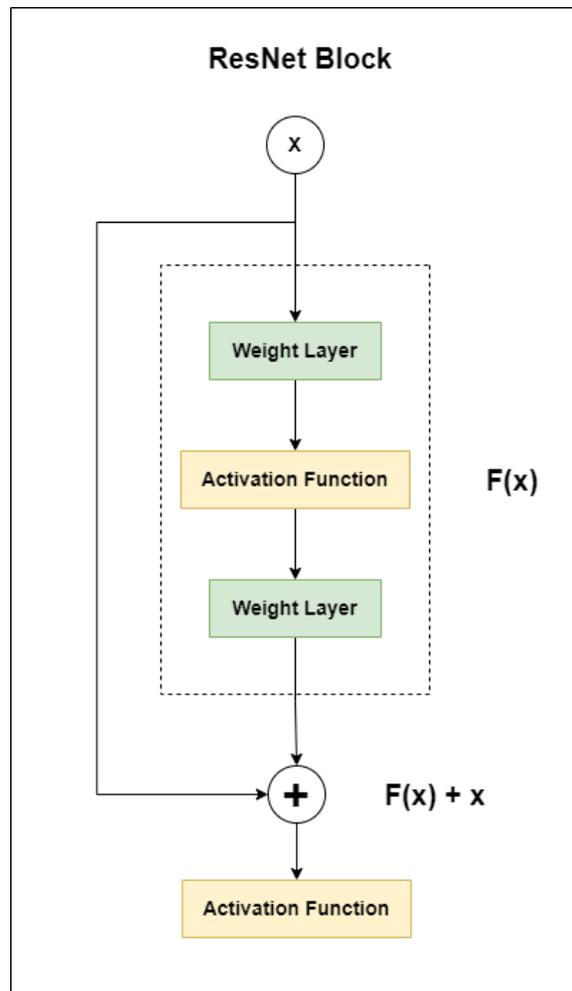
FIGURE 3.3: Residual Neural Network Block

values are between -1 and 1. Figure 3.4 (b) shows a residual network block with a simple skip connection. This gives the network a choice whether to use the residual block or pass the input as output which is also known as an identity mapping. Figure 3.4 (c) shows a customized residual network block. In this network the skip connection consists of a smaller number of operations. This will enable the network to either three convolutions or it can choose to use the skip connection and perform only one convolution step. Instead of an identity mapping the skip connection contains a less computationally expensive operation, Therefore, some useful information can still be learned by the layer.
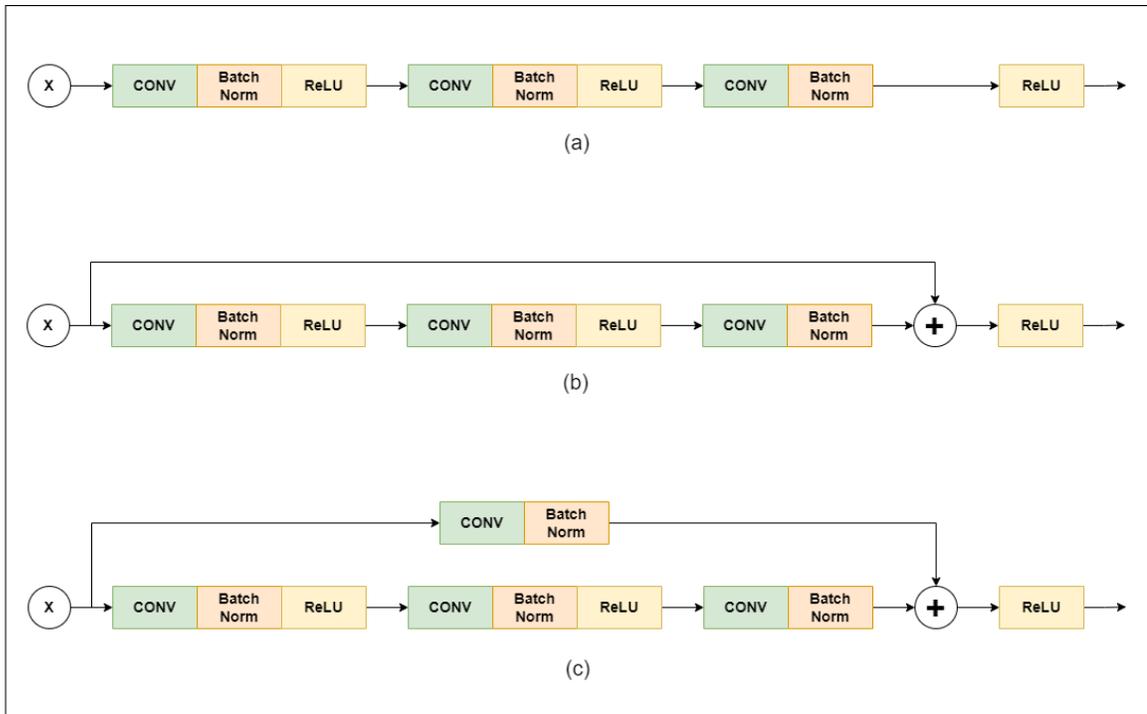
FIGURE 3.4: Convolution Neural Network Block

3.3 Transfer Learning

An important aspect of this study is to evaluate the effect of transfer learning on the problem at hand. The study conducted in [28] uses color and morphological features to classify the colorectal cancer dataset. In this approach, the author uses handcrafted features, which are manually extracted from the data. Although this approach achieves good results. It cannot extract all the unique features from each of the classes. This impacts the overall classification process. To elaborate, We utilize the architecture proposed in [28] on additional datasets from the same domain of histopathological studies, It may not be able to evaluate features that are present in the new unseen data. There may be additional the features belonging to the classes that are ignored by the solution. Only the handcrafted features will be extracted and learned.

In the proposed approach, because of how convolutional neural networks learn, relevant and useful features are extracted automatically by the network from the data. The concept of transfer learning is incorporated as well. Transfer learning as defined by [36] is the application of the knowledge gained by solving one problem and utilizing that knowledge to

solve a different but related problem. For example a neural networks trained on a gastric histopathology dataset is used to train a classifier for thyroid histopathology images in [37]. As these classification challenges are related and of a similar domain the potential to learning better and more relevant features is higher as compared to a network training from scratch. In other cases totally unrelated domains might also benefit from this approach. This is similar to defining a least complex problem to approach a similar or higher complex problem. For example authors of [38] evaluate the effect of intra-domain transfer learning. They study the effect of transfer learning from a number of different domains such as finance, speech, medicine and seismology.

In this study the Residual network used is available as a pre-trained network, already trained on the ImageNet dataset. The ImageNet dataset consists of 14 million sample of images belonging to 1000 individual classes [39]. The last layer of the network is replaced as the number of classes in the colorectal cancer whole slide image dataset is different to that of the ImageNet. The impacts of the transfer learning are later assessed in the next section.

3.4   Proposed Architecture

To explore the potential of Residual Networks in classifying the colorectal cancer dataset, we conduct a total of four experiments. Each individual step of the whole procedure can be seen in the Fig. 3.5. The basic steps include, the collection of dataset, the pre-processing performed on the image dataset, The types of data augmentations applied to add increased variability to the data and the application of two ResNet classification models with two different variations. The details of each step are explained further under their respective headings.

3.4.1   Data Acquisition, Preprocessing and Augmentation

The dataset consists of two sets of image patches of colorectal hematoxylin & eosin (H&E) biopsy images. They contain both cancerous and normal tissue patches. The size of each patch is 224x224. All of the image patches are non-overlapping. The dataset is color nor-
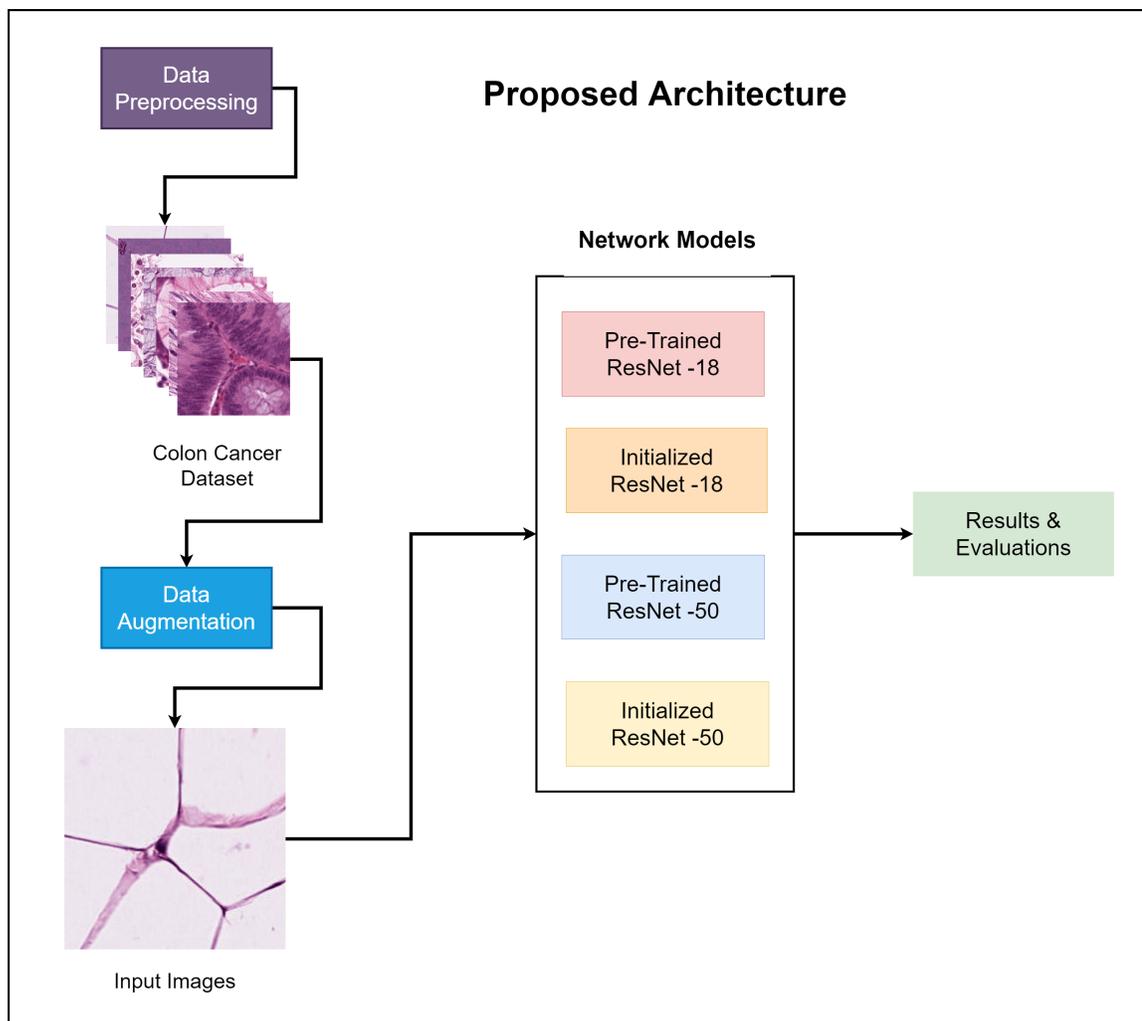
FIGURE 3.5: Proposed Architecture

malized. The type of normalization used is known as Macenko's method of normalization. The datatset has an open access licence and is publicly available at [40]. Traditional data balancing techniques will increase the size of the dataset significantly, resulting in more time and resource consumption. An intelligent data technique using mirror and rotation transformations was introduced. Simple mirroring and rotation of patch allows increased learn-ability in various orientations. To achieve this without increasing the size of the dataset the transformations are done at each epoch. The algorithm will receive a different variation of the dataset each time. Image patches will be flipped and rotated randomly before being given as input. Using this technique, learning of the features can be improved while keeping the resource load to a minimal.

3.4.2   Application of ResNet Models

The application of ResNet models are done in four instances. The dataset prepared in the previous step is provided as input to each of the four residual network models. The first model is a ResNet-18 model that consists of an architecture with eighteen deep convolution layers. The same network is used with a variation in transfer learning. First the architecture has its weights initialized. This means that the network starts from scratch and does not have any weights or learning information stored. The second model of the ResNet-18 model is applied directly using preserved weights from training on the ImageNet dataset. The same steps are taken for the second architecture as well. The second architecture is a ResNet-50 model, it consists of a forty-eight deep convolutional layers. This network is also applied with the same variation as the ResNet-18 model namely, a pre-trained and a non pre-trained model. Similarly, the weights of the pre-trained ResNet-50 model are from the ImageNet dataset.

The results of the experiments are discussed in Chapter 4. It goes into details of the parameters of each experiments and the respective results are reported.

3.4.3   Loss Function

The loss function chosen for the experiments is the cross entropy loss given by:

$$Loss = -\sum_{c=1}^{M} y_{o,c} \log(p_{o,c}) \tag{3.1}$$

There are multiple reasons for selecting Cross Entropy Loss. It is recommended in multi-class classification and is used to update the weights in the training phase. Cross entropy loss is particularly useful when you have an unbalanced training set. [41]

3.4.4   Assessment Metrics

To evaluate the performance of the network model, it is assessed based on the following evaluation metrics. The results of all the network variations are discussed later.

- **Precision:** It is a measure of the correctly predicted positive patterns from the total positive results.

$$Precision = \frac{TP}{TP + FP} \qquad (3.2)$$

- **Accuracy:** It is the measure of the ratio between the correct predictions and the total number predictions.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (3.3)$$

- **Recall:** Used to measure the correctly classified positive patterns.

$$Recall = \frac{TP}{TP + FN} \qquad (3.4)$$

- **F1-Score:** Takes into account the precision and recall of a network and is the ultimate of accuracy of the network.

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \qquad (3.5)$$

Where the terms TP, TN, FP and FN are defined as:

- **True positive (TP):** A test result that correctly indicates the presence of a condition or characteristic.

- **True negative (TN):** A test result that correctly indicates the absence of a condition or characteristic.

- **False positive (FP):** A test result which wrongly indicates that a particular condition or attribute is present.

- **False negative (FN):** A test result which wrongly indicates that a particular condition or attribute is absent.

Chapter 4

Experiment and Evaluation

This chapter describes the details of experiments, the procedure of the data acquisition and the results of the conducted experiments. Further it provides analysis of the results and draws a conclusion from the various findings.

The results discussed in this section are from four experiments. Each experiment is conducted on the same main dataset NCT-CRC-HE-100K. The parameters of each experiment are explained sequentially. The second dataset CRC-VAL-HE-7K is taken as an independent validation set. This means that the dataset doesn't have any overlapping image patches from the main dataset. This approach will result in a more robust validation. After each epoch the validation is performed and the training and testing accuracy are discussed in respective sections.

4.1   Dataset Collection

The dataset consists of two sets of image patches of colorectal hematoxylin & eosin (H&E) biopsy images. They contain both cancerous and normal tissue patches. All of the image patches are non-overlapping. The dataset was color normalized before distribution. The type of normalization used is known as Macenko's method of normalization. This method was specifically developed for the color normalization of hematoxylin & eosin (H&E) stained histopathogical biopsy images [9]. The details of the Macenko's method and the dataset

are explained further 4.1.1. The datatset used in this study has an open access licence and is publicly available at [40].

4.1.1 Macenko's Method

The method described by the authors in [9] is one of the most well known methods in order to overcome the variations that occur during the hematoxylin and eosin (H&E) staining test. These variations are a result of variable conditions under which the slides were prepared. The normalization of the slides allows for an improved analysis. Figure 4.1 shows two raw H&E staining samples **(a)** shows a slide with low saturation and contrast. **(b)** shows a slide with higher saturation and contrast.
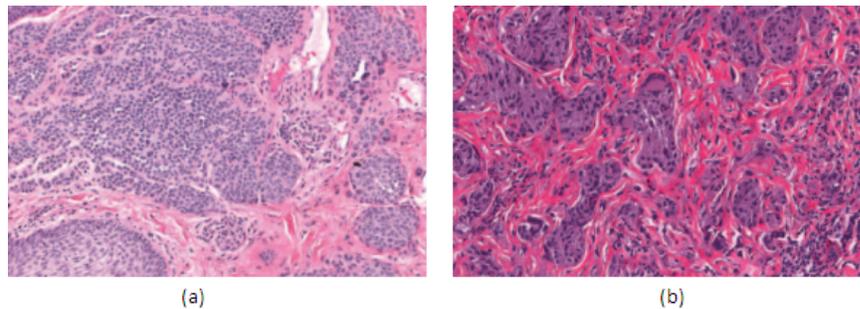
FIGURE 4.1: Shows a sample H&E Slides before applying Macenko's Method [9]

Similarly Figure 4.2 exhibits the result after applying the Macenko's method of normalization on both slides in Figure 4.1. Both the slides having different appearances are now similar to each other in terms of color features.
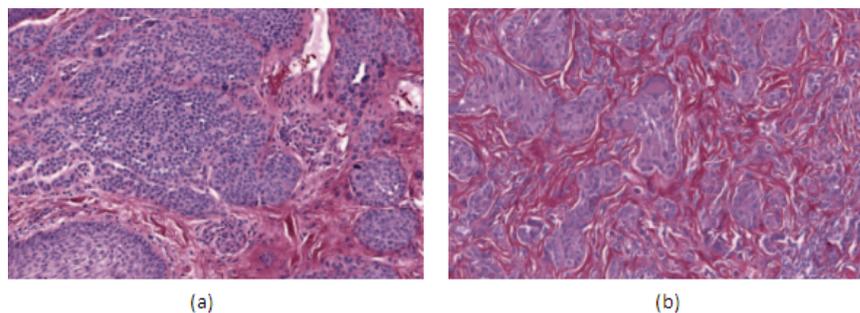
FIGURE 4.2: Shows a sample H&E Slides after applying Macenko's Method [9]

4.1.2   Dataset Description

The dataset provided has two parts. The details of all the classes in the datatset are given in Table 4.1 and 4.2. The first dataset *NCT-CRC-HE-100K* consists of 100,000 non overlapping whole slide image patches of size 224×224, generated from biopsies of 86 patients. The dataset is labelled into 9 classes. *CRC-VAL-HE-7K* is provided as a validation set that consists of 7180 images of size 224×224, generated from biopsies of 50 patients. These images are non-overlapping with *NCT-CRC-HE-100K* dataset. The dataset consists of 9 individual classes. a sample image of each class is given in the Figure 4.3.
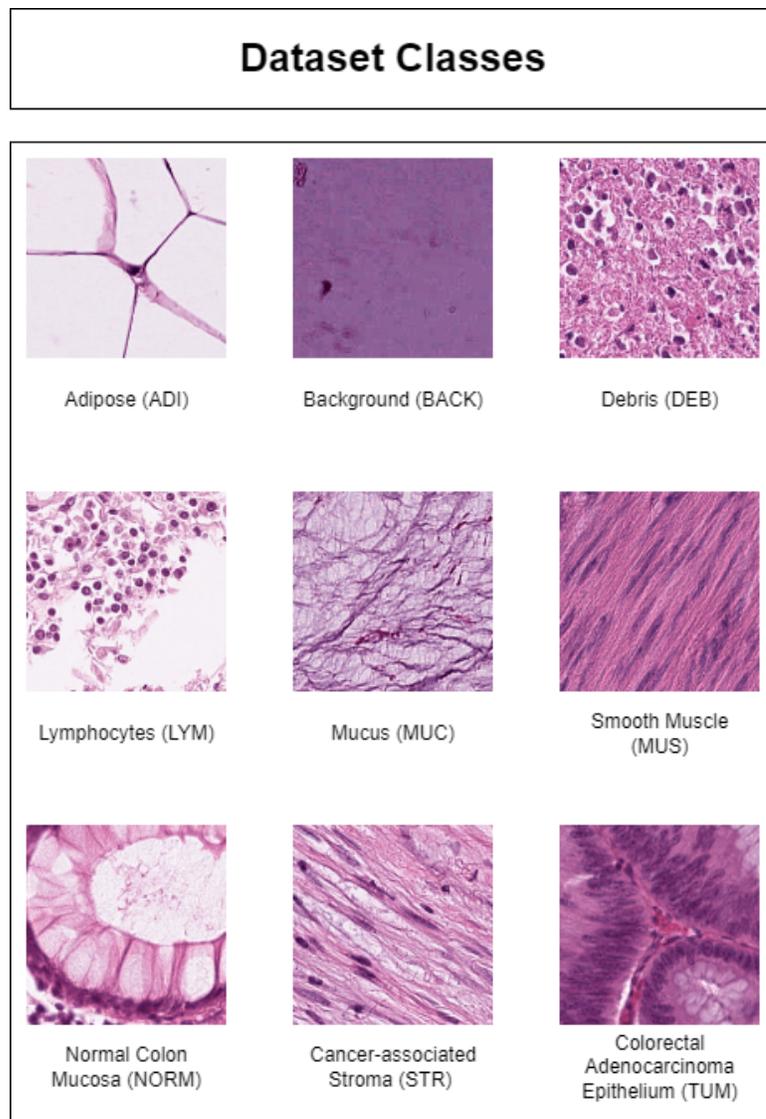
FIGURE 4.3: Sample images from each of the 9 classes in the dataset

Dataset 1 : NCT-CRC-HE-100K

This is the first part of the full dataset. It consists of a set of 100,000 non overlapping images of size 224×224. The data is labbeled and distributed across 9 classes. The details of the number of images in each class and cancerous details are given in the Table 4.1.

TABLE 4.1: NCT-CRC-HE-100K Dataset Details

Class	Samples	Cancerous
Adipose (ADI)	10,407	No
Background (BACK)	10,566	No
Debris (DEB)	11,512	No
Lymphocytes (LYM)	11,557	No
Mucus (MUC)	8,896	No
Smooth muscle (MUS)	13,536	No
Normal colon mucosa (NORM)	8,763	No
Cancer-associated stroma (STR)	10,446	Yes
Colorectal adenocarcinoma epithelium (TUM)	14,317	Yes

Dataset 2 : CRC-VAL-HE-7K

This is the second part of the full dataset. it can be used as an independent validation or testing set. It consists of a set of 7,180 non overlapping images of size 224×224. The data is labbeled and distributed across 9 classes. The details of the number of images in each class and cancerous details are given in the Table 4.2

TABLE 4.2: CRC-VAL-HE-7K Dataset Details

Class	Samples	Cancerous
Adipose (ADI)	1,338	No
Background (BACK)	847	No
Debris (DEB)	339	No
Lymphocytes (LYM)	834	No
Mucus (MUC)	1,035	No
Smooth muscle (MUS)	592	No
Normal colon mucosa (NORM)	714	No
Cancer-associated stroma (STR)	421	Yes
Colorectal adenocarcinoma epithelium (TUM)	1,233	Yes

4.2   Experiments

To evaluate the proposed method, a series of four experiments were concluded. The main purpose of the study is to evaluate the classification of the dataset as well as the effects of using a pre-trained ResNet models. The pre-trained models have weights trained on the ImageNet dataset. The ImageNet dataset consists of 14 million sample of images belonging to 1000 individual classes [39]. The last layer of the network is replaced to suit the number of classes in the NCT-CRC-HE-100K dataset. Each experiment was conducted at a 70:30 train test split. The experiments were limited to 50 epochs to enable a fair comparison between the models. The last 10 epochs of each experiment are then compared for evaluation. According to [42], for most cases, a small batch size of 32 is sufficient. After initial runs a suitable batch size of 64 was selected. The batch size is kept constant for all the experiments. A dynamic approach is applied to achieve a steady rate of learning. A learning rate step scheduler is used to dynamically lower the learning rate on plateau. The initial learning rate is 0.001 and is reduced at a different number epochs for each model. The optimizer is standard for all the experiments. ADAM optimizer is recommended for most applications as it has a faster computation time and requires fewer parameters for tuning it. ADAM optimizer is an extension of the stochastic gradient descent [43]. This setup provides a equal starting ground for all the experiments. This is done partially to learn the effect of the transfer learning. The initialized weights should have an impact on the learning and performance of the Residual architectures. The seed value for all the experiments was kept same during all experiments to achieve reproducibility consistency in experimentation.

Experimentation Platform:
All experiments were conducted on Compute Canada server with NVIDIA Tesla P100 GPUs and a desktop computer with a 2.90GHz Intel Core i7-10700 CPU, 16GB of DDR4 RAM and a dedicated NVIDIA RTX 2070 graphics card with 8GB of VRAM. All implementation is performed in PyTorch.

4.2.1   Experiment 1: ResNet-18

The first experiment was conducted using a ResNet-18 model. The weights of the network are initialized. This results are shown in the Table 4.3. We can observe the highest accuracy

is achieved at epoch 46. The highest training accuracy achieved is 84.6% and validation accuracy of 79.6% respectively.

TABLE 4.3: Accuracy table of the ResNet-18 Model

Epoch	Training	Validation
41	0.82833	0.76812
42	0.84583	0.76886
43	0.84739	0.78750
44	0.84687	0.79492
45	0.84958	0.78710
46	**0.84635**	**0.79648**
47	0.84635	0.79062
48	0.84791	0.79570
49	0.82984	0.78671
50	0.82739	0.79453

4.2.2 Experiment 2: ResNet-18 Pre-trained

The second experiment was conducted using a pre-trained ResNet-18 model. For this experiment the weights of the network are not initialized. Instead the pre-trained weights on the ImageNet dataset are initialized. This results are shown in the Table 4.4. From the table we can observe the highest accuracy is achieved at epoch 50. The highest training accuracy achieved is 85.9% and validation accuracy of 81.4% respectively.

TABLE 4.4: Accuracy table of the pre-trained ResNet-18 Model

Epoch	Training	Validation
41	0.82843	0.79804
42	0.85843	0.79140
43	0.85479	0.79453
44	0.85997	0.79062
45	0.85895	0.77226
46	0.85843	0.78125
47	0.85143	0.79062
48	0.85818	0.79570
49	0.85656	0.81062
50	**0.85895**	**0.81406**

4.2.3   Experiment 3: ResNet-50

The third experiment was conducted using a ResNet-50 model. The weights of the network are initialized. This results are shown in the Table 4.5. From the table we can observe the highest accuracy is achieved at epoch 50. The highest training accuracy achieved is 88.5% and validation accuracy of 81.2% respectively.

TABLE 4.5: Accuracy table of the ResNet-50 Model

Epoch	Training	Validation
41	0.84387	0.77808
42	0.84023	0.78121
43	0.84541	0.77730
44	0.84439	0.75894
45	0.84387	0.76793
46	0.83687	0.77730
47	0.84439	0.78238
48	0.86200	0.79730
49	0.86439	0.80074
50	**0.88491**	**0.81246**

4.2.4   Experiment 4: ResNet-50 Pre-trained

The last experiment was conducted on a pre-trained ResNet-50 model. Again for this experiment the weights of the network are not initialized. Instead the pre-trained weights on the ImageNet dataset are initialized. This results are shown in the Table 4.4. From the table we can observe the highest accuracy is achieved at epoch 47. The highest training accuracy achieved is 91.9% and validation accuracy of 89.5% respectively.

4.2.5   Experiment 5: Testing of All Models

After the initial experimentation, the weights associated with the best results for each experiment were run on the testing set and the accuracies are compared. Fig. 4.4 and Table. 4.8 show the accuracies of the models. During the testing the best accuracy was achieved by the pretrained ResNet-50 model.

TABLE 4.6: Accuracy table of the Pre-trained ResNet-50 Model

Epoch	Training	Validation
41	0.90947	0.85781
42	0.90916	0.84828
43	0.91843	0.86867
44	0.91947	0.85867
45	0.90958	0.89593
46	0.90916	0.89062
47	**0.91958**	**0.89531**
48	0.91437	0.89062
49	0.91758	0.89187
50	0.91583	0.88453

4.3   Results & Comparison

TABLE 4.7: Accuracy table of ResNet Models

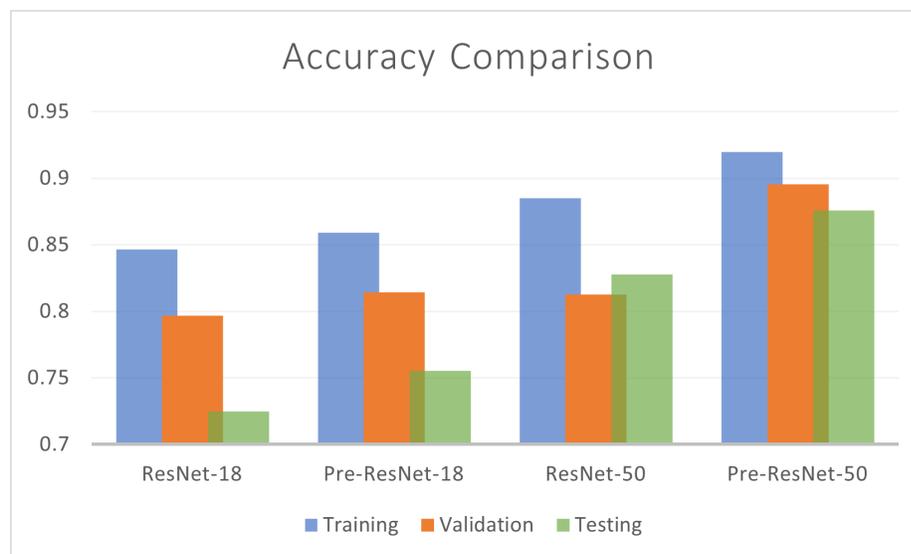| Algorithm	Training	Validation	Testing
ResNet-18	0.84635	0.79648	0.72488
Pre-ResNet-18	0.85895	0.81406	0.75518
ResNet-50	0.88491	0.81246	0.82784
Pre-ResNet-50	0.91958	0.89531	0.88582

FIGURE 4.4: Comparing the Training, Validation and Testing Accuracy of each model

After the initial experimentation the pre-trained ResNet-50 achieved the best overall results. The model converges faster than the initialized ResNet-50. Both the models have identical architectures. Hence the initialized model is less efficient as it requires more time

and resources to achieve similar results.

The first comparison is drawn between the ResNet-18 and its pretrained variation. We can observe the accuracy of each from the Table 4.3 and Table 4.4. The training accuracy of the pretrained ResNet-18 drops after reaching a peak value of 84.6%. The validation accuracy on the other hand reaches a maximum of 79.6%. The testing accuracy attained is 72.4%. The main difference between the networks is that the pre-trained network has the weights preserved from the ImageNet dataset. This should facilitate the learning process to converge quicker to a better accuracy value. In this case the difference is evident in the testing accuracy. The pre-trained network attains a better testing accuracy at 75.5% it means the network weights were adapted to the current classification problem.

The next comparison is drawn between the ResNet-18 pre-trained model and the simple initialized ResNet-50. The accuracies can be observed from the Table 4.4 and Table 4.5. Both the network models achieved a similar accuracy regardless of the second network being substantially deeper than the first one. This can happen because the initialized ResNet-50 model takes a longer time to adapt to the data. The existing weights in the ResNet-18 pretrained model are responsible for quicker convergence. They both reach a similar validation accuracy close to 81%. The difference comes in the testing accuracy shown in the Table 4.8. The pre-trained ResNet-18 achieves a testing accuracy of 75.5% while the ResNet-50 model achieves 82.8%. The Resnet-50 model is better at generalizing. The new unseen samples are better classified with the deeper model. The availability of more trainable parameters in the ResNet-50 are also a factor. The best results are achieved using the ResNet-50 pre-trained model. The training and validation accuracy are close to 92% and 89.5% respectively. This is a result of better adapting to the data because of the pre-existing weights the process is accelerated. The ResNet-50 model is better at learning useful features of the data because of the additional layers. The features extracted at more depths are helpful. The model trained as a result of this is generalized and robust.

4.3.1   Comparison with Previous Work

The multiple level framework proposed by Zhou et al. [10], to classify histopathology dataset. The Cell-Level framework achieved values 73.73%, 98.73%, 74.44% and 84.88%

while the Pretrained ResNet-50 achieved 88.58%, 92.04%, 81.86%, 86.65% for Accuracy, Precision, Recall and F1-Score respectively. It is important to note the the dataset is different among the two results being compared. Overall we can deduce that the Residual Network performs better. It has similar precision and better accuracy. This might be due to the variation in the size of the dataset and the number of classes as the mentioned work is a binary classification problem.

TABLE 4.8: Accuracy table of ResNet Models

Algorithm	Accuracy	Precision	Recall	F1-Score
Cell Level Framework [10]	0.7373	0.9873	0.7444	0.8488
Pre-ResNet-50	0.8858	0.9204	0.8186	0.8665
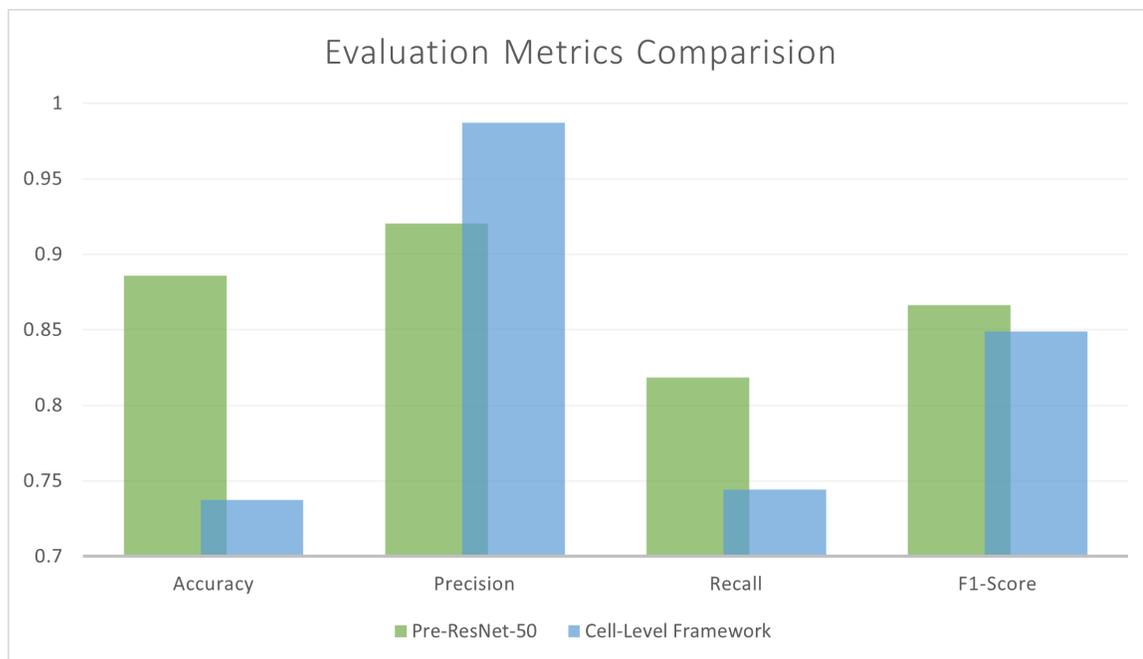
FIGURE 4.5: Comparing the Accuracy, Precision, Recall and F-1 Score with Cell Level Framework [10]

Khvostiko et al. [44] use another deep convolutional neural network to classify multiple histopathological datasets. One of the dataset used is the *NCT-CRC-HE-100K* dataset included in this research. The authors used a pre-existing architecture known as DenseNet-121 consisting of 120 deep convolution layers. This network is substantially deeper than the Residual Network Architectures utilized in this research. The authors claim to have achieved a training accuracy of 91.9% in 20 epochs. Whereas the ResNet-50 architecture in this study reached a similar training accuracy of 91.5%. The DenseNet-121 was retrained

two additional times on the *PATH-DT-MSU* dataset. The dataset consists of H&E whole slide images from colon and stomach biopsies and can be found at [45]. The dataset consists of 5 classes for the stomach and 2 classes for colorectal cancer dataset. After three phases of training, the network achieved 92.6% accuracy. Both the models achieved very similar accuracies on the *NCT-CRC-HE-100K* dataset. The convolutions are performed at various depths in both networks. Both networks are performing similarly while the DenseNet-121 is a much deeper network with more parameters. This is the effect of residual learning taking advantage of optimizing the layers to better suit the data rather than using the full depth of the network.

Chapter 5

Conclusion and Future Works

Haematoxylin and Eosin (H&E) stain tests prove beneficial in the detection of various types of diseases. It is one of the most important tests when it comes to detection of cancerous tissue. Due to the disparity in the growth of each particular type of cancerous tissue, prognosis of the disease becomes a challenging task. By regularly inspecting samples from biopsies of patients, the state and spread of cancerous tissue can be monitored and predicted to a certain degree. Through continuous monitoring the doctor can put forward preemptive measures based on the history of each individual patient. Early detection of cancerous growths reduces the chances of uncontrolled spread of the disease. As a result of histopathological monitoring of cancerous tissue, treatments can be adapted to better suit each particular case, improve the quality of life and the overall life expectancy of the patient.

This research proposes a deep convolutional approach to classify a large histopathological colorectal cancer dataset. Two variations of two different network models were applied to the dataset. The effect of transfer learning was also analyzed. The two network models applied are ResNet-18 and ResNet-50. Both the networks were applied in two variations namely pre-trained and initialized. The pre-trained variation has weights preserved from ImageNet dataset. All the networks were applied to the *NCT-CRC-HE-100K* dataset with a 70:30 train test split. The validation was done using the *CRC-VAL-HE-7K* dataset. All the experiments were performed under similar and reproducible environment.

The ResNet-18 models performed well of a large dataset of 100,000 pathches. The pretrained version of the ResNet-18 model attained an accuracy of 75.51%, overall 3% better than the initialized ResNet-18 model. Similarly the ResNet-50 models achieved an accuracy of 82.78% and 88.58% for initialized and pre-trained models respectively. Over all the pre-trained ResNet-50 performed 6% better under the same conditions. This might not seem like much improvement considering that the larger network is more than twice the size of the first one. But taking into consideration the complexity of this classification problem it is significant. The effect of the transfer learning is evident in the experimentation. This can be observed from the results.

The best performing model in this study is pre-trained ResNet-50. A comparison is drawn with similar state of the art research in histopathology classification. The same dataset used used in this study was classified using DenseNet-121 model by Khvostiko et al. [44]. The DenseNet-121 was trained multiple times to adapt completely to the dataset. The DenseNet-121 model was also trained on additional dataset called *PATH-DT-MSU* in a third round. Compared to our pre-trained ResNet-50 model, the DenseNet-121 model performed better. The ResNet model attained accuracy of 91.5% and the DenseNet-121 performed similarly at 92.6% accuracy. Overall the difference is negligible considering that the DenseNet is substantially deeper and went through multiple rounds of training and also trained on an additional independent dataset.

This thesis successfully improved the cell level framework of the proposed by the authors in [10]. Our ResNet model performs better in every metric described by the authors except for the precision, which is 4% less than the cell level framework. The results show how powerful residual learning can be. The addition of transfer learning with the ResNet model allows the model to converge better than an initialized model. The same ResNet model can be improved with additional rounds of training on unseen datasets from the same domain of research. This would show a significant improvement in the learn-ability of the model.

In the future the same ResNet-50 model which consists of 23 million parameters, can be trained on additional datasets to improve the overall performance and generalization of the model. As the model gets trained on more histopatological datasets the transfer learning will make the model more robust. Convolutional parameters such as stride length and window size can also be adjusted for further experimentation. More depths of the ResNet models

can be explored in the future to find an optimal solution for quicker inference times and over all performance. The application of Residual Networks of Residual Networks (RoR) described in [29] can also be considered to optimize ResNet models.

To summarize, the best performing model is the ResNet-50 model. It achieves the best results among all the models and variations of the models tested in this research. The model is compared to two studies in the same field. It improves upon the cell level framework by a big margin and is comparable to a mush deeper DenseNet-121 model in terms of performance and accuracy.

Finally, this model is specifically adapted to histopatological colorectal datasets and can be evaluated for other types of histopatological datasets such as abdominal or renal patches. This can be considered as a future work of this research.

Bibliography

[1] Joseph Weizenbaum. Computer power and human reason: From judgment to calculation. 1976.

[2] CHI Software. Supervised vs. unsupervised machine learning, May 2019. URL `https://chisoftware.medium.com/supervised-vs-unsupervised-machine-learning-7f26118d5ee6`.

[3] Jessica Pedersen. What is ai?, May 2020. URL `https://www.resquared.com/blog/what-is-ai`.

[4] Mei Wang and Weihong Deng. Deep face recognition: A survey. *Neurocomputing*, 429: 215–244, 2021.

[5] They found colon polyps: Now what?, Feb 2022. URL `https://www.health.harvard.edu/diseases-and-conditions/they-found-colon-polyps-now-what`.

[6] Antonio Galvano, Aurelia Ada Guarini, Valerio Gristina, Nadia Barraco, Maria La Mantia, Marta Castiglia, and Antonio Russo. *Colorectal Cancer: Metastatic Disease*, pages 617–633. Springer International Publishing, Cham, 2021. ISBN 978-3-030-56051-5. doi: 10.1007/978-3-030-56051-5_38.

[7] R Krithiga and P Geetha. Breast cancer detection, segmentation and classification on histopathology images analysis: a systematic review. *Archives of Computational Methods in Engineering*, 28(4):2607–2619, 2021.

[8] Devvi Sarwinda, Radifa Hilya Paradisa, Alhadi Bustamam, and Pinkie Anggia. Deep learning in image classification using residual network (resnet) variants for detection of colorectal cancer. *Procedia Computer Science*, 179:423–431, 2021.

[9] Marc Macenko, Marc Niethammer, James S Marron, David Borland, John T Woosley, Xiaojun Guan, Charles Schmitt, and Nancy E Thomas. A method for normalizing histology slides for quantitative analysis. In *2009 IEEE international symposium on biomedical imaging: from nano to macro*, pages 1107–1110. IEEE, 2009.

[10] Changjiang Zhou, Yi Jin, Yuzong Chen, Shan Huang, Rengpeng Huang, Yuhong Wang, Youcai Zhao, Yao Chen, Lingchuan Guo, and Jun Liao. Histopathology classification and localization of colorectal cancer using global labels by weakly supervised deep learning. *Computerized Medical Imaging and Graphics*, 88:101861, 2021.

[11] Hsiang Chen, Rolf T Wigand, and Michael S Nilan. Optimal experience of web activities. *Computers in human behavior*, 15(5):585–608, 1999.

[12] Hafiz Muhammad Luqman and Mubeen Zaffar. Chess brain and autonomous chess playing robotic system. In *2016 International Conference on Autonomous Robot Systems and Competitions (ICARSC)*, pages 211–216. IEEE, 2016.

[13] Harry Henderson. *Alan Turing*. Infobase Publishing, 2011.

[14] E Alpaydin. Introduction to machine learning fourth edition, 2020.

[15] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[16] Michael A Nielsen. *Neural networks and deep learning*, volume 25. Determination press San Francisco, CA, USA, 2015.

[17] E Roy Davies. *Computer vision: principles, algorithms, applications, learning*. Academic Press, 2017.

[18] Ming-Hsuan Yang, David J Kriegman, and Narendra Ahuja. Detecting faces in images: A survey. *IEEE Transactions on pattern analysis and machine intelligence*, 24(1):34–58, 2002.

[19] Colon and rectal polyps, digestive and liver health university of michigan, Aug 2020. URL https://www.uofmhealth.org/conditions-treatments/digestive-and-liver-health/colon-and-rectal-polyps.

[20] M Titford. The long history of hematoxylin. *Biotechnic & histochemistry*, 80(2):73–78, 2005.

[21] Mary Wong, Joseph Frye, Stacey Kim, and Alberto M Marchevsky. The use of screencasts with embedded whole-slide scans and hyperlinks to teach anatomic pathology in a supervised digital environment. *Journal of pathology informatics*, 9, 2018.

[22] Lee AD Cooper, Alexis B Carter, Alton B Farris, Fusheng Wang, Jun Kong, David A Gutman, Patrick Widener, Tony C Pan, Sharath R Cholleti, Ashish Sharma, et al. Digital pathology: Data-intensive frontier in medical imaging. *Proceedings of the IEEE*, 100(4):991–1003, 2012.

[23] Zitong Wan, Rui Yang, Mengjie Huang, Nianyin Zeng, and Xiaohui Liu. A review on transfer learning in eeg signal analysis. *Neurocomputing*, 421:1–14, 2021.

[24] Muhammet Fatih Aslan, Muhammed Fahri Unlersen, Kadir Sabanci, and Akif Durdu. Cnn-based transfer learning–bilstm network: A novel approach for covid-19 infection detection. *Applied Soft Computing*, 98:106912, 2021.

[25] Muhammad EH Chowdhury, Tawsifur Rahman, Amith Khandakar, Rashid Mazhar, Muhammad Abdul Kadir, Zaid Bin Mahbub, Khandakar Reajul Islam, Muhammad Salman Khan, Atif Iqbal, Nasser Al Emadi, et al. Can ai help in screening viral and covid-19 pneumonia? *IEEE Access*, 8:132665–132676, 2020.

[26] Irfan Ullah Khan and Nida Aslam. A deep-learning-based framework for automated diagnosis of covid-19 using x-ray images. *Information*, 11(9):419, 2020.

[27] Rajesh Mehra et al. Breast cancer histology images classification: Training from scratch or transfer learning? *ICT Express*, 4(4):247–254, 2018.

[28] Sameer Akhtar Syed. *Color and morphological features extraction and nuclei classification in tissue samples of colorectal cancer*. PhD thesis, University of Windsor (Canada), 2020.

[29] Ke Zhang, Miao Sun, Tony X Han, Xingfang Yuan, Liru Guo, and Tao Liu. Residual networks of residual networks: Multilevel residual networks. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(6):1303–1314, 2017.

[30] Masayuki Tsuneki and Fahdi Kanavati. Deep learning models for poorly differentiated colorectal adenocarcinoma classification in whole slide images using transfer learning. *Diagnostics*, 11(11):2074, 2021.

[31] Ruiwei Feng, Xuechen Liu, Jintai Chen, Danny Z Chen, Honghao Gao, and Jian Wu. A deep learning approach for colonoscopy pathology wsi analysis: accurate segmentation and classification. *IEEE Journal of Biomedical and Health Informatics*, 25(10):3700–3708, 2020.

[32] Kuan-Song Wang, Gang Yu, Chao Xu, Xiang-He Meng, Jianhua Zhou, Changli Zheng, Zhenghao Deng, Li Shang, Ruijie Liu, Shitong Su, et al. Accurate diagnosis of colorectal cancer based on histopathology images using artificial intelligence. *BMC medicine*, 19 (1):1–12, 2021.

[33] Michael Gadermayr, Ann-Kathrin Dombrowski, Barbara Mara Klinkhammer, Peter Boor, and Dorit Merhof. Cnn cascades for segmenting sparse objects in gigapixel whole slide images. *Computerized Medical Imaging and Graphics*, 71:40–48, 2019.

[34] Muhammed Talo. Automated classification of histopathology images using transfer learning. *Artificial intelligence in medicine*, 101:101743, 2019.

[35] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[36] Jeremy West, Dan Ventura, and Sean Warnick. Spring research presentation: A theoretical foundation for inductive transfer. *Brigham Young University, College of Physical and Mathematical Sciences*, 1(08), 2007.

[37] Vijaya Gajanan Buddhavarapu et al. An experimental study on classification of thyroid histopathology images using transfer learning. *Pattern Recognition Letters*, 140:1–9, 2020.

[38] Erik Otović, Marko Njirjak, Dario Jozinović, Goran Mauša, Alberto Michelini, and Ivan Stajduhar. Intra-domain and cross-domain transfer learning for time series data—how transferable are the features? *Knowledge-Based Systems*, 239:107976, 2022.

[39] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115 (3):211–252, 2015.

[40] Jakob Nikolas Kather, Niels Halama, and Alexander Marx. 100,000 histological images of human colorectal cancer and healthy tissue, April 2018. URL `https://doi.org/10.5281/zenodo.1214456`.

[41] Zhilu Zhang and Mert Sabuncu. Generalized cross entropy loss for training deep neural networks with noisy labels. *Advances in neural information processing systems*, 31, 2018.

[42] Dominic Masters and Carlo Luschi. Revisiting small batch training for deep neural networks. *arXiv preprint arXiv:1804.07612*, 2018.

[43] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[44] Alexander Khvostikov, Andrey Krylov, Ilya Mikhailov, Pavel Malkov, and Natalya Danilova. Tissue type recognition in whole slide histological images. In -   , volume 31, pages 496–507, 2021.

[45] Alexander Khvostikov, Andrey Krylov, Ilya Mikhailov, Nina Oleynikova, Olga Kharlova, Natalia Danilova, and Pavel Malkov. Path-dt-msu dataset, Aug 2019. URL `https://imaging.cs.msu.ru/en/research/histology/path-dt-msu`.

——————————————————————————————————————- VITA AUCTORIS

——————————————————————————————————————-

Vita Auctoris

NAME: Ali Hassan

PLACE OF BIRTH: Islamabad, Pakistan

YEAR OF BIRTH: 1994

EDUCATION: University of Windsor, Windsor, Ontario

                          2019-2022 M.Sc.