

# Tata Memorial Hodgkin Lymphoma

Aditya Iyer, under the guidance of Dr. Venkatapathy  
Subramanian and Dr. Ganesh Ramakrishnan

October 2022

## 1 Summary

I was part of a team from IIT Bombay that worked on the use of AI/ML to diagnose Hodgkin Lymphoma (HL). The project is being conducted under the aegis of the Koita Centre for Digital Health with the partnership from Tata Memorial Hospital.

The goal of the project is to use AI/ML to evolve a strategy to help select the treatment protocol based on the HL patient's disease characteristics at baseline using PET-CT images and pathology images. I worked with the team on using pathology images for the diagnosis. Coming up with an AI/ML model for diagnosing the disease based on pathology images is challenging because of the complexity involved in processing raw Whole Slide Images (WSI) for AI/ML pipeline; The WSI can be as large as 1GB. To understand the domain better, I, along with the team, worked on datasets that are from similar problem spaces. We worked on the Camelyon17 and TIGER global challenges.

The goal of the Camelyon17 grand challenge is to evaluate new and existing algorithms for automated detection and classification of breast cancer metastases in WSI of histological lymph node sections. I worked on the dataset - Camelyon17-WILDS, in which the data was in the form of patches extracted from the whole slide images in the grand challenge. My role included statistical analysis and manipulation of large datasets using python as well as an understanding of the basics of machine learning (autograd and backpropagation) to train patches of tissues extracted from whole slide images.

The TIGER project is based on the assessment of tumour-infiltrating lymphocytes in breast cancer whole slide images. I was involved in the data processing part of the challenge which encompassed understanding the various data sets which needed to be analyzed as well as extracting patches of relevant areas from a large whole slide image.

Working with Camelyon17 and Tiger challenges enabled us to improve the understanding of the manipulation of large data sets and a better understanding of various possibilities and approaches to breast cancer whole slide images. Additionally, I also spent time understanding some of the basics of breast cancer, which was critical for the project, which involved handling and manipulation of

pathology images.

The project involved a field trip to the Tata Memorial Hospital from where the data (WSI containing Hodgkin lymphoma) were collected. My role in the team was working with large data as well as extracting patches from the images which were learned through the Camelyon17 and TIGER challenges.

## 2 Introduction

### 2.1 Koita Centre for Digital Health

The Koita Foundation has partnered with IIT Bombay in setting up Koita Centre for Digital Health – KCDH is a first of its kind in India focusing on driving academic programmes, research, and industry collaborations in Digital Health. KCDH addresses the need for engineering, clinical professionals, and entrepreneurs in Digital health and creates research and employment opportunities in this domain.

### 2.2 The Project

One of the projects being undertaken under the aegis of KCDH is – **A novel AI-based prognostic approach using PET – CT images and pathology images for advanced stage Hodgkin Lymphoma (HL)**. This is being led by Prof. Ganesh Ramakrishnan, Professor in charge, KCDH and Institute Chair Professor, Department of Computer Science Engineering, Indian Institute of Technology, Bombay. For advanced HL, there are two widely used chemotherapy regimens – ABVD and BEACOPP regimens. ABVD is generally a more common regimen and is well tolerated in most patients, but has lower cure rates than BEACOPP. The BEACOPP regimen is found to be more effective for initial disease control and found to be better with respect to prolonged time to relapse. However, the regimen comes with a higher risk of significant toxicities.

#### 2.2.1 Tata Memorial Hospital

Tata Memorial Centre is amongst the oldest and largest cancer centres in the world, with over 75 years of exceptional patient care, high-quality training, and innovative cancer research. Researchers at the Tata Memorial Centre are world leaders in basic, translational, epidemiologic, and clinical cancer research. Research at TMC includes studies to understand cancer biology, large community-based screening trials for common cancers and cohort studies, neoadjuvant and adjuvant treatment, perioperative interventions, surgical trials, repurposing drugs, and qualitative research to understand the patient journey. The research focuses on impacting survival and quality of life i.e., living longer, or living better – endpoints that really matter to patients. Training and education form a major focus area for TMH, which remains the most sought-after institution for oncology training nationally.

### 2.2.2 Problem Statement

For advanced Hodgkin’s lymphoma, two chemotherapy regimens are often used for treatment, (i) ABVD and (ii) BEACOPP regimens. ABVD regimen has comparatively low cure rates as compared to the BEACOPP regimen, but it is tolerated well in most patients. BEACOPP, though associated with better outcomes, comes with the risk of significant toxicities.

### 2.2.3 Project Objective

With the help of AI / ML, to evolve a strategy to help select the protocol based on the patient’s disease characteristics at baseline using PET-CT images and pathology images. This can help avoid the comparatively toxic BEACOPP protocol where ABVD can give equally good results. The study includes 1) a retrospective aspect, where we use the previously available datasets to develop AI models to predict the outcome of ABVD treatment, 2) the prospective aspect, where we will use the AI models to predict the response to the treatment of ABVD BEACOPP chemotherapy protocols. This is an ongoing project.

## 3 Team

### 1. Prof. Ganesh Ramakrishnan

Professor-in-Charge, Koita Centre for Digital Health, Institute Chair Professor, Department of Computer Science Engineering Ganesh research interestes includes Robust labelled data generation, Data subset selection and summarization, Effective incorporation of domain knowledge for realizing AIs use in resource-constrained operational settings. Ganesh has focused on the adaptation of these contributions to real-world end applications such as optical character recognition and its post-editing, sequence-to-sequence tasks such as automated question generation and machine translation, human activity sensing in audio-visual data, and video summarization, etc. These adaptations have resulted in technology transfers, products, and start-ups.

### 2. Dr. Hasmukh Jain

Dr Hasmukh Jain (Tata Memorial Hospital) in Dr Earnest Borges Marg, Mumbai is one of the most renowned Oncologists at TMH. He is also an Associate Professor at Advanced Centre for Treatment, Research and Education in Cancer. He has co-authored various medical journals on the treatment of cancer and its allied field.

### 3. Dr. Venkatapathy

Venkatapathy is working as Senior Project Research Scientist, at IITB. He works in the field of Artificial Intelligence and has worked in various projects related to AI.

#### 4. Gowtham Khajjayam

Gowtham is an M.Tech 2nd year student at IIT Bombay. Machine learning is his current working area. He is currently working on generative models under Prof. Ganesh Ramakrishnan along with Dr Venkatapathy.

## 4 Background

Hodgkin's lymphoma (HL) is a type of cancer derived from B cell lymphomas characterized by distinctive immunophenotypes and relatively few malignant cells in a non-neoplastic inflammatory background. This disease is significant as it accounts for 15% of all cancers seen in young adults. The treatment of HL depends on the stage it has reached and comprises chemotherapy with or without radiotherapy. This project focuses on identifying prognostic factors in advanced stage HL. Currently, two chemotherapy regimens are used for the treatment of advanced Hodgkin's Lymphoma - ABVD and BEACOPP regimens.

- ABVD regimen - This regimen has cure rates of 70-75 %, which is less than those of the BEACOPP regimen. The regimen is tolerated well in most patients.
- BEACOPP regimen - This regimen has higher cure rates, with survival in the range of 85-90%. However, there is a higher risk of significant toxicities in the patients.

The challenge of the project is being able to identify patients with high specificity and positive predictive value for being treated with BEACOPP based on the history of the patient, PET-CT scan, and pathological images using AI models. The intended outcome of the project is to determine the small proportion of patients (using the AI model) to whom the BEACOPP regimen can be selectively administered while treating the other patients with the simpler ABVD regimen. Thus, the final goal of the project is to use AI and deep neural networks in assisting doctors in their decision-making process of choosing the correct treatment regime for patients having advanced stage HL.

### 4.1 Aim of the Project

- Application of AI models for prognosis in advanced Hodgkin's lymphoma
- Study of individual prognostic values of AI models based on PET-CT and pathology images as well as comparison with available gold standard IPS scores
- Assessment of AI models in predicting interim PET scan positivity

## 4.2 Modules

Three types of data are being used in developing the AI models to assist doctors for staging, treatment surveillance of the advanced-stage HL. They are

1. **Patient history**

These data include basic details collected during the preliminary checkup.

2. **Biopsy of the lymph node**

The biopsy samples of the lymph node are in the form of images consisting of slide micrographs on which morphological examinations are performed. The images capture the microenvironment of the tumour. **This is the part of the project that I worked on. More specifically, I studied the process of extracting relevant patches from WSIs for downstream tasks.**

3. **PET CT scans**

Information provided by the PET-CT scan is important for response evaluation after completion of treatment based on lymphoma guidelines. With a specificity of 90-95%, the scan can be used for staging, interim, end of treatment, and follow-up surveillance. Interim PET is a good predictor of response in Hodgkin Lymphoma as indicated by interim assessment after initial cycles of chemotherapy with the aim of identifying which patients are cured and which need escalation of treatment.

## 4.3 Prognosis

The prognosis itself happens in two stages. The first stage consists of two steps:

1. **Consideration of SME slides from the biopsy of the lymph nodes**

Hodgkin's lymphoma is unique as the bulk of the tumour tissue is composed of the supporting microenvironment, while the tumour cells represent only 1% of the entire cell population. The tumour microenvironment is essential to support the tumour cells and the type of microenvironment influences the outcomes in the case of Hodgkin lymphoma. While the morphological appearance of the tumour cells and microenvironment has been used only to diagnose Hodgkin lymphoma and not to identify the prognosis of the disease, artificial intelligence, and deep learning approaches can be used to predict the severity of Hodgkin lymphoma by looking at the morphological appearance of the tumour cells and the microenvironment, thus identifying the best treatment plan.

2. **Using PET-CT results**

PET-CT results can be used to measure the standardized uptake value of the tumor sites, which can further be used to assess the staging and severity of the disease. The artificial intelligence system will learn about

<b>Score</b>	<b>PET/CT Scan Result</b>
1	No uptake above background
2	Uptake less than mediastinum
3	Uptake greater than mediastinum but less than liver
4	Uptake moderately increased compared to the liver at any site
5	Uptake markedly increased compared to the liver at any site
X	New areas of uptake unlikely to be related to lymphoma

the tumor and provide a score at the end of the first stage, which can be correlated to the severity of the disease from the pathological study. The following table shows the scores corresponding to the PET-CT scan results.

Information provided by the PET-CT scan is important for response evaluation after completion of treatment based on lymphoma guidelines. With a specificity of 90-95%, the scan can be used for staging, interim, end of treatment, and follow-up surveillance. Interim PET is a good predictor of response in Hodgkin Lymphoma as indicated by interim assessment after initial cycles of chemotherapy with the aim of identifying which patients are cured and which need escalation of treatment.

## 5 My Role and Contribution

The basic concept of using machine learning and computer algorithms to assess the presence of disease, became possible because of modern microscopy, resulting in whole slide imaging (WSI). Machine learning works by training an algorithm to make suitable predictions, and the training of the algorithm is based on a comparison with annotations made by expert pathologists. One important aspect of analyzing the images is to create a segmentation map, which effectively assigns some meaning to every single pixel on the image – health tissue versus diseased tissue, foreground versus background, etc. The ability of computer algorithms to generate segmentation maps dramatically improves the quality and reproducibility of the work done by researchers and pathologists. A whole slide image is a digital image produced when glass slides are scanned. There have been significant advances in the hardware and software for whole slide scanning and their access, respectively. Whole slide imaging consists of two processes. The first process employs special scanners to generate large digital images from glass slides. The second process utilizes software to view and analyze the digital files in which the images are contained. Whole slide imaging has been used for a variety of medical purposes such as telepathology for primary diagnosis, remotely interpreting frozen sections and viewing immunostains, showcasing pathology slides at tumour boards, and performing image analysis.

## 5.1 Characteristics of WSIs

### 1. Image dimensions and data size

WSIs are large. A typical sample may be 20mm x 15mm in size, with a resolution of .25 micrometres/pixel. Thus, the resulting image is about 80,000 x 60,000 pixels or 4.8Gp. Images are usually captured with 24-bit colour, so the image data size is about 15GB. Although this is a typical example, larger images may also be captured. For example, sample sizes up to 50mm x 25mm may be captured from conventional 1 x 3 slides, and even larger samples may exist on 2 x 3 slides. Furthermore, images may be digitized at resolutions higher than .25mpp, resulting in data which can take up to 3.75 TB.

### 2. Access patterns

Pathologists cannot observe a complete sample at high resolution on a microscope slide due to the abundance of information there. Instead, they often pan the slide at a low resolution of 5 mpp (2X) or 2.5 mpp (4X) before zooming in at a greater resolution to focus on specific diagnostically important areas. Pathologists often focus while panning and zooming. Hence, it is important for the software for viewing WSIs to provide rapid panning and zooming capabilities.

With a view to encourage researchers to come out with AI models in assisting analysis of WSI for detection, the prognosis of treatment, etc, a few open challenges have been instituted by global research agencies and universities. The research team at IIT Bombay participated in two such challenges, with a view to improve the understanding of building algorithms to process and segment WSIs of cancer – The Camelyon 17 grand challenge and the TIGER grand challenge.

## 5.2 Camelyon17 Grand Challenge

I was initially introduced to the idea of image processing through the Camelyon17 dataset. The challenge focuses on the detection and classification of breast cancer metastases in lymph nodes. Metastatic involvement of lymph nodes is a very important prognostic factor in breast cancer. Lymph nodes are surgically removed for microscopic examination, resulting in a tedious and time-consuming diagnostic procedure for pathologists. Moreover, small metastases are difficult to detect and are often missed. The TNM (Tumour, Nodes, and Metastases) Classification of Malignant Tumors is a globally recognized standard for classifying the extent of the spread of cancer and is one of the most important tools for clinicians to help them select a suitable treatment option and to obtain an indication of prognosis. In the case of breast cancer, TNM staging takes into account the size of the tumor (T-stage), whether cancer has spread to the regional lymph nodes (N-stage), and whether the tumor has metastasized to other parts of the body (M-stage). The Camelyon17 grand challenge only focuses on the pathological N stage.

### 5.3 Camelyon17 WILDS

I worked on a separate dataset, in which the 50 whole slide images present in the original dataset have been broken into 455954 patches. Each whole slide image was manually annotated by pathologists to show tumor regions, and the resulting segmentation masks were used to determine the labels (0-tumor absent, 1-tumor present) for each patch. The data is provided in the metadata.csv file, which I have analyzed using the PANDAS software.

The goal of this challenge is to evaluate new and existing algorithms for automated detection and classification of breast cancer metastases in whole-slide images of histological lymph node sections. The presence of metastases in lymph nodes has therapeutic implications for breast cancer patients. Thus, accurate machine learning and artificial intelligence models hold great promise to reduce the workload of pathologists and reduce the subjectivity in diagnosis.

The goal of Camelyon-17 WILDS is to develop a machine learning model to detect tumours, which could be generalized to a larger number of hospitals. The challenge requires training models that generalize to data from a hospital that is not in the training set. The task is to predict if a given region of tissue contains any tumour tissue, which is considered a binary classification. The input was a 96x96 patch of tissue extracted from a whole-slide image (WSI) of a breast cancer metastasis in a lymph node section and the output would be a label, which is a binary indicator of whether the tissue patch contains any tumour or not.

#### 5.3.1 Data

The data in the dataset is split by the domain (the hospital that the data were taken from):

1. Training: **302,436** patches taken from 30 WSIs, with 10 WSIs from each of the 3 hospitals in the training set.
2. Validation (OOD): **34,904** patches were taken from 10 WSIs from the 4th hospital. These WSIs are distinct from those in the other splits.
3. Test (OOD): **85,054** patches were taken from 10 WSIs from the 5th hospital, which was chosen because its patches were the most visually distinctive. These WSIs are also distinct from those in the other splits.
4. Validation (ID): **33,560** patches were taken from the same 30 WSIs from the training hospitals.

Under the guidance of Dr Venkatapathy Subramanian, I started with fundamental manipulation and statistical analysis of the dataset using the PANDAS software. Initially, I used the command line to download the data as well as install the PANDAS software. Then, I proceeded to write code on PyCharm to find the column headings, number of rows (number of patches), number of patients, etc. and summarized the data in tables.



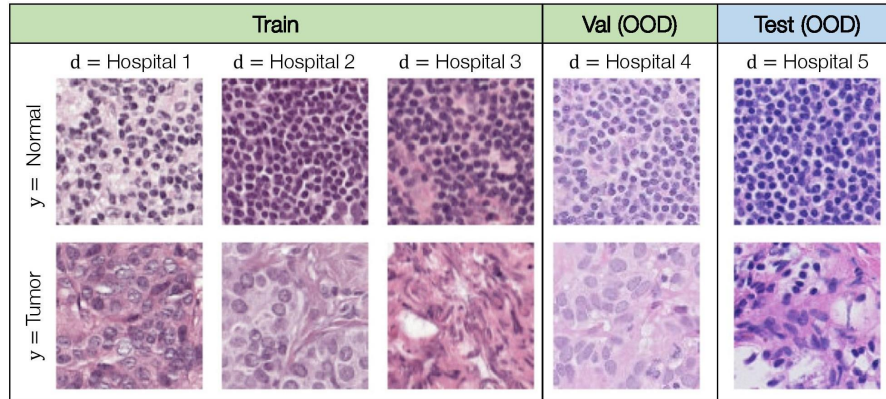


Figure 1: Patches from whole slide images taken from different hospitals

### 5.3.2 Analysis of dataset using python

Please find a link to the colab notebook where I have implemented some code for this challenge:

[Link to Camelyon17 WILDS Tutorial](#)

### 5.3.3 Takeaways

Through the Camelyon-17 grand challenge and the WILDS dataset, I was introduced to the concept of a whole slide image (WSI), and the idea of extracting patches from a large image which can be used for the training of machine learning models. I was also introduced to PANDAS, a python package which is extremely useful to manipulate and analyse large quantities of data. I was able to print the various features and statistics of the WILDS dataset through PANDAS and get a good understanding of the same.

## 5.4 TIGER Grand Challenge

TIGER is an acronym for Tumor Infiltrating Lymphocytes in Breast Cancer. TIGER is the first challenge on fully automated assessment of tumor-infiltrating lymphocytes (TILs) in HE breast cancer slides. It is organized by the Diagnostic Image Analysis Group (DIAG) of the Radboud University Medical Center (Radboudumc) in Nijmegen (The Netherlands), in close collaboration with the International Immuno-Oncology Biomarker Working Group.

### 5.4.1 Goals

- Development of AI models for automated quantification of Tumor Infiltrating Lymphocytes (TILs) in breast cancer, thus aiming at enabling the creation of open AI-based solutions that can automate TILs assessment

- Validation of the prognostic value of AI-based automated TILs scores using a large independent test dataset that includes cases from both clinical routine and from a phase 3 clinical trial

#### 5.4.2 Tasks

- Detection of lymphocytes and plasma cells, which are the main types of cells considered as tumor-infiltrating lymphocytes
- Segmentation of invasive tumour and tumor-associated stroma, which are the main tissue compartments considered when identifying relevant regions for the TILs
- Computation of an automated TILs score based on the output of detection and segmentation

#### 5.4.3 Data

##### **WSI ROIS**

##### **Whole slide images**

There are **195** whole slides images in WSI ROIS

- **151** are from TCGA (The Cancer Genome Atlas Program)
- **26** are from RUMC (Radboud University Medical Center)
- **18** from JB (Jules Bordet Institute)

##### **ROI Size**

- In the RUMC and JB slides, there are 3 regions of interest per slide, with size approximately 500 X 500 microns
- In TCGA slides, there are variable number and smaller regions of interest

##### **Annotations**

There are various labels as a part of the annotations:

- Label 1 – Invasive tumour
- Label 2 – Tumour-associated stroma. Tumour-associated stroma is connected tissue associated with the tumour. Some stroma may resemble healthy stroma.
- Label 3 – In situ tumor such as Ductal Carcinoma In Situ (DCIS). DCIS is the presence of abnormal cells inside a milk duct in the breast. DCIS is considered the earliest form of breast cancer. DCIS is noninvasive, meaning it hasn't spread out of the milk duct and has a low risk of becoming invasive). In-situ means in the original place, so the cancer cells have not spread out yet.

- Label 4 – Healthy glands having healthy epithelial cells
- Label 5 – Necrosis not in situ. Regions of necrotic tissue not a part of in-situ tumour. For example, ductal carcinoma in situ (DCIS) often presents a typical necrotic pattern, which can be considered as part of the lesion itself, such a necrotic region is not annotated as "necrosis" but as "in-situ tumor".
- Label 6 – Inflamed Stroma. This class contains tumour-associated stroma that has a high density of lymphocytes (i.e., it is "inflamed"). When it comes to assessing the TILs, inflamed stroma and tumour-associated stroma can be considered together, but were annotated separately to take into account differences in their visual patterns; so for the programming purpose, these can be considered as the same probably.
- Label 7 – Rest. This class contains regions of several tissue compartments that are not specifically annotated in the other categories; examples are healthy stroma, erythrocytes, adipose tissue, skin, nipple, etc.

Additionally, most ROIs contain annotations of lymphocytes and plasma cells in the form of bounding boxes. Cells were annotated using point annotations and then squared bounding boxes of 8x8 microns were constructed centred on the point annotation. This fixed-size bounding box size is inspired by previous work on lymphocyte detection, where an average equivalent diameter of 8 microns was used for lymphocytes.

There are 2 types of annotation, in 2 subfolders:

#### WSI level annotation

- This contains several folders. The Images folder contains images, which are named on the basis of their origin. B represents biopsy and S represents a surgical resection.
- The annotations are in the tissue mask folder and can be opened through ASAP as an overlay. Visualization should be enabled to exclude the background pixels in the algorithm.
- Annotations in the mask folder are TIF files and the other is XML file type. In this, we get the ROIs. In the XML files, we also have the bounding boxes which are manual annotations of lymphocytes and plasma cells.
- TCGA annotations derived from two data sets – BCSS and NuCLS. BCSS provides manual annotations of tissue compartments only in both formats (masks and xml). NuCLS provides manual annotations of different types of cells including lymphocytes and plasma cells

#### ROI level annotation

- There are 2 folders – Tissue Cells and Tissue BCSS. The annotations are the exact same as WSI images except that these are restricted to the ROI. The rest has been cropped out. This may be easier to use than WSI annotations while formulating the algorithm.

- The images are in the form of png files and the cell annotations are in coco format.

### **WSI BULK**

#### **Whole slide images**

There are **93** whole slide images in WSI BULK, all of which are confined within the manually annotated region.

#### **Annotations**

- The images with annotations are in the Tissue Mask folder. They have to be opened ASAP using an overlay.
- The images have coarse annotation, which means that all tumour cells are inside the annotated part. However, there are also other cells and healthy tissue inside the annotated part. So the algorithm has to be able to differentiate between tumour cells and other non-clinically relevant areas on the image.

### **WSI TILS**

#### **Whole slide images**

There are **82** whole slide images in WSI TILS. There are no annotations, but the TIL score is provided, and several pitfalls are mentioned, which may need to be studied while developing the algorithm.

The information can be used as calibration for TIL score as each slide has a TIL score based on visual analysis by an expert. The colours represent various things like invasive tumours, tumour-associated stroma etc.

After I had a detailed understanding of the dataset, I worked on extracting patches from a WSI in the wsibulk folder. Like for the Camelyon-WILDS data, I used the command line to create a virtual environment and install necessary software programs like matplotlib (which would help print the output of the patch extracted as well as the WSI), Openslide, Jupyter lab, as well as the TIF file of the image from AWS. Then, I used the Jupyter notebook and code from the tutorials on the grand challenge website to extract patches of different resolutions and sizes from the WSI. I encountered errors often while running the code which I have explained below.

#### **5.4.4 Troubleshooting**

*Registrant 'openslide' is not found in the register of class 'WholeSlideImage-Backend' with an empty register.*

This error is caused due to:

- **Inability to install OpenSlide using Port**

To fix, use Homebrew to install OpenSlide with the command: brew install openslide

- **Starting the notebook with the wrong python executable**

To fix, type the following command in the terminal: `python -m ipykernel install --user --name=tigervenv`

Then, in the Jupyter notebook, change the python executable from Python 3 (ipykernel) to tigervenv using the drop down towards the top right of the page.

#### 5.4.5 Takeaways

In the TIGER challenge, I was introduced to the installation of whole slide images and extracting patches from them. I was introduced to terms like spacing, width, height, and centre, in the context of image manipulation, which I could link to the WILDS metadata file. I also learned to debug code, since I faced many errors along the way. Overall, I received a thorough understanding of the TIGER dataset and tried my hand at extracting patches from a WSI.

### 5.5 Tata Memorial Project

The project involved a field trip to the Tata Memorial Hospital, the aim of which was to discuss the goals and feasibility of the project with the doctors and collect relevant data. I got an idea of how doctors assess Hodgkin lymphoma and the challenges that the project would present.

- Size of the slide is variable and depends upon the patient. It is up to us to choose the standard dimension of the size and trim out what is outside the area.
- Lymphoid cells are of different shapes with the presence of plasma cells, epithelial cells, eosinophils, etc.
- Variations in the size of tumour cells present from what is known. Hence bystander cells may be picked up or tumour cells can be missed in the annotations.
- While cutting micron thick sections, cells appear to look bigger or smaller as the cuts are taken from separate planes. We have to consider parameters such as nuclear outline along with nuclear size.
- Neoplastic cells in Hodgkin lymphoma are slow growing and there is no longitudinal data.

I also received an elaborate understanding of Hodgkin lymphoma. I learnt that it is composed of tumour cells and the tissue environment.

- Tumor cells - Criteria for the model to assess the tumour cells could be size, the pattern of distribution, density, and number.

- Tissue environment - There is no hard and fast way to analyze the tissue environment. The model could consider the number and density of eosinophils, macrophage reach, and mitotic activation.

I realized the differences between the TIGER and Camelyon17 challenges and this project. In the case of Hodgkin lymphoma, the neoplastic cells are in the minority and the tissue environment is variable. However, this is not true for the grand challenge, as the main focus is the tumour cells, which are the majority. The goal for the team after speaking to the doctor is to design a model based on all of the factors listed above. The idea is to use digital microscopy to get more information which will help decide the treatment given, thus making the choice more objective in nature.

The data we received from the doctors in the hospital was in the iSyntax image format, which is designed to combine the medical grade image quality of JPEG 2000 with the speed of JPEG.

#### 5.5.1 Advantages of iSyntax files

- While iSyntax employs the same recursive wavelet transformations as JPEG 2000, it replaces the entropy coding part with a completely new method, which finds a more compact way to store the same information.
- The entropy coder has no influence on the image quality of the compressed image, but only on the compression ratio and speed.
- iSyntax' speed and cloud-enabled data model allow the creation of true cloud-deployed solutions for digital pathology that benefit from the cloud's virtually infinite scalability and cost-effectiveness.

## 6 Takeaways

Overall, the project was extremely informative and gave me a large degree of experience in using programming for medical purposes. The structure of the project, which began with understanding grand challenges such as the Camelyon and TIGER challenges equipped me with the skillset necessary for contributing effectively to the TATA Memorial Project. I received plenty of knowledge about the pathological field as well as ways to integrate pathology and technology to improve treatments. I also learned a plethora of skills such as the analysis of large quantities of data, the manipulation of images using python, extracting relevant patches from whole slide images, as well as debugging code, which will prove to be extremely important for future projects I may undertake. Moreover, the opportunity to visit a hospital, and be a part of the discussion of a large-scale project gave me first-hand experience of being part of a team working together to solve a problem in the real world.

Using information from what I learnt in this project, I contributed to the [PAN-DAS 1](#) and [PANDAS 2](#) tutorials in the book titled "Informatics AI in Healthcare" by Dr. Ganesh Ramakrishnan.

## 7 References

Press, Dove, et al. “Whole Slide Imaging in Pathology: Advantages, Limitations, and Emergin — PLMI.” Whole Slide Imaging in Pathology: Advantages, Limitations, and Emergin — PLMI, [www.dovepress.com](http://www.dovepress.com)

“Classic Hodgkin Lymphoma.” Pathology Outlines - Classic Hodgkin Lymphoma, [www.pathologyoutlines.com](http://www.pathologyoutlines.com), 1 Dec. 2021

DIAGNijmegen. “Pathology-Whole-Slide-Data/Notebooks at Main · DIAGNijmegen/Pathology-Whole-Slide-Data.” GitHub, [github.com](https://github.com)

“Datasets.” Datasets, [wilds.stanford.edu](http://wilds.stanford.edu)

“TIGER - Grand Challenge.” Grand-Challenge.Org, [tiger.grand-challenge.org](http://tiger.grand-challenge.org)

“CAMELYON17 - Grand Challenge.” Grand-Challenge.Org, [camelyon17.grand-challenge.org](http://camelyon17.grand-challenge.org)

“Digital Health — NGO Transform.” Digital Health — NGO Transform, [www.koitafoundation.org](http://www.koitafoundation.org)

“Koita Centre for Digital Health.” Koita Centre for Digital Health, [www.kcdh.iitb.ac.in](http://www.kcdh.iitb.ac.in)

“AI for Breast Cancer Detection Using Whole-Slide Images.” AI for Breast Cancer Detection Using Whole-Slide Images, [www.kantify.com](http://www.kantify.com)

“Home - International TILS Working Group.” International TILS Working Group, [www.tilsinbreastcancer.org](http://www.tilsinbreastcancer.org), 20 Jan. 2022

Thepathologist.Com, 2022