AI meets Pathology Education: Teaching how to use Vision Transformer as a Quality Assurance tool to rule out classical Hodgkin Lymphoma

Andy Nguyen, M.D., M.S.

Vice Chair, Digital Pathology
Dr. John D. Milam Endowed Chair in Pathology
Professor of Pathology and Laboratory Medicine
UTHealth, McGovern Medical School

Intelligence Amplified: AI Meets Med Ed

2025 [pending]

UTHealth | McGovern Medical School
The University of Texas Health Science Center at Houston

Outline of talk

- In this study we describe our attempt to teach our pathology residents how to use AI to assist with diagnosis and use AI as a QA tool

- We focus on a single goal of classifying a case as classical Hodgkin lymphoma (cHL) versus non-classical Hodgkin lymphoma (non-cHL)

Introduction to Vision Transformer as a QA Tool

- In the field of haematopathology, artificial intelligence (AI) models have shown promising results for the detection and classification of lymphomas from whole-slide images (WSIs) with areas under the receiver operating characteristic curve (AUROC) often exceeding 0.90

- However, all these algorithms focus on diagnosing only a few subtypes of the most common lymphomas, such as diffuse large B-cell lymphoma (DLBCL), follicular lym phoma (FL), and chronic lymphocytic leukaemia/ lymphocytic lymphoma (CLL), and are far from being able to diagnose the range of lymphoid lesions identified in international classifications. Therefore, the AI approaches currently published do not replicate, to date, the diagnostic approach of the pathologist who, in daily practice, must be able to recognize nearly 100 different subtypes of lymphomas in addition to the reactive lesions, which are also very varied.

Introduction to Vision Transformer as a QA Tool (cont'd)

In this initiative we attempt to narrow down on the scope of AI coverage. We focus on a single goal of classifying a case as classical Hodgkin lymphoma (cHL) versus negative for classical Hodgkin lymphoma. The rationale behind this goal is three-fold:

- 1.cHL is a relatively common lymphoma that practicing pathologists encounter rather often
- 2.Our group already designed and validated a Vision Transformer module to predict cHL versus anaplastic large cell lymphoma (ALCL). Since ALCL is not a common lymphoma, the module can be best utilized to rule out cHL. A prediction of cHL is treated as a confirmatory one for cHL; wheras a prediction of ALCL would be considered as non-cHL lymphoma
- 3.The preliminary results from selected cases can be used as teaching materials to show our pathology residents and fellows how to apply AI in routine practice. This application, if used on a routine basis, may be useful as a QA tool to prevent misdiagnosis.

Introduction to Vision Transformer

- While the Transformer architecture (such as in ChatGPT, Copilot, etc., based on large language model) has become the de-facto standard for natural language processing tasks, its applications to computer vision remain limited.

- Alexey Dosovitskiy et al, *An image is worth 16x16 words: transformers for image recognition at scale.* International Conference on Learning Representations 2021

  Apply transformer model directly to images-> Vision Transformer (ViT)

- Self-attention in ViT allows each part of image to relate (pay attention) to other parts of image, regardless of the distance between them.

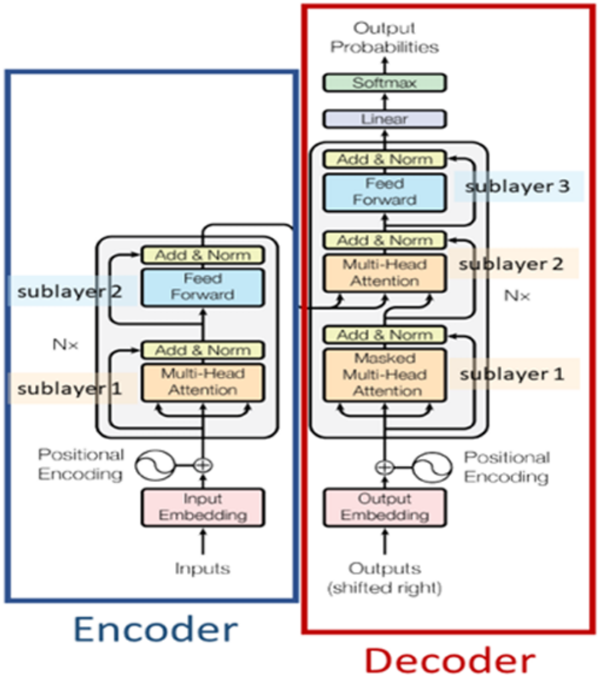
- ViT has been used to build Foundation models which are trained on very large datasets, using self-supervised learning, which do not require labeled labels.

  They can be finetuned for a wide range of downstream tasks using a modest amount of task-specific labeled data for training.

Transformer Model

Transformers are neural networks that use a self-attention mechanism to capture relationships between input elements, especially in long sequences. They can process and learn from all data types, including text, and speech

2017-TRANSFORMERS



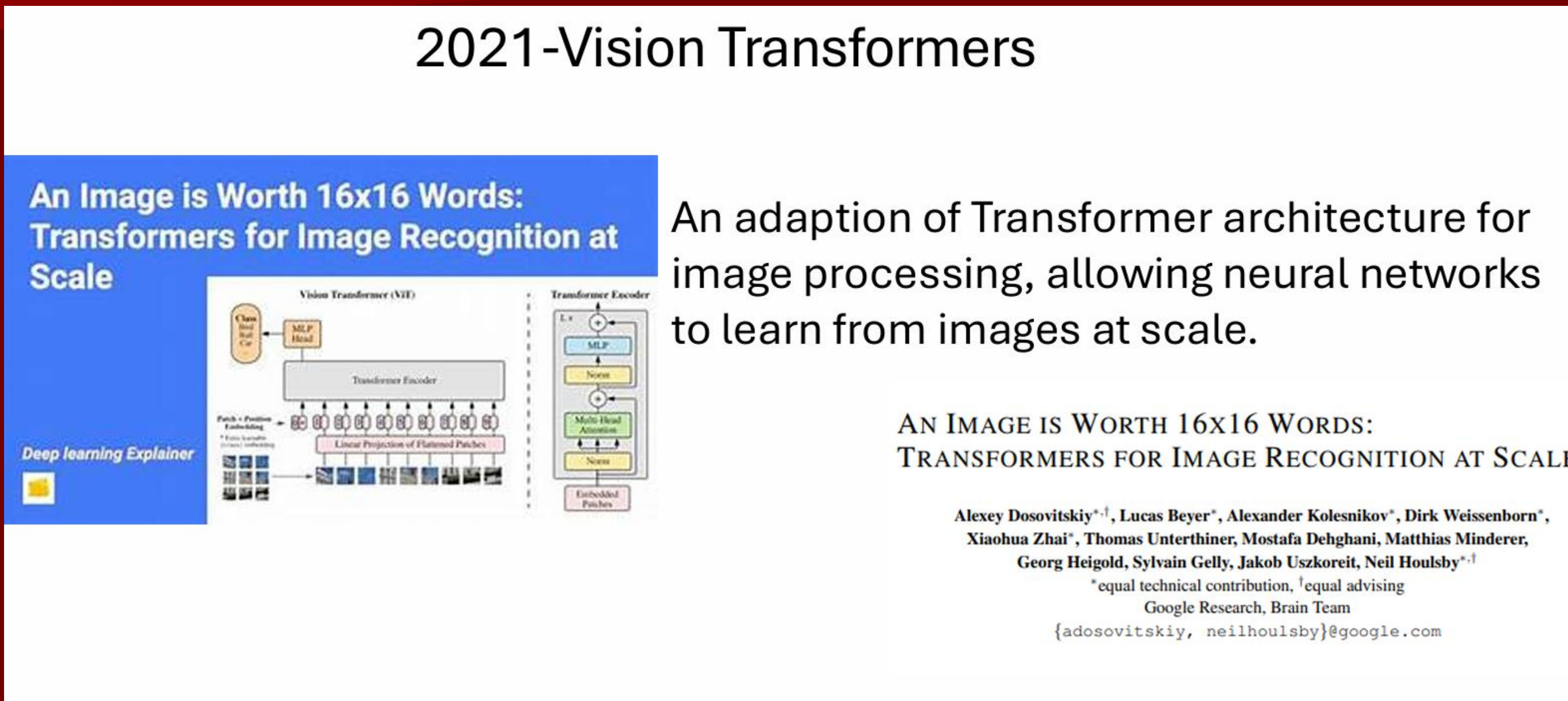
- Unlike CNN, train on data via parallel processing rather than sequentially and a mechanism called self-attention
- Can be scaled massively allowing it to train on unlimited text datasets

Attention Is All You Need

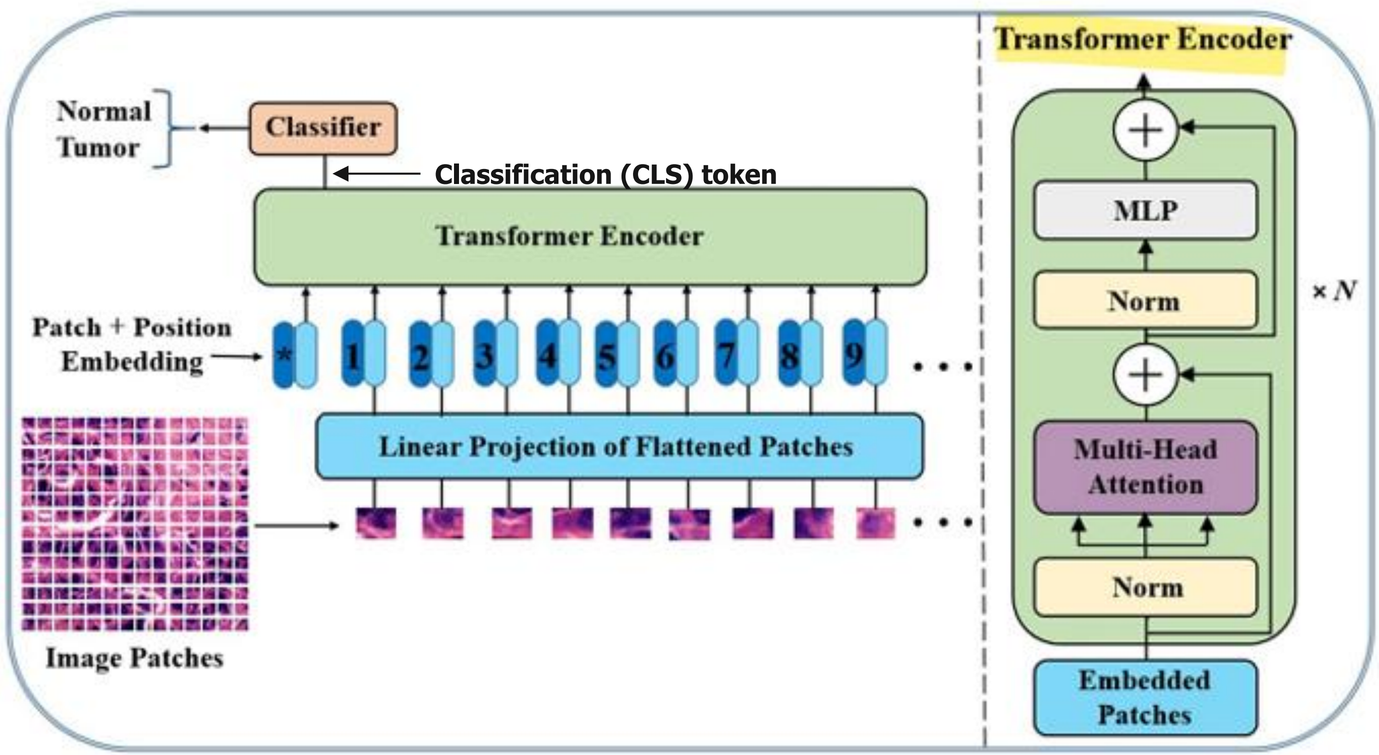
Ashish Vaswani* Google Brain avaswani@google.com
Noam Shazeer* Google Brain noam@google.com
Niki Parmar* Google Research nikip@google.com
Jakob Uszkoreit* Google Research usz@google.com
Llion Jones* Google Research llion@google.com
Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu
Łukasz Kaiser* Google Brain lukaszkaiser@google.com
Illia Polosukhin* ‡ illia.polosukhin@gmail.com

Ashish Vaswani, 31st Conference on Neural Information Processing Systems (NIPS 2017)

Vision Transformer Model



Alexey Dosovitskiy et al, *An image is worth 16x16 words: transformers for image recognition at scale.* International Conference on Learning Representations 2021



A schematic diagram of a standard ViT model. Sequential image patches are used as the input, which is then processed with a transformer encoder and uses an MLP head module to generate a class prediction

- ViT model: encoder only, inherited architecture from that of a standard transformer.
- It splits the image into a grid of non-overlapping patches before passing them to a linear projection layer as tokens. These tokens are then processed by a series of multi-headed self-attention layers to capture global relationship

Validating a Vision Transformer module to predict cHL vs. ALCL

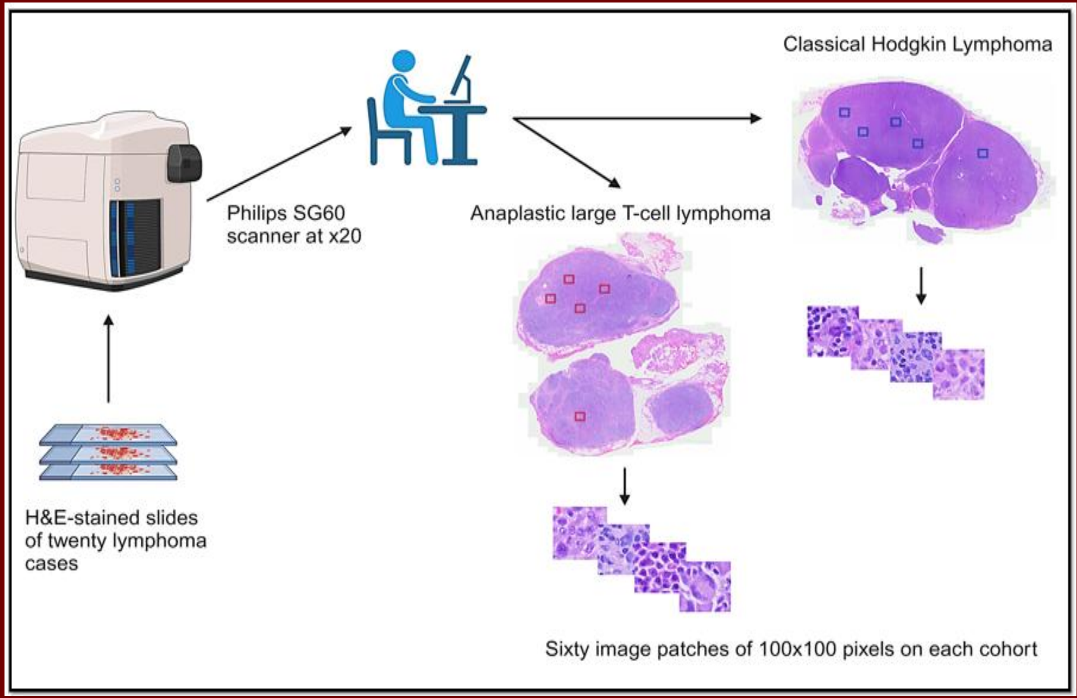
- We conducted a retrospective compilation of cases with diagnosed cHL and ALCL by current World Health Organization criteria at our institution from 2017 to 2024. We reviewed the morphological characteristics of each case and selected the hematoxylin and eosin- (H&E) stained slides from 20 cases, which were scanned using the SG60 scanner (Philips Corporation, Amsterdam, Netherlands) at 40x magnification.
- The SG60 scanner has capacity for 60 glass slides, produces high-quality images, full automation (for focus, calibration, brightness and contrast settings), with tissue shape detection to outline and scan non-rectangular regions of interest for shorter turnaround times.
- The total scan time of a slide for a 15 × 15 mm benchmark scan area at a 40 x resolution is ≤ 62 seconds. The images were acquired and stored in iSyntax2 format. Philips Image Management System was used to display the images.

Materials and Methods

- From each WSI, 60 image patches of 100x100 pixels (at 20x magnification, 0.5 µm/pixel) were obtained for feature extraction with SnagIt software (TechSmith Corp, Okemos, Michigan, USA).

- A total of 1200 image patches were obtained from which 1079 (90%) were used for training, 108 (9%) for validation, and 120 (10%) for testing.

- The cases were divided into two cohorts, with 10 cases for each diagnostic category. For each test set of 5 images, the expected diagnosis was combined from the prediction of five images, i.e., majority voting, at least three or more must agree to be considered as the predicted result.



SG-60 Scanners in DPALM Digital Pathology

Microscopic Lens

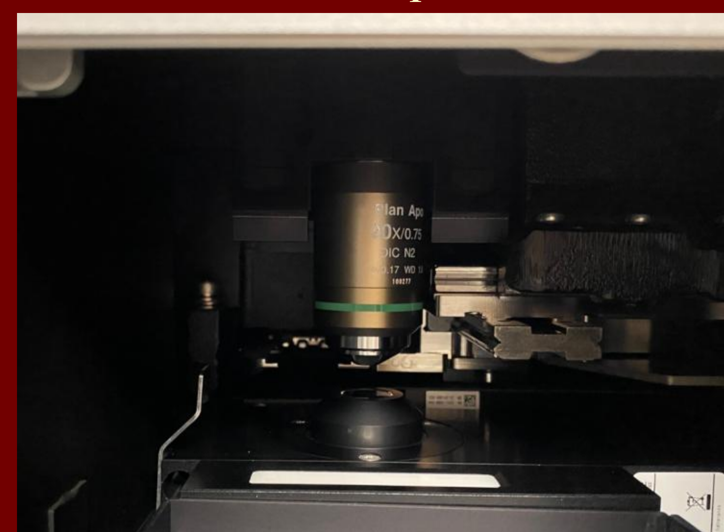
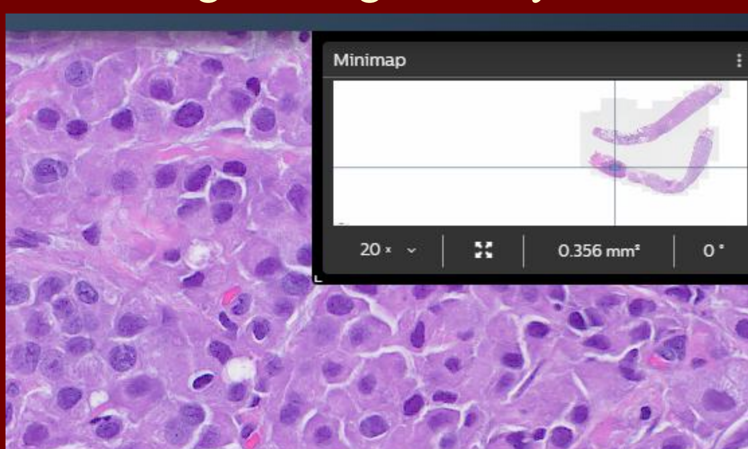


Image management system



Materials and Methods (cont'd)

- Hardware:
  Intel Xeon Gold 5222 CPU; 48 GB RAM

  GPU: NVIDIA RTX A4000 (16 GB, 6144 CUDA cores)

- Software:
  Windows 11-64 bit
  Python, Torch, Torchvision to build Vision Transformer model
  (using supervised training with labeled data)
  CUDA for parallel processing

A Snippet of Python Code

```python
# PART 1***: CORE CODE FOR THE ViT ENGINE:

import torch
import torch.nn as nn
import math
import torch.multiprocessing as mp
import matplotlib.pyplot as plt
import numpy as np
import torchvision
import torchvision.transforms as transforms
import torch.optim as optim
import warnings

# Suppress the display of many unloaded modules in beginning of run
warnings.filterwarnings("ignore", message="Reloaded")

# use GPU if available; otherwise use CPU; display the device used (GPU or CPU)
device = torch.device("cuda" if torch.cuda.is_available() else "cpu")
print(f"Using device: {device}")

#Initialize the model parameters (need to set as needed before each run):
img_size = 100    # size of images, for example 100 for 100x100 images
patch_size = 20    # size of image patches, for example 5 for 5x5 image patches
d_model = 128     # the dimensionality of the model: commonly used values for d_model in practice are 128, 256, 512, or 1024
num_heads = 4     # number of attention heads; it evenly divides the d_model dimension; i.e. d_model/num_heads=integer
num_layers = 6    # number of transformer layers; commonly used values are between 6 and 12
lr=0.001          # learning rate
num_epochs = 200   # number of epochs
num_classes = 2   # number of prediction classes
batch_size=32     # number of cases in each reading batch for datasets

# Create a Transformer model instance
transformer_model = nn.Transformer()

# Print the default configuration
# print(transformer_model) # suppressed display for now (very verbose)

# Determining the dim_feedforward parameter
# dim_feedforward: the dimensionality of the hidden layer in the feed-forward neural network
dim_feedforward = transformer_model.encoder.layers[0].linear1.weight.size(0)

# Positional Encoding with Sine and Cosine
class PositionalEncoding(nn.Module):
    def __init__(self, d_model, max_len=5000):
        super(PositionalEncoding, self).__init__()
        pe = torch.zeros(max_len, d_model)
        position = torch.arange(0, max_len, dtype=torch.float).unsqueeze(1)
        div_term = torch.exp(torch.arange(0, d_model, 2).float() * (-math.log(10000.0) / d_model))
        pe[:, 0::2] = torch.sin(position * div_term)
        pe[:, 1::2] = torch.cos(position * div_term)
        pe = pe.unsqueeze(0).transpose(0, 1)
        self.register_buffer('pe', pe)

    def forward(self, x):
        return x + self.pe[:x.size(0), :]
```

The scary part:
Most of the coding is assisted and debugged by AI (ChatGPT, CoPilot, etc.)

RESULTS

- Display from execution of the ViT model:

```
+++++++++++
VISION TRANSFORMER MODEL
Written in Python to assist diagnosing Anaplastic Large Cell Lymphoma versus
Classical Hodgkin Lymphoma

This model is based on pytorch, torchvision, and is designed with Multi-head Attention,
sine/cosine embedding function, and residual connections
.........
PARAMETERS OF MODEL:
Number of images in Trainning set:  1080
Number of images in Testing set:  120
Size of images:  100
Size of image patches:  20
Dimensionality of the model:  128
Number of attention heads:  4
Number of transformer layers:  6
learning rate:  0.001
Feedforward dimension:  2048
Number of epochs:  200
Number of prediction classes:  2
Number of cases in each reading batch for datasets:  32
+++++++++++++
TRAINING THE VISON TRANSFORMER MODEL:
Loss in each epoch, with number of test images in each Batch:
[Epoch 1, Batch of: 30] loss: 0.234
[Epoch 2, Batch of: 30] loss: 0.203
[Epoch 3, Batch of: 30] loss: 0.201
[Epoch 4, Batch of: 30] loss: 0.201
[Epoch 5, Batch of: 30] loss: 0.201
[Epoch 6, Batch of: 30] loss: 0.197
[Epoch 7, Batch of: 30] loss: 0.198
[Epoch 8, Batch of: 30] loss: 0.200
[Epoch 9, Batch of: 30] loss: 0.197
[Epoch 10, Batch of: 30] loss: 0.187
[Epoch 11, Batch of: 30] loss: 0.187
[Epoch 12, Batch of: 30] loss: 0.178
[Epoch 13, Batch of: 30] loss: 0.197
[Epoch 14, Batch of: 30] loss: 0.177
[Epoch 15, Batch of: 30] loss: 0.171
[Epoch 16, Batch of: 30] loss: 0.177
[Epoch 17, Batch of: 30] loss: 0.168
[Epoch 18, Batch of: 30] loss: 0.165
[Epoch 19, Batch of: 30] loss: 0.170
[Epoch 20, Batch of: 30] loss: 0.158
```

← Display of software introductory info

← Display of dataset info, hyperparameters used in execution

← Display of Loss in each Epoch

RESULTS Cont'd

- The test results from ViT model showed a diagnostic accuracy at 100% for 120 test images.

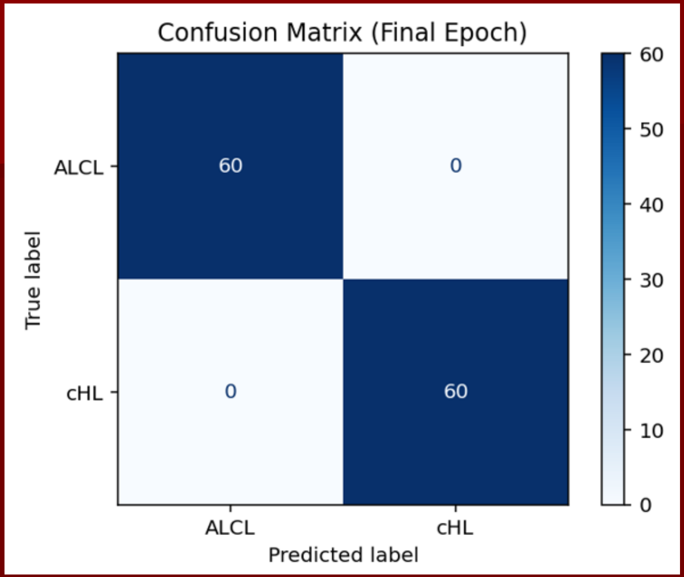
```
[Epoch 496, Batch of: 30] loss: 0.001
[Epoch 497, Batch of: 30] loss: 0.001
[Epoch 498, Batch of: 30] loss: 0.001
[Epoch 499, Batch of: 30] loss: 0.001
[Epoch 500, Batch of: 30] loss: 0.001
+++++++++++++++++
- Training was completed
- Results: accuracy of the network on 120  test images:  100.0 %
+++++++++++++++
- DX code: Anaplastic Large cell Lymphoma->0; Classical Hodgkin lymphoma->1
```

← Display of Loss in each Epoch (last part)

← Display of accuracy of prediction on 120 test images (100%)

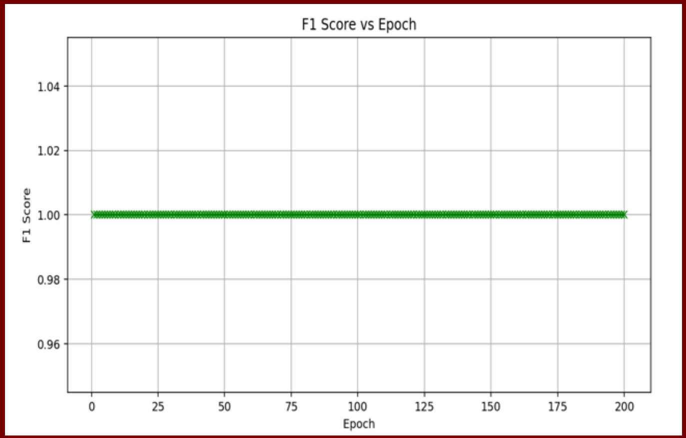
← Display of all 120 test cases (image, DX, predicted DX)

Measurement Metrics: Accuracy and F1 Score



Accuracy= (TN + TP) /(TN + FN + FP + TP)
=100%

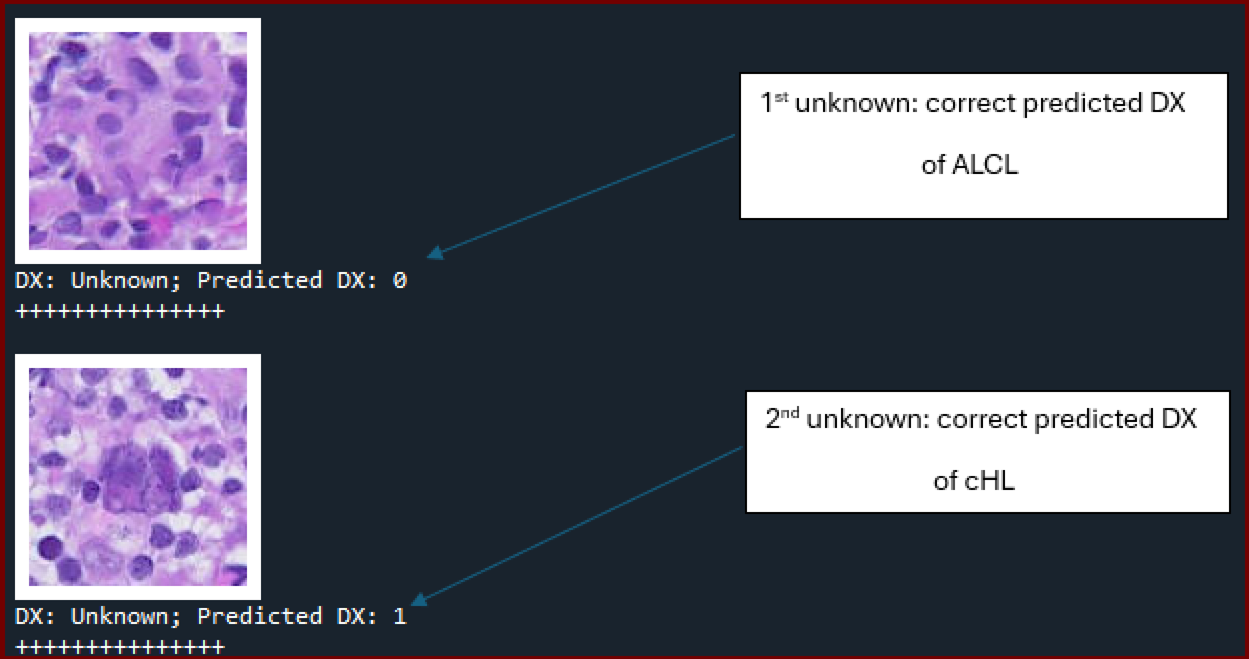
Where: TN=True Negative
TP=True Positive
FN=False Negative
FP=False Positive



F1 Score = 2 (Precision x Recall)/(Precision + Recall)
= 100%

Where: Precision=TP/(FP +TP)
Recall= TP/(FN + TP)
(F1 Score is useful for datasets containing uneven class distributions)

Production Protocol for Testing Unknown Images: Used in this teaching initiative



Two unknowns (“Test01.jpg” and “Test02.jpg”)

Teaching Module

- A total of 5 cases are included in this teaching project (2 cases of cHL and 3 cases of other lymphoma types). These cases are processed for H&E slides and scanned for whole-slide-image (WSI) at the Digital Pathology Laboratory at University of Texas Medical School-Houston. These cases have morphology findings such that cHL cannot be ruled out. They typically include cases with nodular lymphocyte predominant Hodgkin lymphoma, large B cell lymphoma, and others

- Two image patches of size 100x100 pixels were obtained from each cases (at 20x20 magnification) from whole-slide-image (WSI). They were included as two test images in the AI module which is based on the Vision Transformer model which had previously been developed, validated and tested with an accuracy of 100% (prediction of cHL vs ALCL).

Demo: Case 1

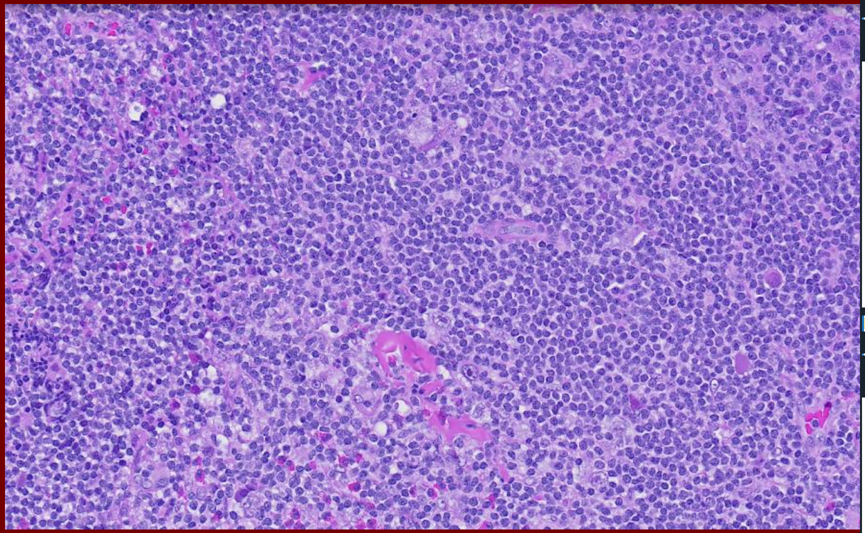
- S25-xxxx35: classical Hodgkin Lymphoma-> predicted: classical Hodgkin Lymphoma

SURGICAL PATHOLOGY REPORT

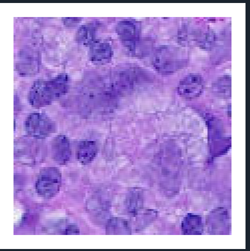
DIAGNOSIS:

Neck Mass Lymphoma Workup, Left (Excision):

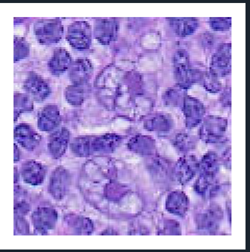
Classic Hodgkin lymphoma, nodular sclerosis subtype



```
+++++++++++++++
DX code:
- Anaplastic Large cell Lymphoma->0
- Classical Hodgkin lymphoma    ->1
***All unknown images are displayed with predicted DX:
+++++++++++++++
```



```
DX: Unknown; Predicted DX: 1
+++++++++++++++
```



```
DX: Unknown; Predicted DX: 1
+++++++++++++++
```

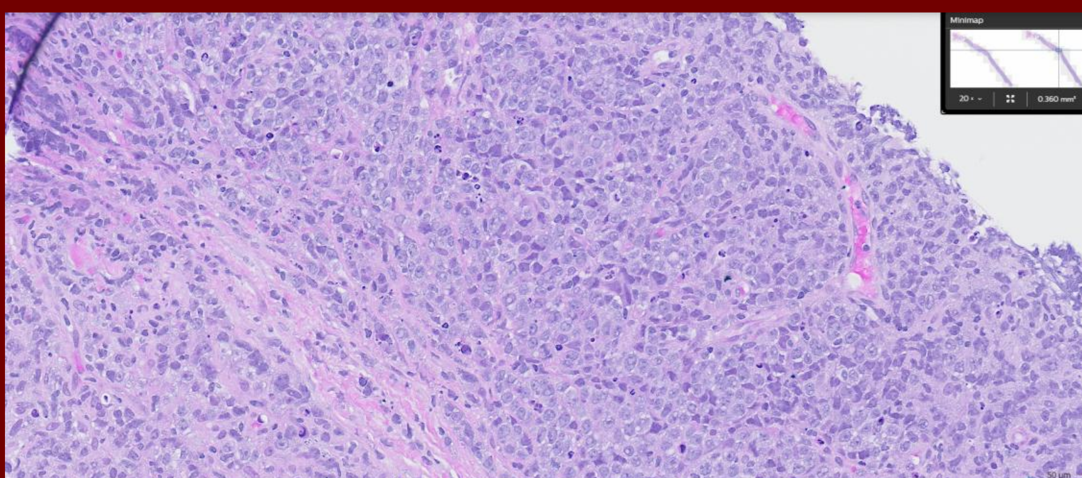
Demo: Case 2

- S25-xxx61: Diffuse Large B Cell Lymphoma-> predicted: non classical Hodgkin Lymphoma

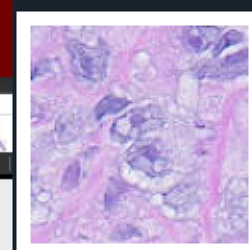
BIOPSY PATHOLOGY REPORT

DIAGNOSIS:

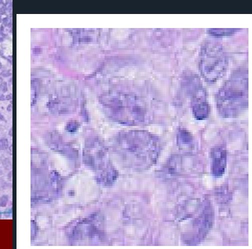
1. **Neck Core, Left (Biopsy):**
   Diffuse large B-cell lymphoma, germinal-center subtype with Ki67 at 80%, see comment



```
+++++++++++++++
DX code:
- Anaplastic Large cell Lymphoma->0
- Classical Hodgkin lymphoma    ->1
***All unknown images are displayed with predicted DX:
+++++++++++++++
```



```
DX: Unknown; Predicted DX: 0
+++++++++++++++
```



```
DX: Unknown; Predicted DX: 0
+++++++++++++++
```

Demo: Case 3

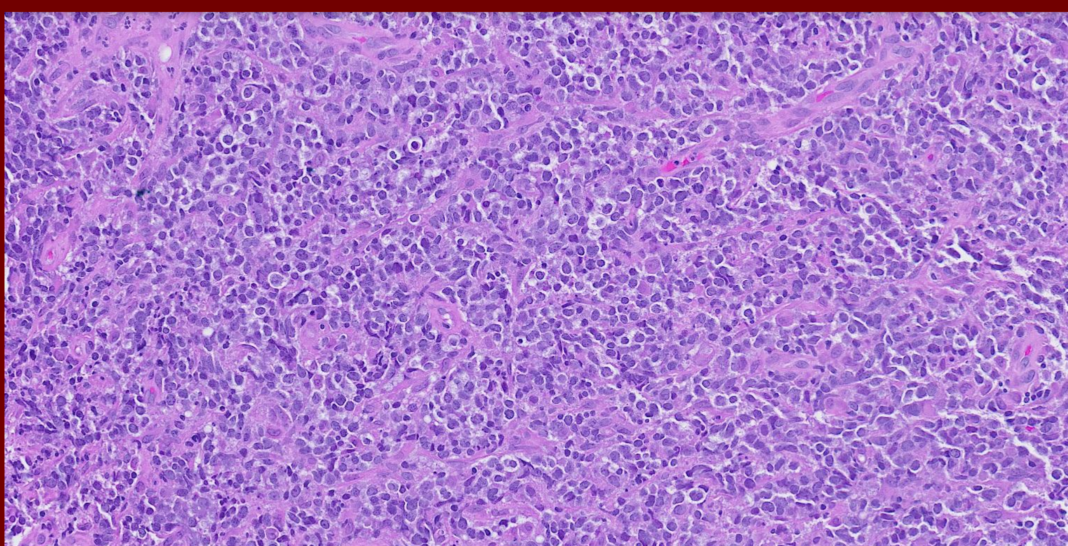
- S25-xxx66: Diffuse Large B Cell Lymphoma-> predicted: non classical Hodgkin Lymphoma

BIOPSY PATHOLOGY REPORT

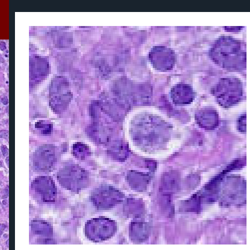
DIAGNOSIS:

Nasopharynx (Biopsy):

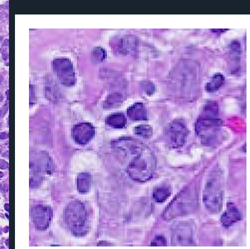
- Diffuse large B-cell lymphoma, non-germinal center immunophenotype, EBV-negative (see comment)
- Proliferative index of approximately 70-80% by Ki67 immunostaining



```
+++++++++++++++
DX code:
- Anaplastic Large cell Lymphoma->0
- Classical Hodgkin lymphoma    ->1
***All unknown images are displayed with predicted DX:
+++++++++++++++
```



```
DX: Unknown; Predicted DX: 0
+++++++++++++++
```



```
DX: Unknown; Predicted DX: 0
+++++++++++++++
```

Demo: Case 4

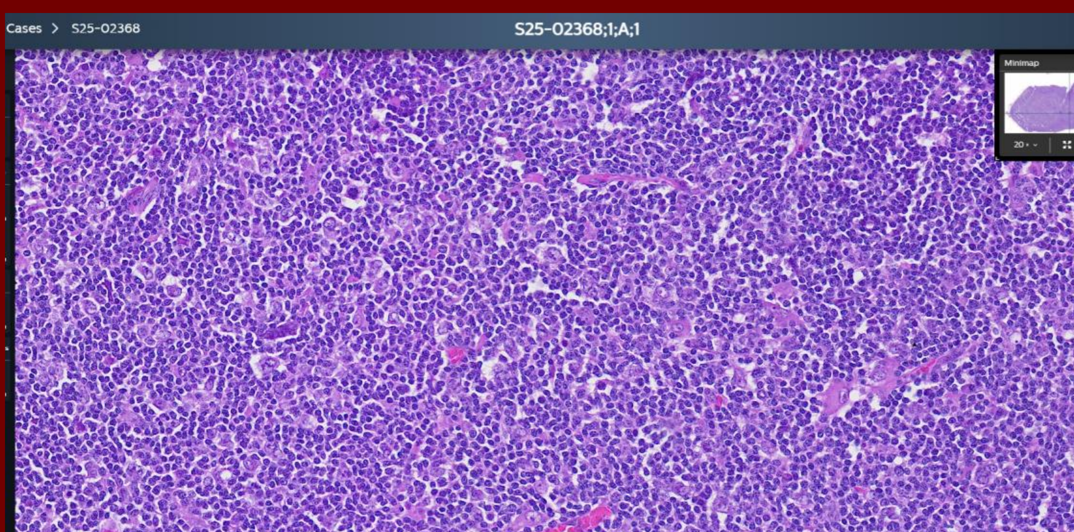
- S25-xxx68: Nodular Lymphocyte-Predominant Hodgkin Lymphoma-> predicted: classical Hodgkin Lymphoma (review of histology: typical for cHL, IHCs supportive of NLPHL)

BIOPSY PATHOLOGY REPORT

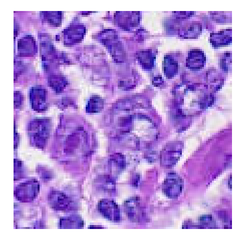
DIAGNOSIS:

Neck Mass, Left (Excision):

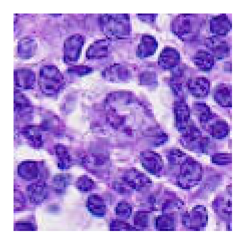
- **Nodular lymphocyte-predominant B-cell lymphoma/nodular lymphocyte-predominant Hodgkin lymphoma, patterns A and B (grade 1)**
- **EBV-negative by EBER in situ hybridization**




```
++++++++++++++
DX code:
- Anaplastic Large cell Lymphoma->0
- Classical Hodgkin lymphoma    ->1
***All unknown images are displayed with predicted DX:
++++++++++++++
```



```
DX: Unknown; Predicted DX: 1
++++++++++++++
```



```
DX: Unknown; Predicted DX: 1
++++++++++++++
```

Demo: Case 5

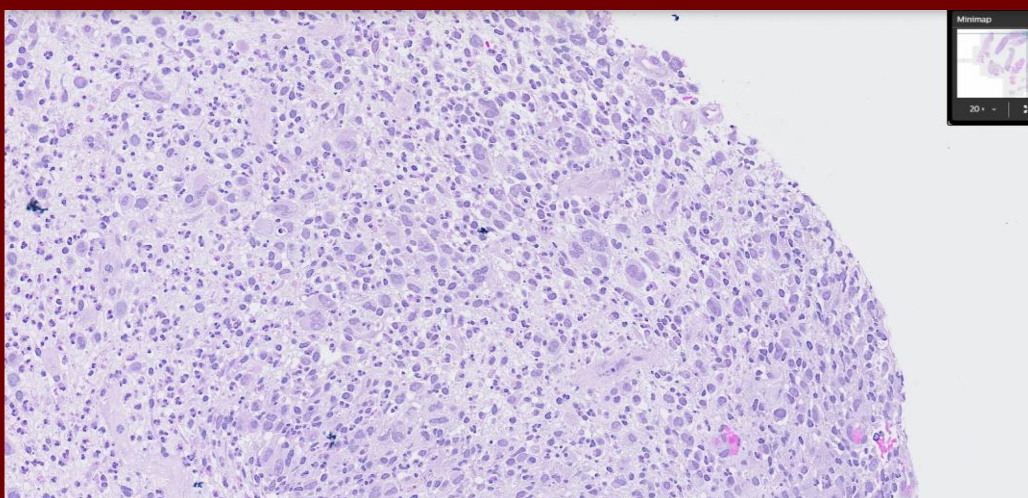
- S25-xxxx96: classical Hodgkin Lymphoma->predicted: non classical Hodgkin Lymphoma (review of histology: atypical for cHL, IHCs supportive of cHL)

DIAGNOSIS:

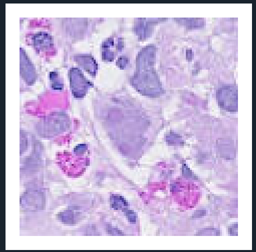
Neck Mass, Right (Excision):

Classical Hodgkin Lymphoma.

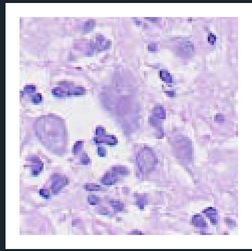
See microscopic description and comment.



```
+++++++++++++++
DX code:
- Anaplastic Large cell Lymphoma->0
- Classical Hodgkin lymphoma    ->1
***All unknown images are displayed with predicted DX:
+++++++++++++++
```



```
DX: Unknown; Predicted DX: 0
+++++++++++++++
```



```
DX: Unknown; Predicted DX: 0
+++++++++++++++
```

SUMMARY

- These 5 teaching cases were shown to residents on hemepath rotation to show how useful AI tool could help with diagnosis and the QA process.

- To our best knowledge, this study presents the first attempt to use a Vision Transformer model to provide preliminary teaching materials for pathology residents on hematopathology rotation in practical application of AI for screening to rule out cHL in clinical practice using WSI.

- This AI application may be useful as a QA tool to prevent misdiagnosis if used on a routine basis. Note that this Vision Transformer model is currently used for teaching only; no effort to revise cases with discrepant diagnosis after review is made.

DISCUSSION

- ViT serves as useful QA tool for further review of difficult cases. Note that the final diagnosis is still established by the expert hematopathologists with further testing (immunohistochemical stains or molecular testing)

- Residents with interests are encouraged to use the same tool for more cases (retrospective or prospective) to learn AI application in pathology QA initiatives. The AI software is available in the PC in DPALM Digital Pathology Lab

- The current module is still rudimentary and time-consuming to predict diagnosis of individual case. A plan has been in place to refine this module, with better interface, and to streamline the prediction process to make it more practical and simplified to use as a QA module on a daily basis.

Acknowledgement
(Projects on Deep Learning in Pathology)

- Faculty at U of Texas-Houston, Medical Scholl, Pathology:
  L Chen, MD
  Z Hu, MD
  H El Achi, MD
  A Wahed, MD
  I Wang, MD
  B Mai, MD
  J Armstrong, MD

- Faculty at U of Texas-Houston, Medical Scholl, Oncology:
  A Rios, MD
  Z Kanaan, MD

- Residents at U of Texas-Houston, Medical Scholl, Pathology:
  T Belousova, MD
  D Rivera, MD
  Y Gao, MD

- Staff in Digital Pathology Laboratory, U of Texas-Houston, Medical Scholl, Pathology:
  K Ali, BS
  R Zhang, MD

A Snow Day in Houston, Jan 2025