Classification of Sub-type of Lymphoma using Deep Learning

MSc Research Project
Data Analytics

Prasad Balasaheb Thorat
Student ID: x18185711

School of Computing
National College of Ireland

Supervisor: Dr. Christian Horn

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Prasad Balasaheb Thorat
Student ID:	x18185711
Programme:	Data Analytics
Year:	2020
Module:	MSc Research Project
Supervisor:	Dr. Christian Horn
Submission Due Date:	28/09/2020
Project Title:	Classification of Sub-type of Lymphoma using Deep Learning
Word Count:	6109
Page Count:	20

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	28th September 2020

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Classification of Sub-type of Lymphoma using Deep Learning

Prasad Balasaheb Thorat
x18185711

Abstract

The accurate identification and categorization of cancer structure or sub-type is an important task because of considerable workload and expertise in pathological skills. The current digitization in diagnostic system of disease provides huge quantities of image data enabling a faster and accurate diagnosis through the development of automatic techniques for image classification. Previous studies have provided evidence for automatic cancer tissue analysis by using deep learning strategies that retrieve and organize discriminating insights from the images automatically. Therefore, in this study an innovative and empowered deep learning framework is proposed to classify three types of lymphoma as Follicular Lymphoma (FL), Chronic Lymphocytic Lymphoma (CLL) and Mantle Cell Lymphoma (MCL). In this research, a technique is followed with the help of the Inception_V3 and DenseNet121 along with a CNN. A publicly available dataset from National Institute of Ageing (NIA) is used in this study. This study performs the histogram normalization on all the images to enhance the performance of model. The data augmentation have been carried out on the dataset so that overfitting can be avoided. The achieved accuracy 92.51% from study shows that Inception_v3 managed to achieved better results in comparison with DenseNet121 and CNN in terms of categorical accuracy.

1 Introduction

Lymphoma is a blood diseases which develop in the immune cells, the human body's infection-fighting cells. Lymphoma disease is mainly categorized into two different types as Hodgkin lymphoma and non-Hodgkin lymphoma. These types of lymphoma are growing at significant rates which should be given special treatment. Non-Hodgkin lymphoma accounts for about 5% cases in men, and accounts for 4% in women, as per the study by American Cancer Institute. Lymphoma is very common disease, but still researchers have not identified what causes lymphoma. The effective treatments depend on the accurate type and level of the lymphoma, so an accurate diagnosis must be made by the doctor. Doctors have considered trying many ways to make lymphoma diagnosis more accurate. Microscopic lymphoma diagnosis remains stressful. Many pathologists and doctors collectively tried to increase the accurate detection rate of lymphoma diagnosis Sheng et al. (2020).

The current identification of sub type of lymphoma rely on histopathological analysis of tissues at various magnification level by an expert pathologist whose diagnosis is based on morphological characteristics observed in haematoxyl and eosine (HE) staining. Approximately 10 percent of cases remain complicated and take additional tests such as in-situ fluorescent hybridisation or polymerase chain reaction methods that are commonly unavailable in some labs. In addition to the methods of histopathological examination, more and more medical image processing devices and systems have been proposed for diagnosing the disease. These new methods are non-invasive compared to conventional methods of clinical examination and could get the dynamic information of a cancer Syrykh et al. (2020). Computer vision methods are new tools to bridge the gap between cancer diagnostic and analysis of pattern. Almost all classifications of cancer based on computer vision methods depend on the identification of individual cells, requiring segmentation or pre-selected regions of interest. The benefit of deep learning is that major hierarchical relationships inside the data can be identified algorithmically without manual interventions. (Orlov et al.; 2010).

To detect the breast cancer from histopathological slides, Robertson et al. (2018) employed several machine learning algorithms like random forest and multiple instance learning. The deep learning based diagnostic system can detect cancer regions better than tradition methods because it does not require any manual intervention and it can extract the features automatically. Recently, the diagnostic system based on deep learning have made automated methods for histopathology image processing possible, trough which an objective analysis may reduce the inter- and intra-observer variation in diagnosis of lymphoma. In addition, digital microscopy allows the identification of previously unidentified features by the human being and, paired with deep learning techniques, it can identify new morphological features.

This research mainly targets the three major types of non-Hodgkins lymphoma. These types are Mantle Cell lymphoma(MCL),Chronic lymphocytic leukemia(CLL) and follicular lymphoma(FL). As an image processing method, Histogram normalization is used in this study. The processed image is then augmented with basic rotation and flipping methods to avoid overfitting and biases issue.

This paper is structured in following manner: 2nd Section explains the previous work performed on this topic, Section 3 provides details of methodology and implementation, Section 4 gives details of evaluation metrics carried out and conclusions are discussed at the end in Section 5.

1.0.1 Research Question and Objective

How well the sub types of lymphoma can be classified using Deep learning and Transfer learning approach with limited computation power?

The motive of this research project is to develop an accurate classification model with limited computation power which gives a higher classification accuracy. Therefore, CNN and transfer learning with DenseNet121 and Inception_v3 is proposed in this research. This research project has following three setps.

- To perform preprocessing on images: The dataset used in this research has a less images. To achieve the better performance of the model, the histogram normalization and data augmentation is applied.

- Application of deep neural networks: Convolution Neural Network (CNN) which is developed from scratch and transfer learning with Inception-v3 and DenseNet121.

- To compare the performance of model on the basis of classification accuracy and execution time.

2 Related Work

2.1 Deep Learning in Medical Domain

The effective use of deep learning in the medical sector has the ability to develop new tool that surpasses the present techniques in terms of classification accuracy and time complexity. Image classification was one of the major areas in which neural networks made an important contribution to medial image analysis. The analysis done by Litjens et al. (2017) gives an overview of different image processing techniques applied in medical images. This study summarises the CNN model and pre-processing techniques used in image classification. Dimitriou et al. (2019) conducted a survey on deep learning for whole slide images. The most effective approaches to training models of deep learning on whole slide images is not to use the entire image as data and rather select only small patches from images. The majority of approaches in the previous decade concentrated on seeking ways to directly derive features from images for subsequent use of models. Thus, the extraction of features and the design of models were two separate, independent stages that were conducted sequentially. It has been shown that automating this method by the use of Convolutional Neural Networks (CNNs) results in more appropriate feature suited to the nature of the problem. This is one of the factors behind the success of deep learning and more specifically, learning based on neural networks in medical domain.

The deep learning method is used in Vo et al. (2019) for image classification. In this study, the feature extraction is done with the help of Deep Convolutional Neural Network. This results in the improvement of multiple class classification accuracy. Gradient boosting is used for the classification purpose. Nahid et al. (2018) applied CNN, a Long-Short-Term-Memory(LSTM) and a combination of CNN and LSTM for classification of histological breast images. After deriving the useful features from images, Softmax and Support Vector Machine(SVM) layers are used for decision-making. In the study, an accuracy of 91% is achieved on 200x dataset and precision of 96% is achieved on 40x dataset. Guan et al. (2019) used the Inception-v3 Deep Convolutional Neural Network (DCNN) to differentiate lymphadenopathy. In the analysis there were few misclassified cases which shows that most challenging cytology type for DCNN to differentiate is lymphoma. The Inception-v3 DCNN model demonstrated great roles in enhancing the diagnosis of lymphadenopathy using cytological images after training with a huge dataset.

2.2 Need for Lymphoma Diagnosis

The most important diagnostic method for lymphoma is analysis of histopathological images which analyses the morphological characteristics of tumour under microscope. Even after the analysis being performed by expert hematologists and pathologists, the diagnosis is difficult due to numerous factors such as human skill requirement, complexities, variation in slide staining etc. The application of image processing and machine learning techniques in different cancer detection methods appears as the key tool. The researchers used these techniques to identify CLL, FL and MCL sub types of lymphoma, although the existing systems are not so effective due to the complex characteristics of these sub types, and progress is still being made in improving lymphoma classification. In this section the previous work done in this area in past ten years will be discussed.

As per the mutation of cells and the required course of treatment, Lymphoma is categorised as Hodgkin's lymphoma and non-Hodgkin's lymphoma. The non-Hodgkin's lymphoma starts to grow when the B or T types of blood cells becomes abnormal. According to World Health Organization, there are more than 40 types of lymphoma. in this research three main types which are CLL, FL and MCL are considered.

Lippi et al. (2020) used the concept of multiple-instance learning to classify types of malignant lymphoma. In the analysis Support Vector Machine (SVM) and Random Forest (RF) are used as classifier. In multiple-instance learning, the example to be classified consists of collection of instances, and then the label is typically given to a bag or collection rather than a single instance. In the study, the bag corresponds to the patients and single instance corresponds to the volume of interests. The system implemented in has a pipeline of stages. In the first stage, the segmentation is performed on raw image for volume of interests, subsequently texture analysis is performed on extracted regions of interests. At the end, the machine learning classifier is trained to learn a function that is capable to divide lymphoma subtypes. The proposed system undergoes mutiple stages, which requires manual intervention to segment the image. This process is time-consuming because it requires experts to extract the volume of interests. Therefore it is very important to develop a system that can extract features directly from images without need of segmentation.

A range of previous studies have used the automated method for classifying lymphoma images on publicly accessible datasets. Each study used different techniques for the preprocessing stage of images, extraction of features, classification of images. Image preprocessing is very fundamental and an important step in classification. Basically, this is used to eliminate noise, improve contrast, image sharpening. In Orlov et al. (2010) to distinguish H and E color of staining, Color deconvolution approach is used. Ruifrok et al. (2001) suggested deconvolution method for seperation of overlapping spectra. In h and e staining H targets cell nuclei (blue), and cytoplasm (red) is stained with eosin. The another important step is segmentation anf feature extraction. Feature extraction methods are applied to capture the feature from images that will be useful in detection or classification. The segmentation and texture analysis is done with the help of Gabor filters. Gabor filters were used to detect edges, and to reflect texture. The functions used in Gabor filter are four-fold Gaussian transformation, and four-fold harmonic function transformation.

Dabeer et al. (2019) classified breast histopathology images into benign and malignant using deep learning. The Network was trained with 2480 benign and 5429 malignant images of the RGB color model. The CNN is made up of two types of layers as Convolutional layer and Pooling layer. The classification method is performed by taking the flattened weighted feature map derived from the final pooling layer, which is used as input to the fully connected network, which estimates the loss which subsequently modifies the weights of the hidden internal nodes.

2.3 Application of Transfer Learning in Histopathology

Standard deep learning techniques study the training dataset for future inferences, assuming that future data are in the same space or are distributed in the same way as the training. Because of the access restrictions to human-labeled training data, training data remaining in the same feature space or providing the same distribution as future data can not be assured to be sufficient to prevent the overfitting problem. Besides the data in the target domain, similar data can also be used in real-world applications in a different domain to extend the scope of our prior information about the target future data. Transfer learning is used as a solution to solve this challenges. Transfer learning extract the information from another or related domain and transfers them in target tasks Shao et al. (2014). In the medical domain it is very difficult to get sufficient amount of data for training. Therefore tarnsfer learning is growing in medical image diagnosis. Chang et al. (2017) proposed a transfer learning approach based on Google's Inception v3 model. The result from the research shows the capability of Transfer learning with AUC of 0.93. The trained model in the study performed classification with 0.83 accuracy for benign and 0.89 for malignant.

The similar approach for transfer learning is used by Xie et al. (2019) to perform multiclass classification. In the study Inception_V3 and Inception_ResNet_V2 are used as a pre-trained model. These two classifier architectures are trained on huge dataset of ImageNet. Their structure of learning and parameters used are frozen. The neurons in the fully-connected layer is modified based on the task and parameters in the fully-connected layers are retrained. The study indicates that Inception_ResNet_V2 is superior in performance to Inception_v3. The model used in the study indicates that Inception_ResNet_V2 was able to extract more useful features. The transfer learning solved the issue of insufficient training data. In Vogado et al. (2018) VGG and AlexNet were used to extract the feature. the proposed methodology aims to accurately classify the images with features extracted from images without need of segmentation. Motlagh et al. (2018) presents the deep learning models ResNet and Inception to differentiate microscopic cancerous images. The system proposed in the study also applies additional techniques like advanced preprocessing and data augmentation. The model ResNet_V1_152 was able to classify the malignant and benign breast cancer with accuracy of 98.7%. Wu et al. (2019) used pre-trained model Densenet121 and ResNet50 to derive features from images and then connected with classifiers like SVM. The result from experiment shows that transfer learning combined with SVM is capable to detect the cancer area with accuracy of 97.07%.

2.4 Image Processing Techniques

2.4.1 Preprocessing

The very important step in any image analysis is to eliminate insignificant cell structures, remove noise, and increase contrast. Basically the aim is to improve the image quality for segmentation. This step is widely used in studies related to lymphoma. IN Orlov et al. (2010) color deconvolution is used as the preprocessing approach on the dataset of 40 images stained with hematoxylin and eosin. Vu et al. (2019) discussed the effect of preprocessing techniques on convolutional neural network. In the study, the images were divided into four different preprocessing classes which are later processed with three different techniques. The classes include original, normal preprocessing, over preprocessing and other normal preprocessing. The histology images from these four classes includes cancerous and non-cancerous patches. in the normal preprocessing class, cell enhancement and noise reduction are applied. in over preprocessing thresholding and morphological operations are applied. The experimental results shows that normal preprocessing algorithm has better effect on classification.

Anghel et al. (2019) performed stain normalisation with the help of unsupervised-stain vector method. The comparative study shows differences before applying stain normalisation and after applying stain normalisation. The results show that the accuracy of CNN for classification is improved by 8 percent. in Wang et al. (2019), a traditional image thresholding approach called as OTSU algorithm is used to remove noise while maintaining the tissue regions. The blind color deconvolution approach is used in Hidalgo-Gavira et al. (2018). This method is based on bayesian technique. the analysis shows that this method is faster and easier to implement. Histogram normalisation is useful in improving image quality and appearance of cancerous cells. It is very essential investigate this approach which can enhance the performance of medical image classification models.

Inconsistent staining of the biopsy and nonstandard imaging may cause color variation in histopathology images. There are multiple research which examine normalization method to accurate image colors that may differ due to various stain producers, tissue storing requirements, and tissue sample coloring and digitization processes. It has been noted that in the preprocessing step no analysis of the lymphoma segmentation algorithms examined the color normalization. It is necessary to investigate this method that can enhance performance of subsequent image analysis algorithms.

2.4.2 Segmentation

The segmentation is being used to identify objects of interest which lead to an efficient analysis of particular structures. These identified regions can be used to detect or classify the cancer types. Segmentation is performed based on the related regions by means of homogeneity parameters such as color and texture. The techniques for segmentation of lymphomas could be classified into methods based on the histogram and edge detection. There are multiple region of interests in lymphoma images such as lymphocytes, follicular regions, centrocytes and neoplastic cells. Methods based on regions, thresholding, and K-means clustering algorithm have been proposed for segmentation of follicular regions. OTSU method is one of the most popular threshold method.

Tosta et al. (2017) used the K-mean clustering algorithm with the Euclidean distance to recognize follicular regions by classification of pixels.

2.4.3 Data Augmentation

Data Augmentation addresses the issues of overfitting by increasing the data size. This is done on the belief that additional information can be derived through augmentations from the training image. There are many data augmentation techniques available in the literature, but it is very essential to maintain the label of image while applying augmentation. Most of the analyzed images are extracted from computerized tomography (CT) and magnetic resonance imaging (MRI) scans, all of which are expensive to collect and labor intensive. Because of the rarity of diseases, the confidentiality of patients, the need for medical experts to mark, and the cost and manual effort needed to carry out medical imaging processes, it is especially difficult to create large medical image datasets Shorten and Khoshgoftaar (2019). Prior study on data augmentation for computational pathology identified important groups of techniques of augmentation : (1) morphological and (2) color transformation.

Convolutional Neural Networks have achieved great attention in disease classification. It is very important for a model to train on a huge data. When a neural network is trained on a small data, there is a chance of overfitting. A model is considered overfit when it perform well on training data but poor on any new dataset (Shorten and Khoshgoftaar; 2019). Different approaches have been recommended over the period of time to reduce overfitting. The most common ways to avoid the problem of overfitting are adding regularization term or dropout technique. There is also another technique called batch normalization that is helpful in this scenario . In (Namozov and Im Cho; 2018) , the training dataset was increased with the help of techniques like flipping the images , or rotating by 90°. Three new images were created for each original image using data augmentation.

Mikołajczyk and Grochowski (2018) discussed the importance of style transfer in data augmentation. Style transfer is connected to color and text transfer. This method resample the pixel from one image to other image. Two histological images were combined in this study to make a new one with a solid structure and different color. The following figure shows the effect of style transfer on histopathologial image.

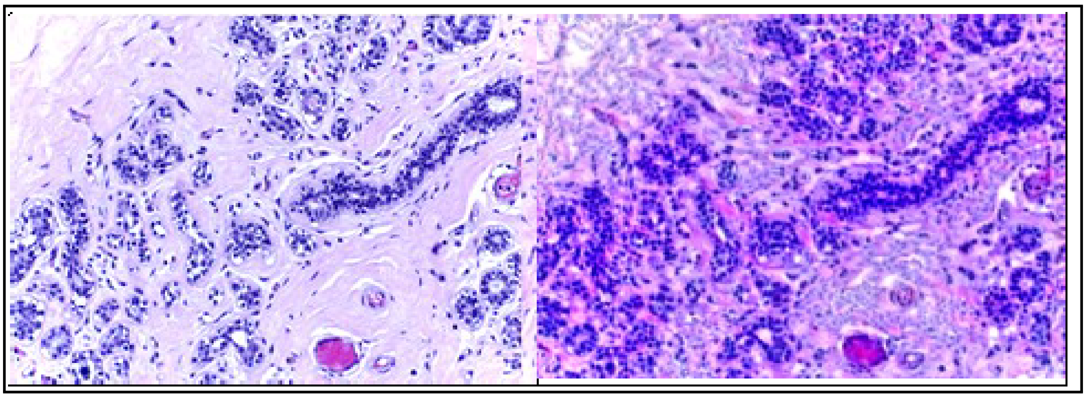

Figure 1: Left Image: Original Right Image: After Transformation Vu et al. (2019)

3 Methodology

The goal of the project is to establish a system that can accurately classify the sub types of lymphoma. To achieve this, there is a need to design a workflow. After analysing and gathering the requirement of study, CRISP-DM is selected as a methodology. The impacting factor about CRISP-DM is that it can be modified at any stage according to the requirement of research. The CRISP-DM involved multiple steps as shown in Figure 2. Each stage involved in this process is explained further Wirth and Hipp (2000).

The research shall use image dataset of tissues which is structured into different folders based on classes. Such data will be stored and pre-processed by resizing them to an appropriate dimensions which is then divided into testing and training dataset. The deep learning models are constructed on train dataset, which are also expanded to avoid overfitting, thus facilitating accurate prediction on a new dataset. The model that is developed in previous step is the evaluated on test dataset using different evaluation metrics.

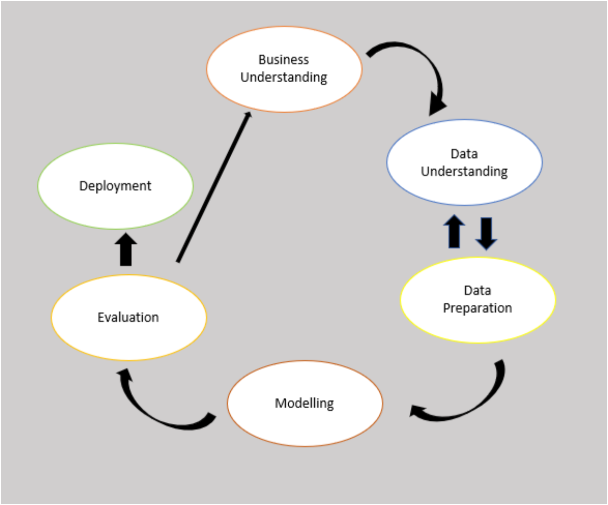

Figure 2: Methodology

3.1 Business Understanding

The current method to detect the lymphoma type is based on the analysis of biopsy images under microscope. These technique to detect and diagnose the exact type of lymphoma have some shortcomings. The accurate and early diagnosis of lymphoma is necessary for deciding the correct course of treatment for patients. But the current method for diagnosis requires manual intervention and multiple reviews before finalizing the decision. With the increase in number of cases of cancer, it is becoming more challenging to address this issue. Hence it has become important to build a model that can speed up the process of diagnosis. As discussed in the literature, a CNN model is used in this study. The CNN model will directly extract the features from images without any manual intervention resulting in speedy diagnosis.

3.2 Data Understanding

It is very important for research study to select an appropriate dataset. This study use the dataset that is publicly available for lymphoma sub type classification. This dataset is provided by National Institute of Ageing (NIA). This includes three types of lymphoma: FL, CLL and MCL Janowczyk and Madabhushi (2016). The images are stained with Hematoxylin and Eosin, and the data was collected from various sites to introduce a high variation to the dataset. There are total 374 images in the dataset of size 1388 X 1040. Out of 374 images, 112 belongs to CLL, 140 to FL and remaining 124 to MCL. The sample from each class is shown in following figure.

3.3 Data Preparation

After choosing the set of data, the images should processed and modified so they are fully compatible to achieve the greater efficiency of the designed model. For feeding input images to the model, images are normalized for better readability. The reshaping of images was done according to the requirement of different models. Data augmentation is performed on the training dataset to avoid overfitting.

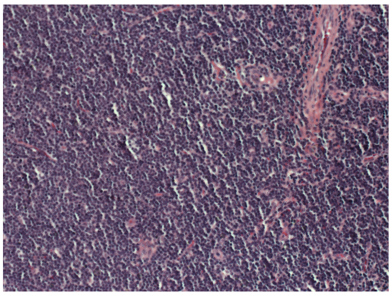

Figure 3: Chronic Lymphocyte Lymphoma

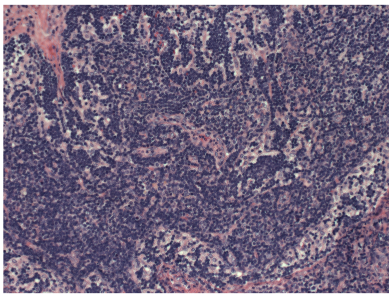

Figure 4: Follicular Lymphoma

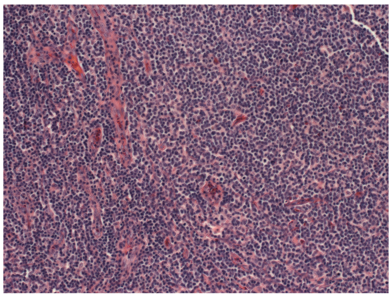

Figure 5: Mantle Cell Lymphoma

3.4 Modelling

3.4.1 Convolutional Neural Network (CNN)

A deep Learning based Convolutional Neural Network is primarily used for image classification. A CNN model is an improved version of traditional neural network which permits a network to derive global as well as local features from images. This helps in developing model that has better decision making system. A few intermediate layers are added to manage the functionality of a CNN network, with a convolutional layer. As discussed in (Nahid et al.; 2018), there are five main layers present in the CNN. As mentioned earlier, convolutional layer is used to extract the feature. To compute an activation map consisting of neurons, each filter is transformed with the volume of inputs. The another layer called as Batch normalization is responsible for receiving input from earlier layer and providing it to next layer. Each neuron gives linear output. When the output of one neuron is fed to another neuron it again produces linear output. to overcome this situation there are several nonlinear activation functions available in CNN. These functions are Sigmoid, TanH, ReLU, Leaky ReLU. A convolutional Model produces a large number of features. As the number of features extends, the computational complexity also increases significantly and the model becomes more sensitive. To overcome this situation, a sampling process is introduced. A sampling or pooling process reduces the overall dimensionality and complexity. For determining class count each fully connected layer connects to the previous layer. The figure gives the flow diagram for this research study

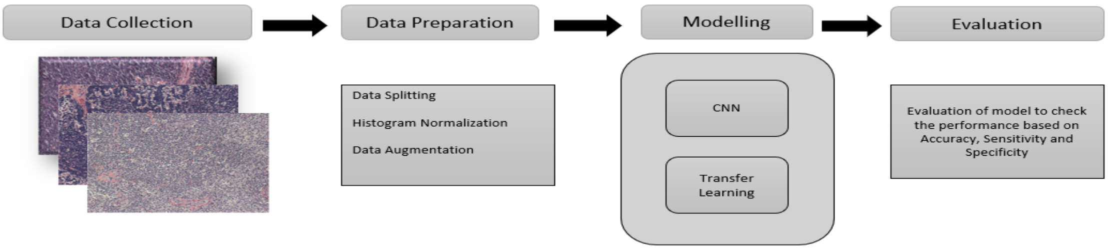

Figure 6: Modelling

3.4.2 Transfer Learning with Inception-v3 and DenseNet121

In this research transfer learning is used to solve the measure issue of lack of data. The transfer learning is used with the help of Inception-v3 and DenseNet121. The Architecture of Inception-V3 is as shows in Table 1 as explained in (Guan et al.; 2019). The Inception network is developed in a way that convolution modules can produce discriminatory characteristics and decrease the parameters required. Each Inception module consists of multiple convolution layer and parallel pooling layers.

Layers	**stride**	**Input**
convolutional layer	3x3/2	224x224x3
convolutional layer	3x3/1	111x111x32
Padded	3x3/1	109x109x32
Pool	3x3/2	109x109x64
convolutional layer	3x3/1	54x54x64
convolutional layer	3x3/2	52x52x80
convolutional layer	3x3/1	25x25x192
Inception A x3	-	25x25x288
Inception B x5	-	12x12x768
Inception C x2	-	5x5x1280
Fc	51,200x1,024	5x5x2,048
Fc	1,024x1,024	1,024
Fc	1,024x4	1024
Softmax	Classifier	4

Table 1: Architecture Of Inception_v3

The all layers before Fully Connected(Fc) layer form a foundation network for Inception. The difference between two operations is known as stride.

Similarly an architecture of DenseNet121 is discussed in Ho and Gwak (2019). DenseNet121 can enhance the intake information from data and process them into deeper dimension. Because of having less layers, the network could be thinner and compact. DenseNet could learn more richer and globalized data features, as the DenseNet classifier utilizes features from across all levels of complexity that offer clearer decision boundaries when the training set is limited.

3.5 Evaluation of Applied Model

The efficiency of all the model is measured with the help of evaluation metric like classification accuracy. The details are given in Section 6.

4 Design Specification

4.1 Convolutional Neural Network (CNN)

These are the primarily used algorithm in medical disease classification and visual learning task, because of its exceptional characteristic of preserving image relations, while carrying out dimensionality reduction. It takes necessary features from an image and minimize the parameters that algorithm has to evaluate, enhancing the computational efficiency. One of the major issue that impact the performance of CNN is feature selection. The CNN gets histopathological slides as input and transforms with the help from Convolutional,Polling and RELU layer as shown in Figure 7. The basic assumption behind CNN efficiency lies in multidimensional representation, which means that the features of the higher level offer stronger data information (Ker et al.; 2017).

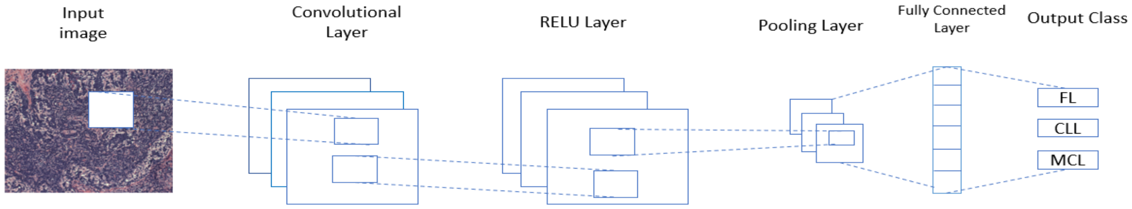

Figure 7: CNN Workflow

4.2 Convolutional Layer

The fundamental layer in model is Convolutional Layer. A convolution is an operation consists of 2 functions. In classification, one function contains pixel values , and second function contains kernel size. Calculating the product among the functions gives result. Then the filter is moved towards the succeeding position in the image. It is called as stride length. This process is done unless the whole histopathological slide is completed.

4.3 Fully Connected Layer

The Final layer is fully connected layer, which means that each neuron of earlier layer is attached to neuron in fully connected layer. There can be one or more fully connected layers in the CNN depends on image classification task. This layer gets the output from the previous layer which are Convolutional, Pooling or RELU as its input, and calculates a probability value for classification into different classes. Essentially, this layer take a look at the combination of the most activated features which would suggest that the image corresponds to a specific class. For example , cancer cells with high DNA to cells ratio compared with normal cells on pathology glass slides. If characteristics of cancerous cell were firmly identified from the previous layer, the CNN will detect the presence of abnormal cells.

4.4 Rectified Linear Unit (RELU) Layer

The RELU layer is a non linear function in which negative values are fixed to 0. This optimizes and improves calculations and learning, promotes to prevent an issue of disappearing gradients.

4.5 Pooling Layer

The main function of this layer is to decrease the number of parameters and the dimension of image. Pooling only reduces width and height and not the depth. The position of pooling layer is between Convolutional and RELU layer. The widely used layer is max-pooling layer. The other pooling layers are L2-normalization and average pooling.

4.6 Optimization

The performance of classification model can be enhanced by making changes in parameters. The one such kind of parameter is optimizer. There are many optimizers available such as Stochastic Gradient Descent (SGD), RMSprop, Adam. The model is evaluated by comparing the actual outcome and the output obtained.

4.7 Learning Rate

It is very essential factor that has a huge influence on performance of model. The minimal learning rate is necessary and number less than that leads to unsatisfactory performance of model. The gradual increase of learning rate with each epoch results in better performance (Ker et al.; 2017).

4.8 Early Stopping

Early stopping is a parameter that helps to avoid overfitting when number of epochs are high. Basically it allows to give a high number of epoch to train the model and stop training once model starts to overfit.

5 Implementation

This section summarises about the system implication and lists out all the task carried out to classify the sub types of lymphoma. The first subsection will discuss about the tools used and required environment for research. In the second subsection, the data transformation will be explained followed by details about the applied model.

5.1 Development Environment

The two most powerful programming languages used in deep learning are Python and R. This project is entirely implemented in Python because of available libraries. Python has multiple libraries like tensorflow, keras which are dedicated for neural networks. The google colaboratory is used for experiment. The data stored in google drive is accessed with the help of Python API. Other necessary libraries like NumPy , Matplotlib and Python Imaging Library(PIL) are also installed for data transformation and data visualization.

5.1.1 Data Transformation and Data Preprocessing

As discussed in methodology section, the dataset publicly available by National Institute of Ageing is used for this research. Different operation like image resizing, histogram normalization and data augmentation are performed on the entire dataset. The original images were resized to 224 × 224 × 3 for Inception-v3 and CNN, and to 64 × 64 × 3 for transfer learning With DenseNet. Histogram normalization is performed on the entire dataset. A function is written which takes all the images as input and write the transformed images into different folder. This technique is used to enhance the brightness of input images without disturbing the cell structure. The following figure shows the effect of histogram normalization on sample image.

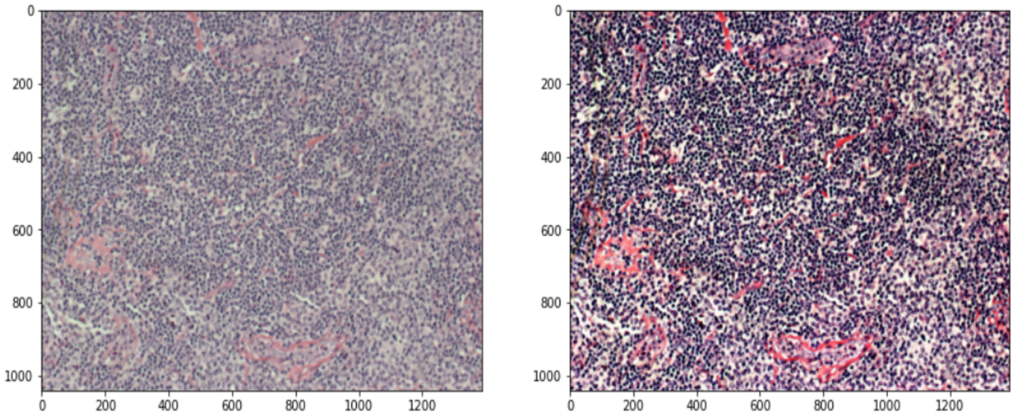

Figure 8: Left: Original Image Right: Transformed Image

Data augmentation was carried out on entire dataset to enhance the performance of model. There are several data augmentation methods provided by tensorflow like rotating images with different angles, cropping image randomly and adjusting brightness, etc. This study uses rotation by 90°and flipping to increase training set without compromising region of interest.

5.1.2 CNN Built From Scratch

Layers	**Information**
1st Layer	Convolutional layer, ReLu, Kernels = 16, Drop out(0.3)
2nd Layer	Convolutional layer, ReLu, Kernels = 32, Drop out (0.3), Max Pooling(3x3)
3rd Layer	Convolutional layer, ReLu, Kernels = 64, Drop out (0.3), Max Pooling(3x3)
4th Layer	Convolutional layer, ReLu, Kernels = 128, Drop out (0.3), Max Pooling(3x3), Drop out(0.3)
Dense	512 Nodes
Output	Softmax(3 Classes)

Table 2: Layers of Proposed CNN

The Table 2 summarise the layers included in CNN. To execute the model on training and validation data, model.fit generator() is used. The different number of epochs were given depending on the requirement of experiment. Adam Optimizer is used with learning rate of 0.0001.

5.1.3 Transfer learning using Inception-v3 and DenseNet121

There are multiple architectures available in keras which are trained on ImageNet data. These models can be utilized for transfer learning. The ImageNet dataset has large number of training and validation images. There are total 1000 categories in the ImageNet dataset. Inception_v3 and DenseNet121 are used in this study. The weights of these architecture are first installed from keras. These models are then changed based on the requirement of image classification. The 4 top layers are trainable and the remaining layers were freezed. The multiple combination of parameters like number of epochs, batch size were tried for analysis. The learning rate was also different for various experiments. The 0.0001 learning rate gave better accuracy with high number of batch size. The original input size of Inception_v3 is 299. But in this study the images were resized to 224 × 224. The model gives better results for image size of 224 × 224. The Inception_v3 architecture has thousand classes, but this study targets only three classes which are FL, CLL and MCL. Hence, in this study, the number of output class is modified to three. The weights of DenseNet121 are initialized with a pre-trained model using Adam standard parameters. The last layer of DenseNet121 has softmax function.

6 Evaluation

6.1 Case Study 1

The CNN model was developed from scratch. This model consist of four convolutional layers. The number of epochs varied for different experiments. The below Table 3 demonstrates the achieved classification accuracy with different approaches.

Accuracy	**CNN(Original Dataset)**	**CNN(Histogram Normalization)**	**CNN(After Augmentation)**
Training Accuracy	66.52%	68.57%	74.53%
Testing Accuracy	57.89%	67.10 %	71.65%

Table 3: CNN Evaluation

When the CNN is applied without any pre-processing on original dataset then the situation called as overfitting occurs. To overcome this problem, Data Augmentation was performed and training image size was increased from 242 images to 1214 images. By augmenting the images the accuracy increased to 71.65% from 57.89%. The model gives an accuracy of 71.65% on augmented dataset which is highest in all three experiment. The accuracy was increased approximately by 10% after histogram normalization performed on original dataset. The Figure 9 shows the behaviour of training accuracy and validation accuracy.

The number of epochs were further increases to 100 to check whether it increases the performance of model or not. After increasing the number of epochs, the classification accuracy increases by 5%.

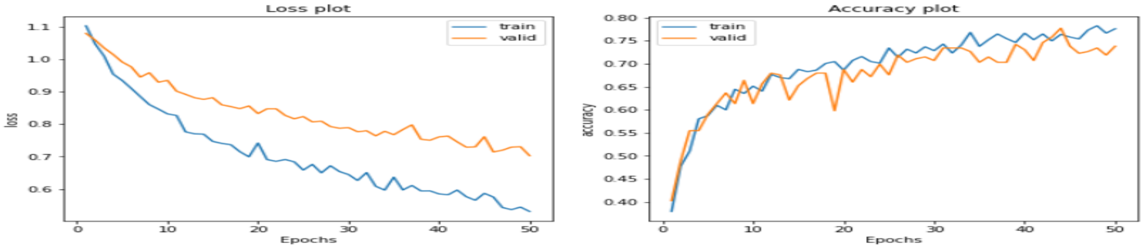

Figure 9: Epochs (50) – CNN

6.2 Case Study 2

Transfer learning with the help Inception_v3 is applied on original dataset without any pre-processing. In this case the training set contains 242 images, validation set of 56 images and rest 76 images were in testing set. The model was trained for different set of epochs. The model managed to achieve an accuracy of 57.89% for 10 epochs.

Evaluation Metric	**Inception_v3(Original Dataset)**	**Inception_v3(After Augmentation)**
Training Accuracy	72.73%	92.75%
Test Accuracy	57.89%	92.51%

Table 4: Inception_v3 Evaluation

The model Inception is kept same but the size of training dataset was increased from 242 images to 1214 images with the help of data augmentation. First the histogram normalization is performed on entire dataset to increase the brightness of images. After generating normalized images, basic augmentation methods like flipping(left to right, up down) and rotation performed on the entire data. The model has achieved an accuracy of 92.51% on the original test after building the classifier for 10 epochs. The behaviour of training and validation accuracy is as shown in Figure 15.

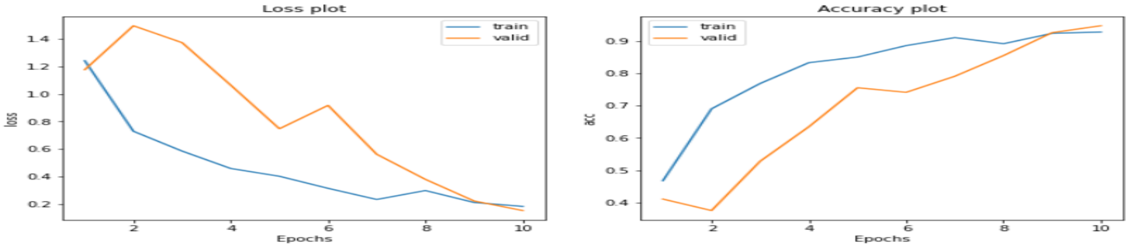

Figure 10: Epochs (10) – Inception_v3

6.3 Case Study 3

DenseNet121 model with Adam optimizer at learning rate of 0.0001 does not perform well on original dataset. The DenseNet121 gives an accuracy of 40% on original dataset having 242 training images. The Figure 11 Shows the behaviour of training and validation accuracy with respect to number of epochs.

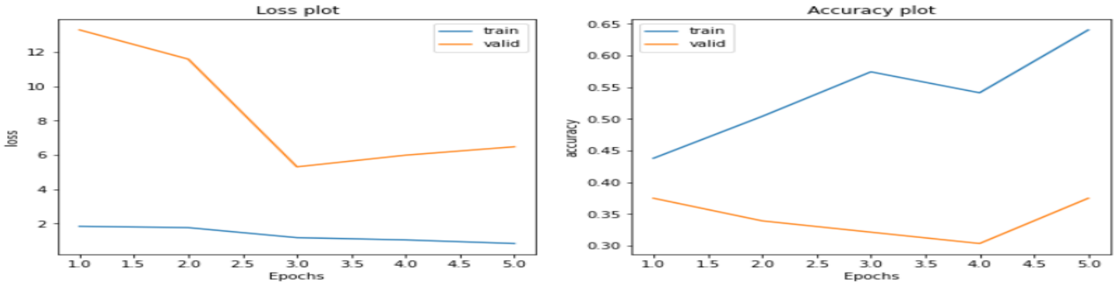

Figure 11: DenseNet121 on Original Dataset

To increase the performance of model, the input images were preprocessed with the help of histogram normalization and data augmentation was performed on entire dataset. The dataset was again splitted into 80:20 ratio. The training dataset contains 1214 images. Different image size were given as an input to DenseNet121. The highest accuracy obtained was 88.15% for input size of 64x64.

7 Discussion

The objective of this study was to accurately classify sub types of lymphoma with deep learning. A significant drawback when deep learning is implemented is the shortage of training data that allows the model to learn features. Since the basic requirements for deep learning were not fulfilled, it was difficult to classify sub types of lymphoma. In this study 374 images were used to obtained the accuracy of 92% with Inception_v3. Image Normalization technique in pre-processing has shown a great effect on neural networks. While performing the experiment, different data augmentation methods have been tried on sample images to check the significance. It has been observed that flipping, rotation and cropping are the label-preserving transformations. The following graph in Figure 12 shows the comparative study of model in terms of accuracy.

Besides evaluating the models based on accuracy, a training time is also compared and understood how each one of these models perform against one another. As the CNN model was built from scratch, it took more time to train as compared with Incepiton_v3 and DenseNet121. The results shows that transfer learning models were better in terms of execution time and accuracy. The limitation of the study has been to obtain more set of histopathological images for lymphoma that could help train the model more. The field of research may be extended to many other types of disease.

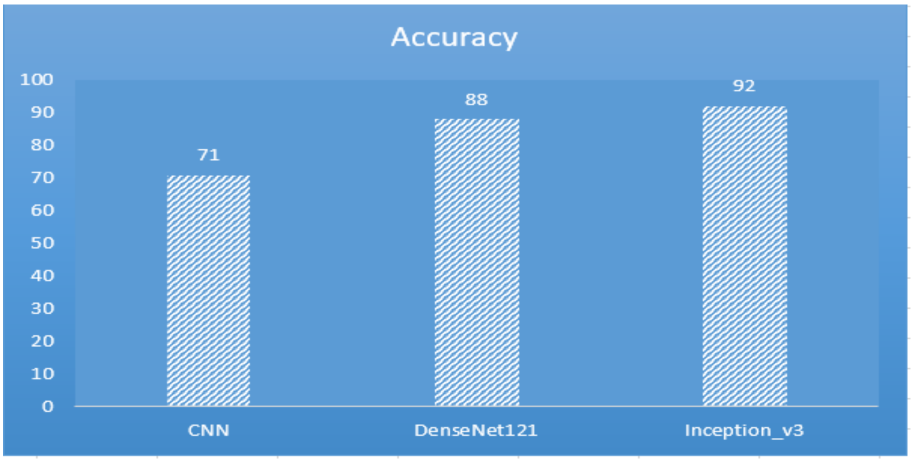

Figure 12: Comparative study of model in terms of accuracy

8 Conclusion and Future Work

To classify the sub types of lymphoma, we used deep learning convolutional neural network and transfer learning with Inception_v3 and DenseNet121. The dataset has been tested with each model. The preprocessing was performed on images to improve the performance of model. It is very important to apply label-preserving transformations while augmenting the data. Because of the system's limited computing capacity it was not possible to perform more image operations. At last, it has been observed that Inception_v3 with transfer learning obtained an accuracy of 92.51%. It is concluded that this image data is best suited for Imagenet pre-trained weight and the softmax output activation function. Performing all of the experiments it has been observed that performance can be significantly increase by increasing training data and epoch number. In future the performance of DenseNet121 can be improved with the help of binary search connection. The hyperparameter testing and tuning was not carried out on large-scale due to time constraints and resources. In future work, attention should be given to a large numbers of images and parameters which might benefit the classifiers. Due to lack of data availability this research considered only three types of lymphoma. In future the targeted class can be increased.

9 Acknowledgment

I am grateful to have Dr. Christian Horn as my supervisor for research project. He helped me in every stage of this research project which made it possible. I am also thankful to my parents for their continuous support.

References

Anghel, A., Stanisavljevic, M., Andani, S., Papandreou, N., Rüschoff, J. H., Wild, P., Gabrani, M. and Pozidis, H. (2019). A high-performance system for robust stain normalization of whole-slide images in histopathology, *Frontiers in Medicine* **6**.

Chang, J., Yu, J., Han, T., Chang, H.-j. and Park, E. (2017). A method for classifying medical images using transfer learning: A pilot study on histopathology of breast cancer, *2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom)*, IEEE, pp. 1–4.

Dabeer, S., Khan, M. M. and Islam, S. (2019). Cancer diagnosis in histopathological image: Cnn based approach, *Informatics in Medicine Unlocked* **16**: 100231.

Dimitriou, N., Arandjelović, O. and Caie, P. D. (2019). Deep learning for whole slide image analysis: An overview, *Frontiers in Medicine* **6**.

Guan, Q., Wan, X., Lu, H., Ping, B., Li, D., Wang, L., Zhu, Y., Wang, Y. and Xiang, J. (2019). Deep convolutional neural network inception-v3 model for differential diagnosing of lymph node in cytological images: a pilot study, *Annals of translational medicine* **7**(14).

Hidalgo-Gavira, N., Mateos, J., Vega, M., Molina, R. and Katsaggelos, A. K. (2018). Blind color deconvolution of histopathological images using a variational bayesian approach, *2018 25th IEEE International Conference on Image Processing (ICIP)*, IEEE, pp. 983–987.

Ho, T. K. K. and Gwak, J. (2019). Multiple feature integration for classification of thoracic disease in chest radiography, *Applied Sciences* **9**(19): 4130.

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

Ker, J., Wang, L., Rao, J. and Lim, T. (2017). Deep learning applications in medical image analysis, *Ieee Access* **6**: 9375–9389.

Lippi, M., Gianotti, S., Fama, A., Casali, M., Barbolini, E., Ferrari, A., Fioroni, F., Iori, M., Luminari, S., Menga, M. et al. (2020). Texture analysis and multiple-instance learning for the classification of malignant lymphomas, *Computer methods and programs in biomedicine* **185**: 105153.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Van Der Laak, J. A., Van Ginneken, B. and Sánchez, C. I. (2017). A survey on deep learning in medical image analysis, *Medical image analysis* **42**: 60–88.

Mikołajczyk, A. and Grochowski, M. (2018). Data augmentation for improving deep learning in image classification problem, *2018 international interdisciplinary PhD workshop (IIPhDW)*, IEEE, pp. 117–122.

Motlagh, M. H., Jannesari, M., Aboulkheyr, H., Khosravi, P., Elemento, O., Totonchi, M. and Hajirasouliha, I. (2018). Breast cancer histopathological image classification: A deep learning approach, *BioRxiv* p. 242818.

Nahid, A.-A., Mehrabi, M. A. and Kong, Y. (2018). Histopathological breast cancer image classification by deep neural network techniques guided by local clustering, *BioMed research international* **2018**.

Namozov, A. and Im Cho, Y. (2018). An improvement for medical image analysis using data enhancement techniques in deep learning, *2018 International Conference on Information and Communication Technology Robotics (ICT-ROBOT)*, IEEE, pp. 1–3.

Orlov, N. V., Chen, W. W., Eckley, D. M., Macura, T. J., Shamir, L., Jaffe, E. S. and Goldberg, I. G. (2010). Automatic classification of lymphoma images with transform-based global features, *IEEE Transactions on Information Technology in Biomedicine* **14**(4): 1003–1013.

Robertson, S., Azizpour, H., Smith, K. and Hartman, J. (2018). Digital image analysis in breast pathology—from image processing techniques to artificial intelligence, *Translational Research* **194**: 19–35.

Ruifrok, A. C., Johnston, D. A. et al. (2001). Quantification of histochemical staining by color deconvolution, *Analytical and quantitative cytology and histology* **23**(4): 291–299.

Shao, L., Zhu, F. and Li, X. (2014). Transfer learning for visual categorization: A survey, *IEEE transactions on neural networks and learning systems* **26**(5): 1019–1034.

Sheng, B., Zhou, M., Hu, M., Li, Q., Sun, L. and Wen, Y. (2020). A blood cell dataset for lymphoma classification using faster r-cnn, *Biotechnology & Biotechnological Equipment* **34**(1): 413–420.

Shorten, C. and Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning, *Journal of Big Data* **6**(1): 60.

Syrykh, C., Abreu, A., Amara, N., Siegfried, A., Maisongrosse, V., Frenois, F. X., Martin, L., Rossi, C., Laurent, C. and Brousset, P. (2020). Accurate diagnosis of lymphoma on whole-slide histopathology images using deep learning, *npj Digital Medicine* **3**(1): 1–8.

Tosta, T. A. A., Neves, L. A. and do Nascimento, M. Z. (2017). Segmentation methods of h&e-stained histological images of lymphoma: a review, *Informatics in Medicine Unlocked* **9**: 35–43.

Vo, D. M., Nguyen, N.-Q. and Lee, S.-W. (2019). Classification of breast cancer histology images using incremental boosting convolution networks, *Information Sciences* **482**: 123–138.

Vogado, L. H., Veras, R. M., Araujo, F. H., Silva, R. R. and Aires, K. R. (2018). Leukemia diagnosis in blood slides using transfer learning in cnns and svm for classification, *Engineering Applications of Artificial Intelligence* **72**: 415–422.

Vu, Q. D., Graham, S., Kurc, T., To, M. N. N., Shaban, M., Qaiser, T., Koohbanani, N. A., Khurram, S. A., Kalpathy-Cramer, J., Zhao, T. et al. (2019). Methods for segmentation and classification of digital microscopy tissue images, *Frontiers in bioengineering and biotechnology* **7**.

Wang, X., Chen, H., Gan, C., Lin, H., Dou, Q., Tsougenis, E., Huang, Q., Cai, M. and Heng, P.-A. (2019). Weakly supervised deep learning for whole slide lung cancer image analysis, *IEEE Transactions on Cybernetics* .

Wirth, R. and Hipp, J. (2000). Crisp-dm: Towards a standard process model for data mining, *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, Springer-Verlag London, UK, pp. 29–39.

Wu, C., Ruan, J., Ye, G., Zhou, J., He, S., Wang, J., Zhu, Z., Yue, J. and Zhang, Y. (2019). Identifying tumor in whole-slide images of breast cancer using transfer learning and adaptive sampling, *2019 Eleventh International Conference on Advanced Computational Intelligence (ICACI)*, IEEE, pp. 167–172.

Xie, J., Liu, R., Luttrell IV, J. and Zhang, C. (2019). Deep learning based analysis of histopathological images of breast cancer, *Frontiers in genetics* **10**: 80.