AI Breakthrough in Computational Pathology

AN ENABLER IN PERSONALISED CANCER TREATMENT

MUHAMMAD SHABAN

POSTDOCTORAL RESEARCH FELLOW

BRIGHAM AND WOMEN'S HOSPITAL, HARVARD MEDICAL SCHOOL

Outline

- Introduction to Computational Pathology (CP)
- Whole Slide Images
- Application of AI in CP
- Importance of AI in CP
- AI based CP Workflow
- Top Tier Publications
- Challenges
- State CP in Pakistan
- Two case studies
- Conclusion

Definitions

Pathology

- The science of the causes and effects of diseases
- Especially, the branch of medicine that deals with the laboratory examination of samples of body tissue for diagnostic or forensic purposes

Histopathology

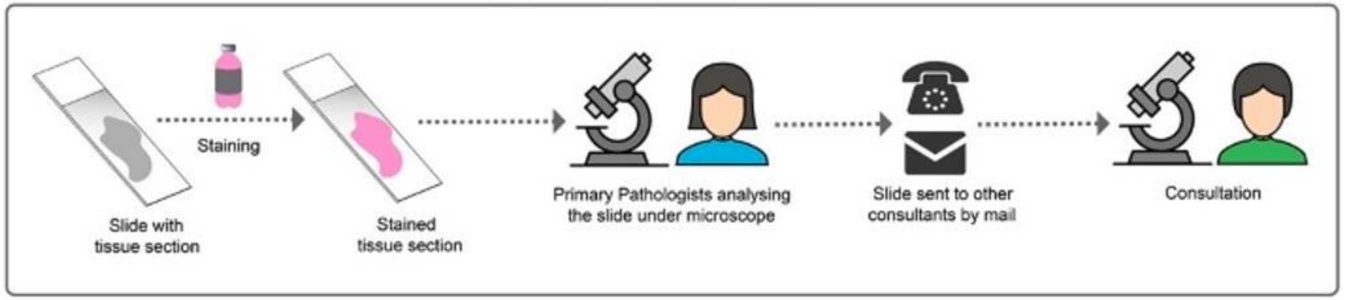
- Sub-branch of pathology
- It is concerned with the investigation of disease by examining cells and tissues

Pathologist/Histopathologiest

- A scientist/doctor who studies the causes and effects of diseases

Glass Slides

- A glass slide is a thin, flat, rectangular piece of glass that is used as a platform for microscopic specimen observation.



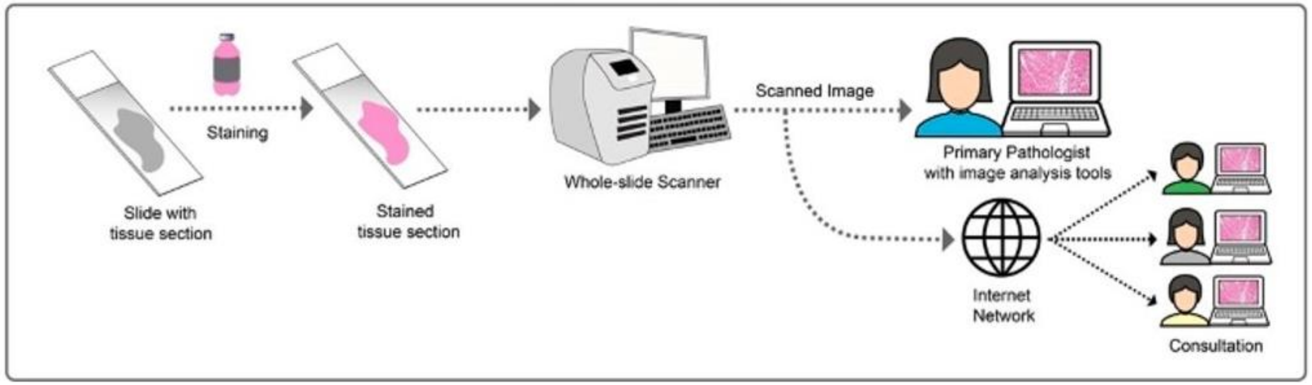
Definitions

Digital Pathology

- It incorporates the acquisition, management, sharing and interpretation of pathology information — including slides and data — in a digital environment.

Digital Slides/ Whole Slides Images

- Whole slide imaging, also known as virtual microscopy, refers to scanning a complete glass slide and creating a single high-resolution digital file.



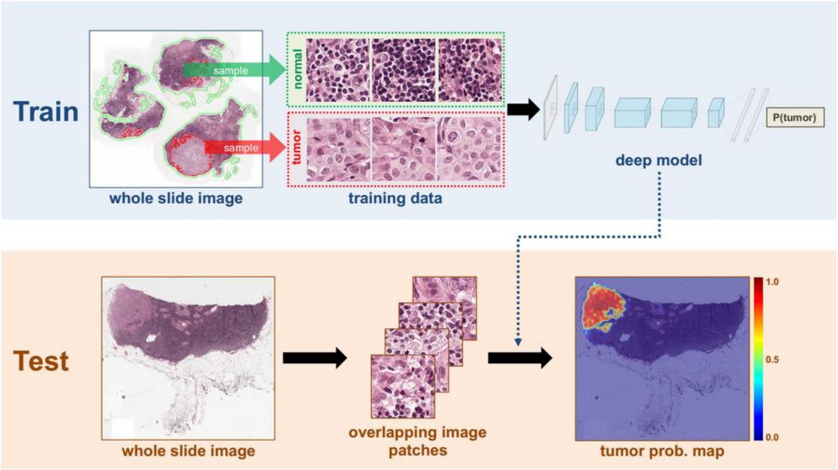
Definitions

Computational Pathology (CP)

- It is the analysis of digitized pathology images with associated metadata, typically using artificial intelligence (AI) methods.

Deep learning is CP

- It is a type of AI method, commonly used in computational pathology, that is able to "learn" how to perform tasks based on examples.



Motivation to work in Computational Pathology

Computational Pathology is an applied field of Computer Vision/Artificial Intelligence

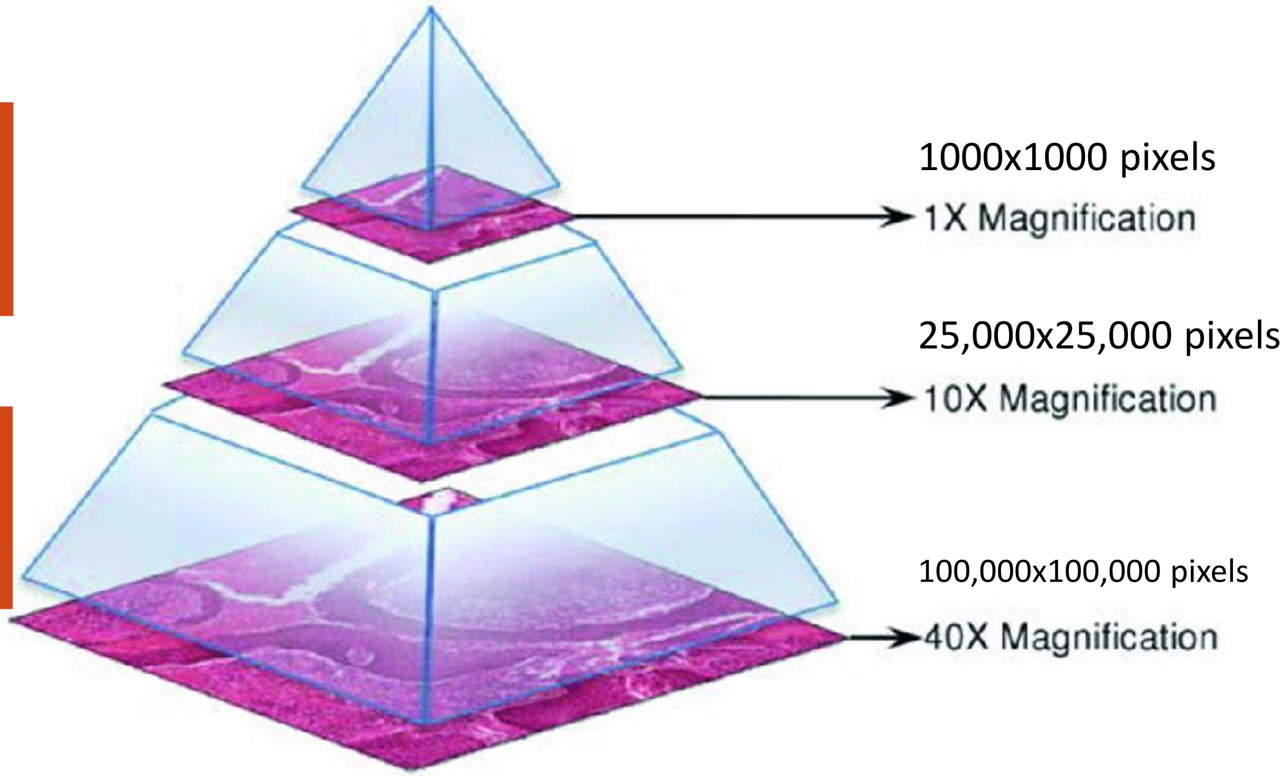
It is quite new as compared to other medical imaging based applied fields

In this filed competition is low therefore, probability to get a PhD position or Job in a top university/organization is high ☺

Whole Slide Images

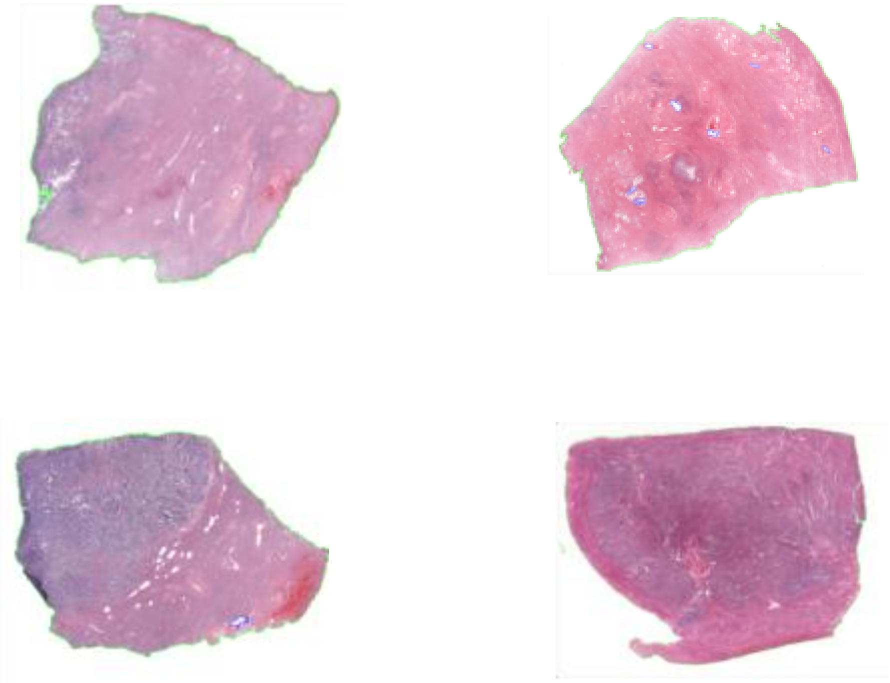
Takes 1-4 GB of space on disk when save in compressed format

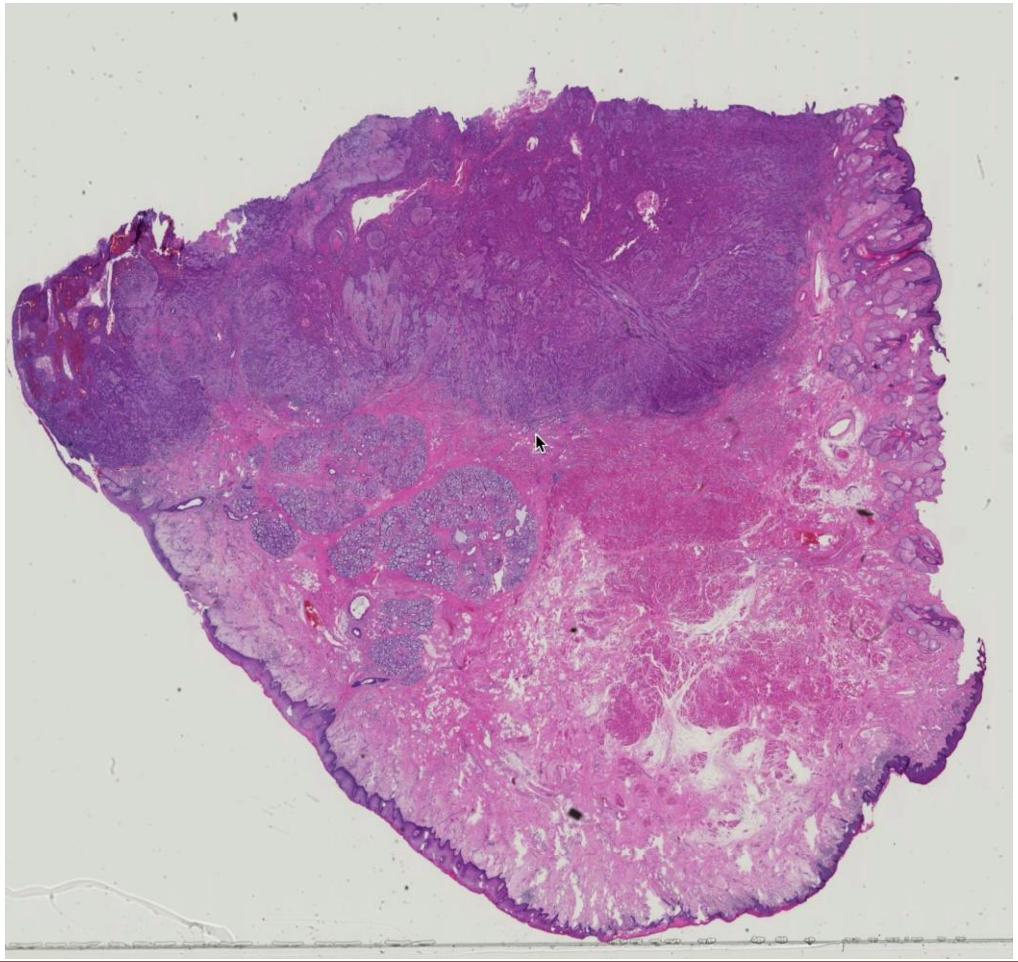
Don't try to load these images into memory at highest magnification



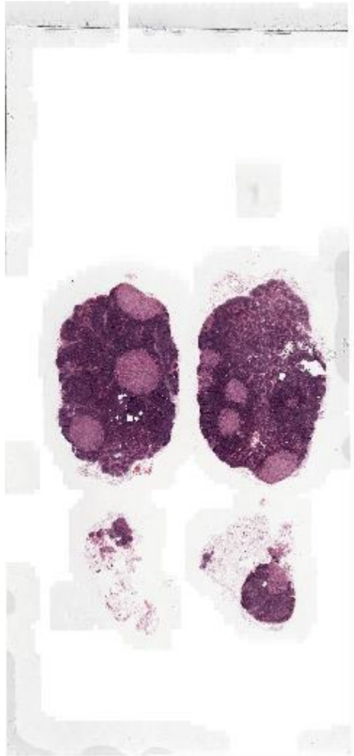
Whole Slide Images

Whole Slide Images

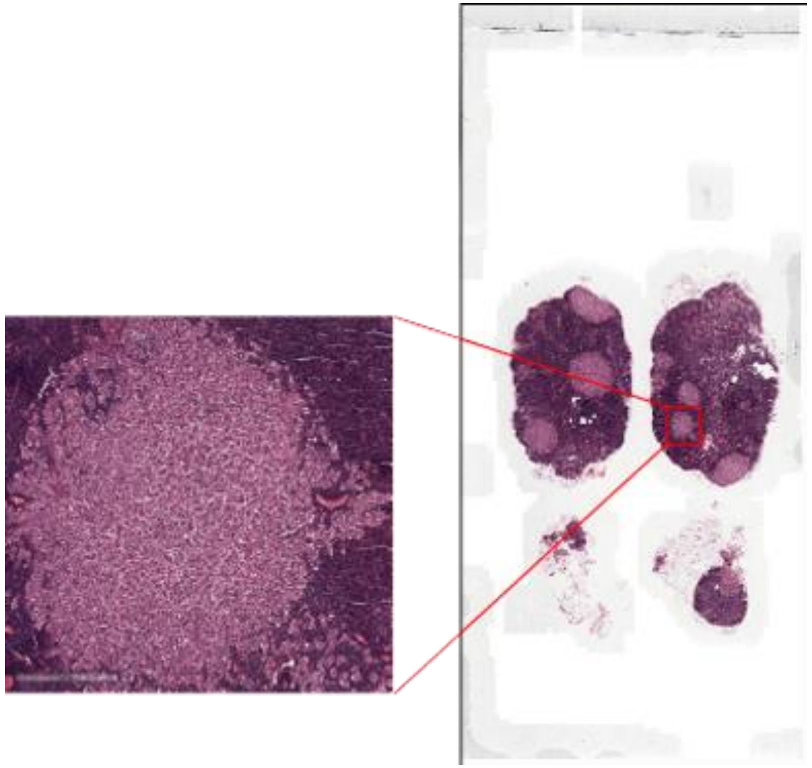




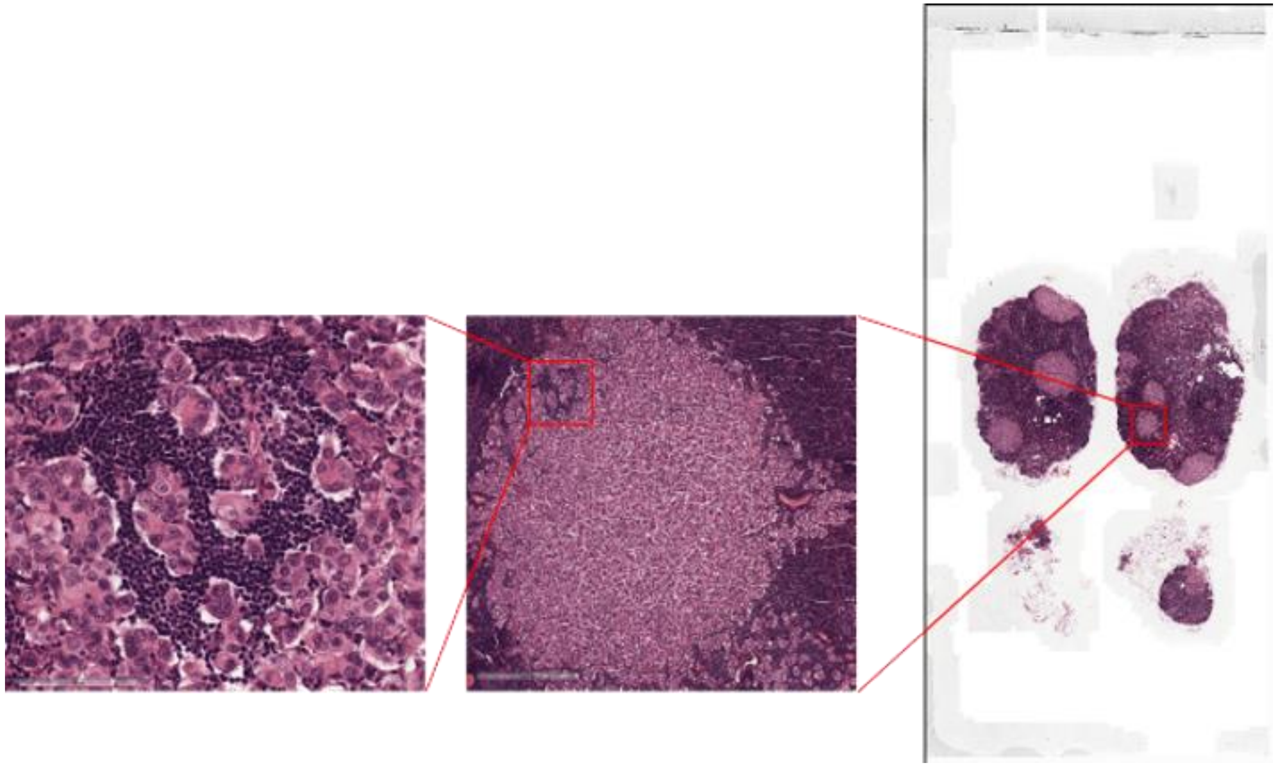
Whole Slide Images



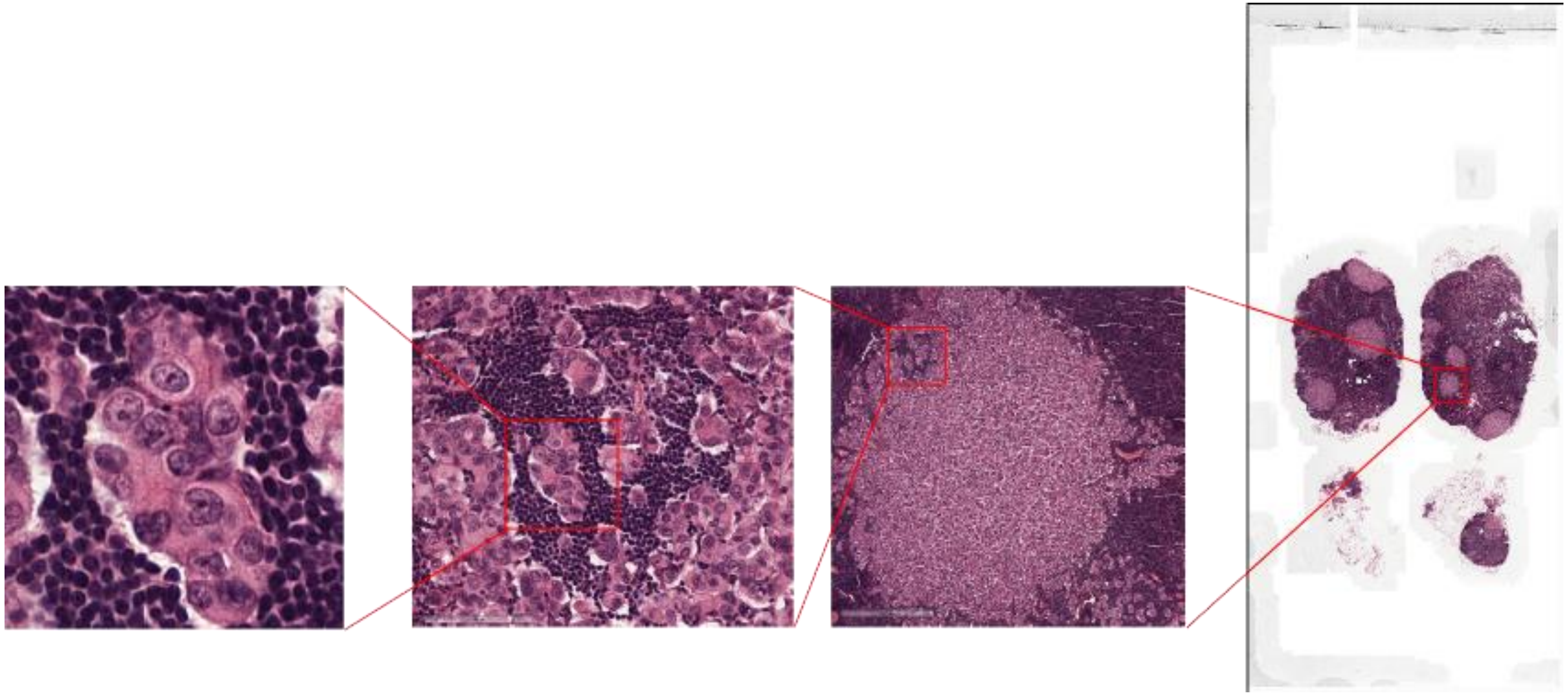
Whole Slide Images



Whole Slide Images



Whole Slide Images



Application of AI in CP

Cancer Diagnosis/Detection

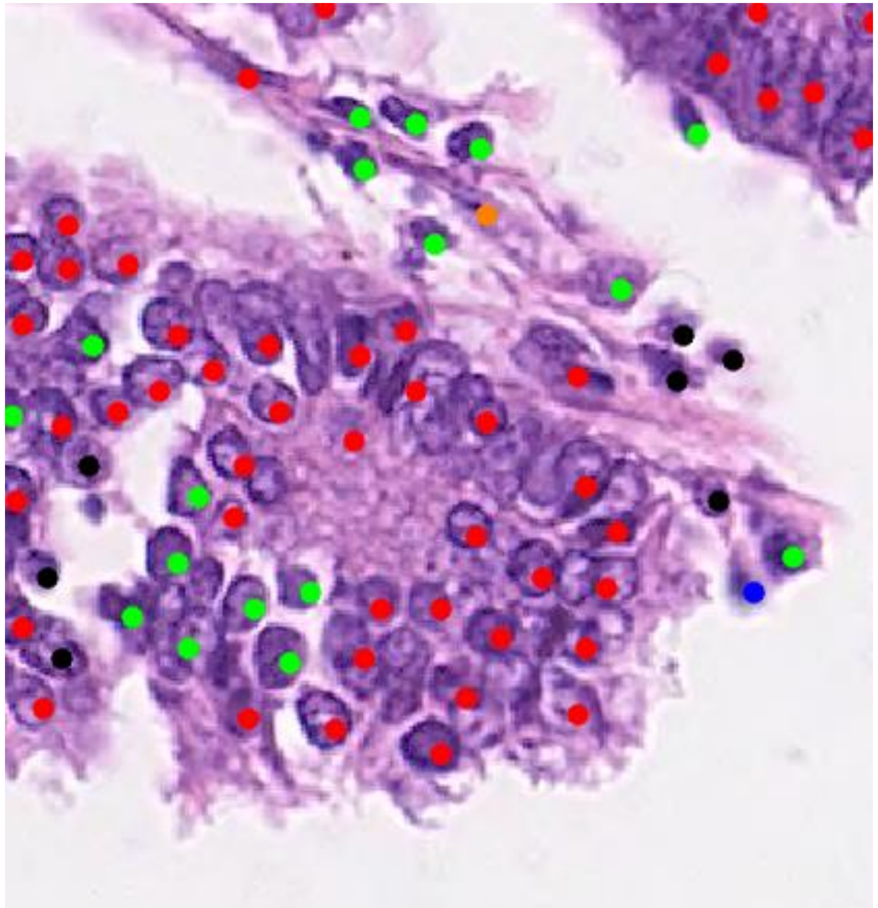
- A pathologist looks at the tissue under a microscope to see if the tissue has tumor cells or regions

Application of AI in CP

Cancer Diagnosis/Detection

- A pathologist looks at the tissue under a microscope to see if the tissue has tumor cells or regions

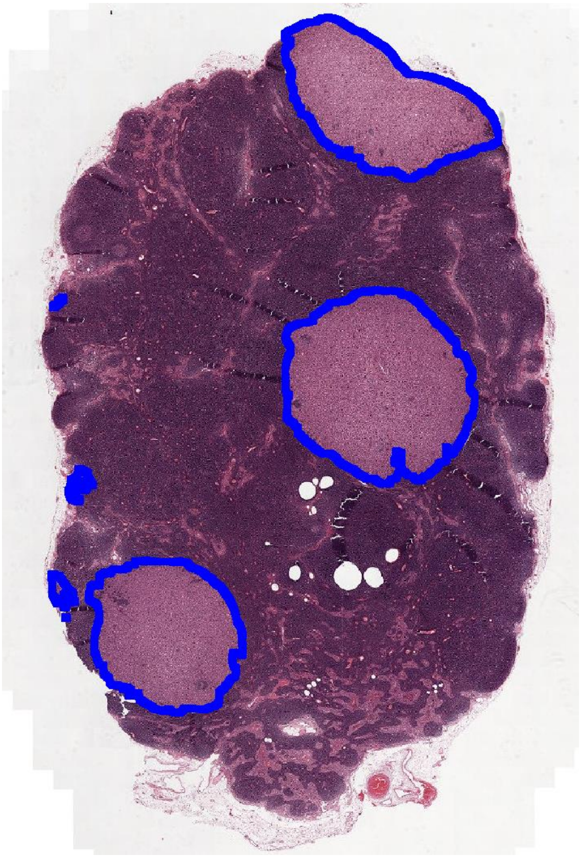
- Cancer Diagnosis is basically an **objection detection and classification problem** where object of interest are tumor cells and normal cells



Application of AI in CP

Cancer Diagnosis/Detection

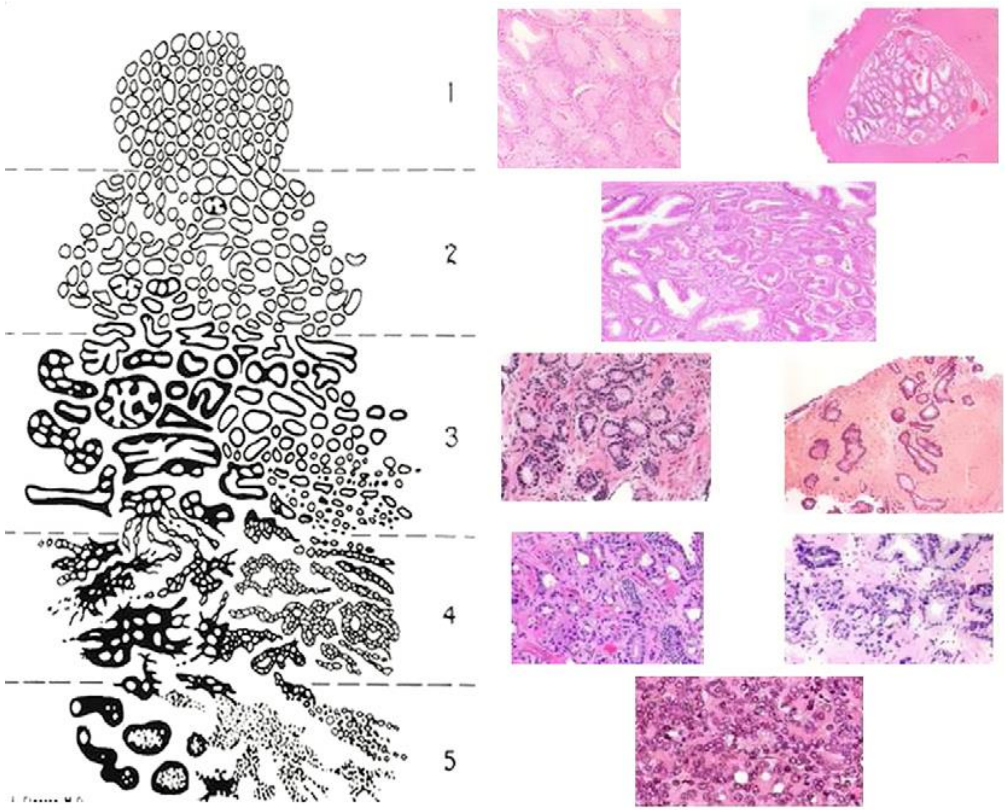
- A pathologist looks at the tissue under a microscope to see if the tissue has tumor cells or regions
- Cancer Diagnosis is basically an **objection detection and classification problem** where object of interest are tumor cells and normal cells
- Cancer detection can be considered as **segmentation problem** at WSI level



Application of AI in CP

Cancer Grading

- A pathologist looks at the tissue under a microscope and grade how different the cancer cells look compared to normal, healthy cells

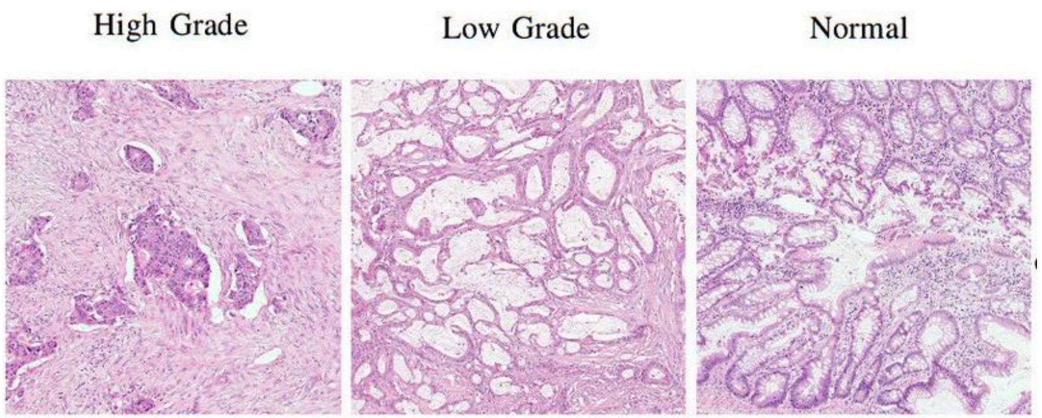


Application of AI in CP

Cancer Grading

- A pathologist looks at the tissue under a microscope and grade how different the cancer cells look compared to normal, healthy cells.

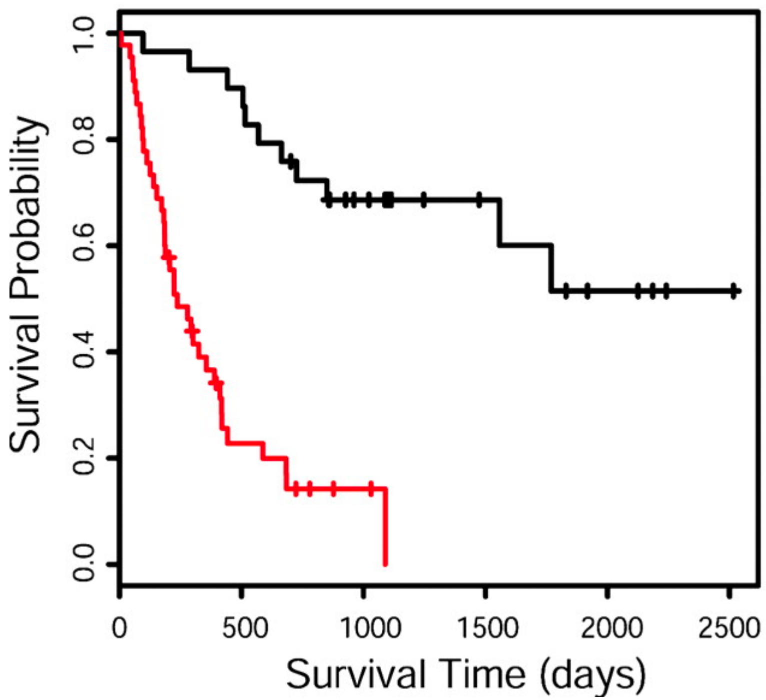
- Cancer grading is an **image classification problem** where the image should be reasonably large to capture region of interests.



Application of AI in CP

Survival Analysis

- A statistician analyze the clinical and pathological parameters of a patients and develop a mathematical model to predict the probability of survival for a certain period of time

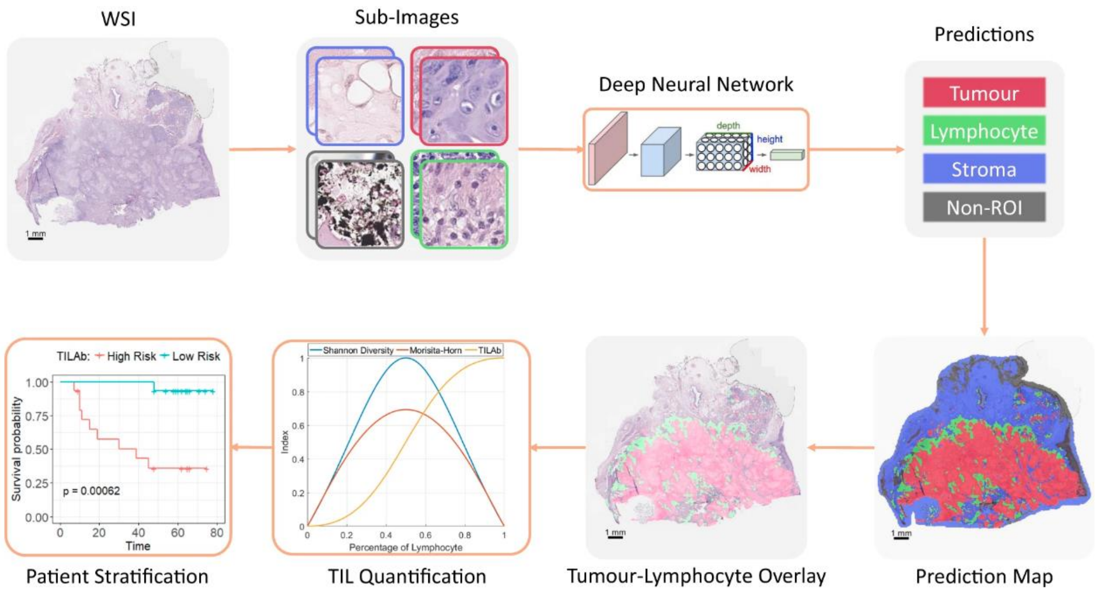


Application of AI in CP

Survival Analysis

- A statistician analyze the clinical and pathological parameters of a patients and develop a mathematical model to predict the probability of survival for a certain period of time

- Instead of relying only on clinical and pathological parameters we can use AI based methods to find novel feature which related to patient survival



Importance of AI in CP

Workload

- Cancer detection in a WSI requires significant amount of time (4-5 minutes)
- There could be multiple WSIs for a single case
- Number of cancer cases is increasing globally, and the number pathologist is decreasing in the US
- AI based cancer detection can reduce pathologist workload

Consistency in Detection

- Two different pathologist may assign different diagnosis label to same case (inter observer variability)
- Same pathologist may assign different diagnosis label to same case at different time (intra observer variability)
- AI based methods will be deterministic and will assign same label to same case every time

Importance of AI in CP

Accurate Prediction

- AI based methods are more accurate in some CP tasks e.g., tumor segmentation, cell detection/counting

Discovering New Clinicopathological Relationships

- Pathologist usually evaluate each case based on predefined set of parameters e.g., tumor stage or grade
- AI has the potential to find novel patterns which may lead to precise treatment plan for each patient

AI based CP Workflow

Preprocessing

- Patch extraction from WSIs
- Stain/Color Normalization
- Data Augmentation

AI Model

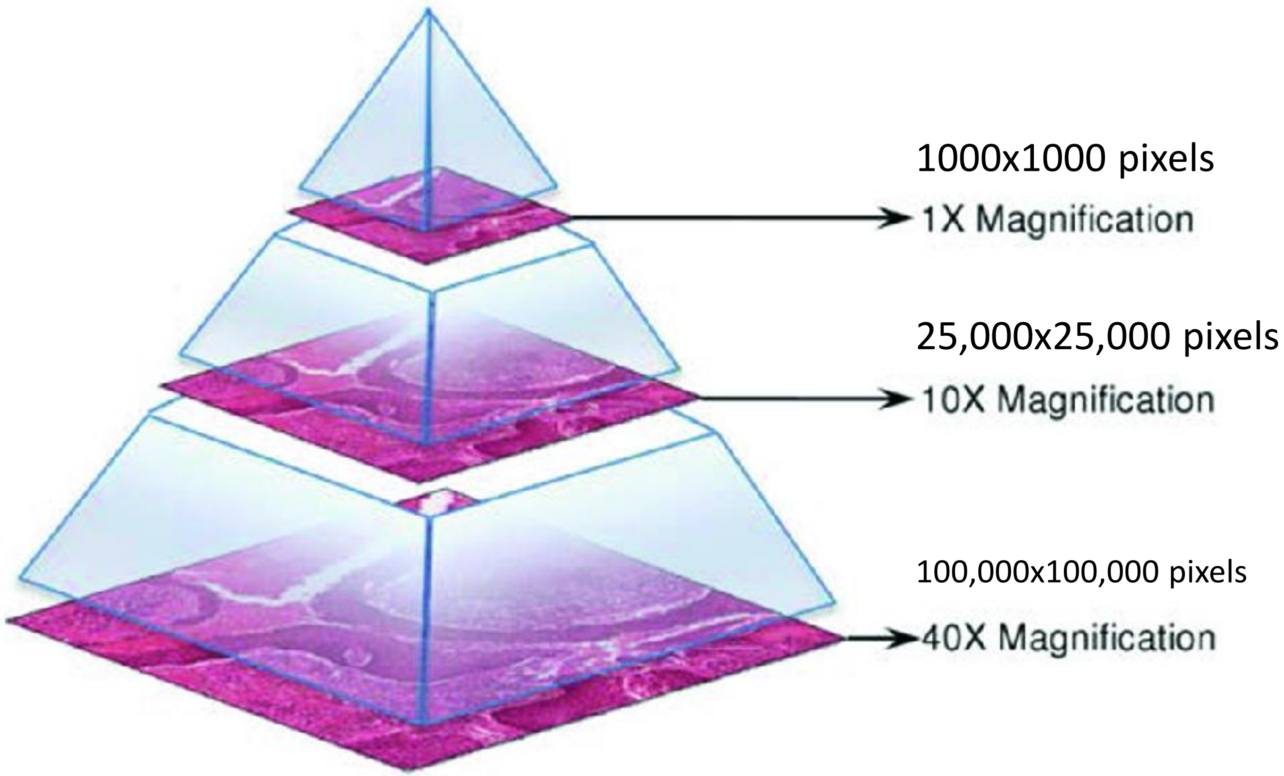
- Selection of Suitable Network Architecture

Post Processing

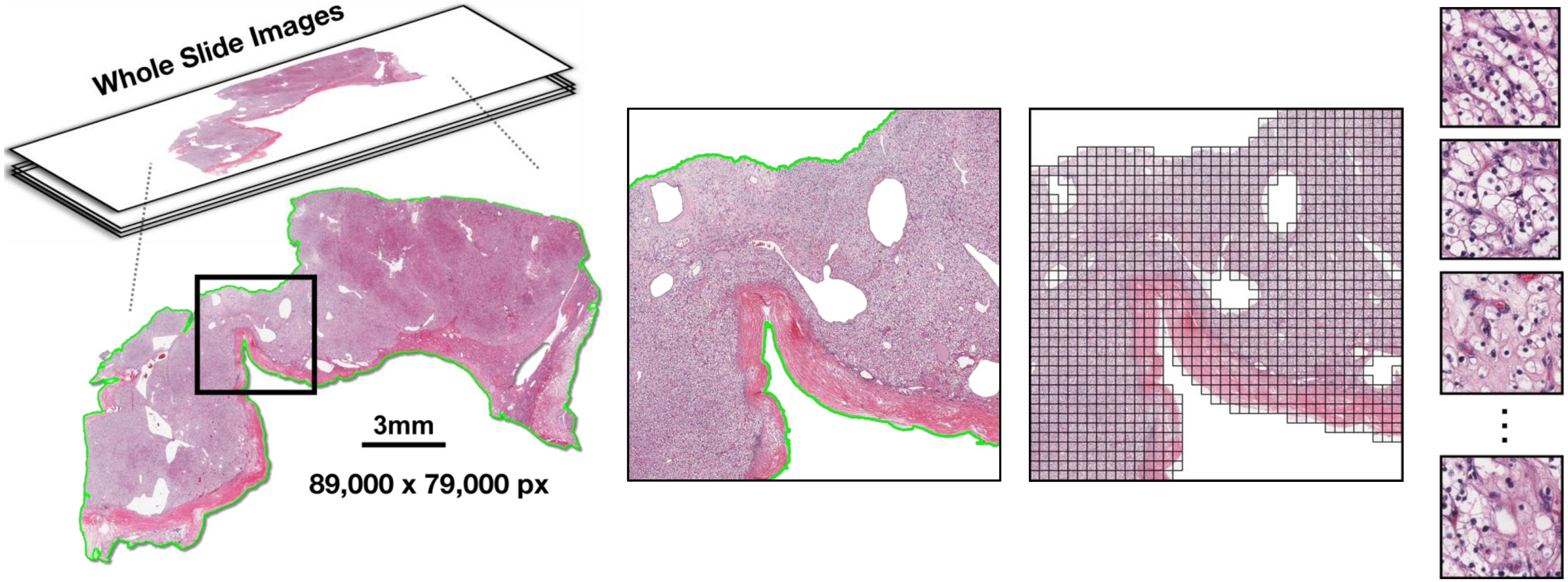
- Patch prediction intergration to generate WSI level prediction

Preprocessing - Patch extraction from WSIs

- WSIs are very large and cannot be directly used for training
- Need to select a magnification level to extact patches
- Model trained on one magnification and tested on another magnification usually shows inferior performance
- Need to select subset of patches from each WSIs as each WSI may contains tens of thousandas of patches



Preprocessing - Patch extraction from WSIs



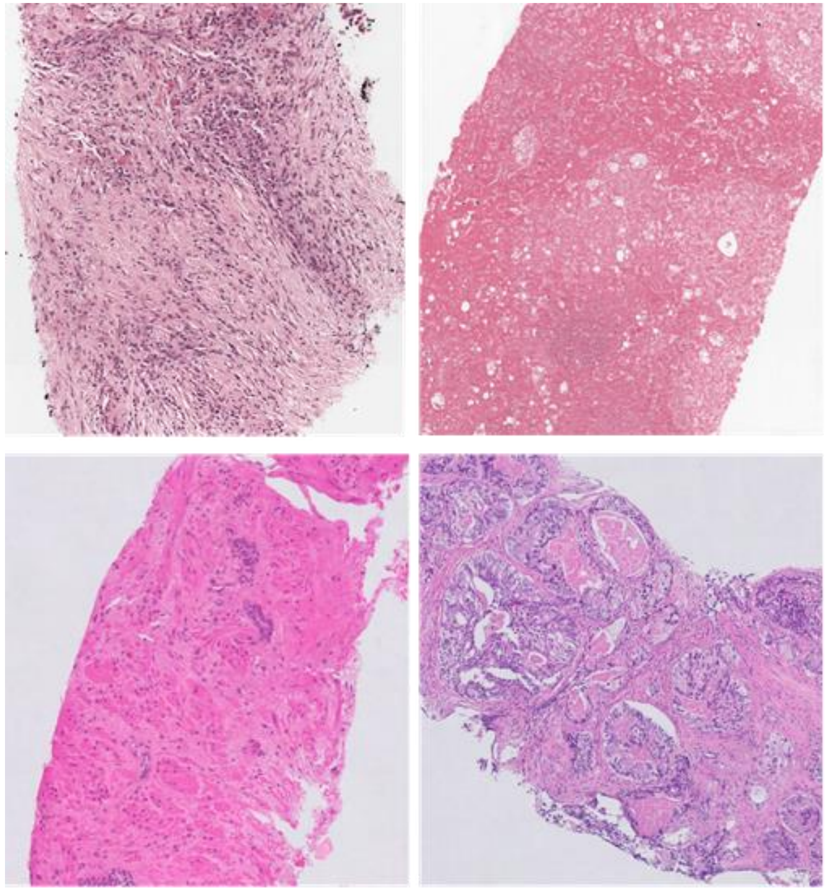
Preprocessing - Stain/Color Normalization

WSIs from different institutions/hospitals may have significant color variation due to difference in slide preparation or scanning protocols

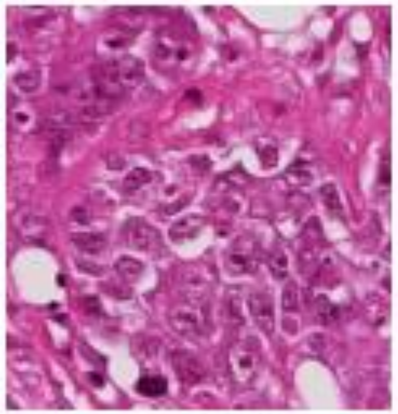
Color normalization helps in better performance during test time

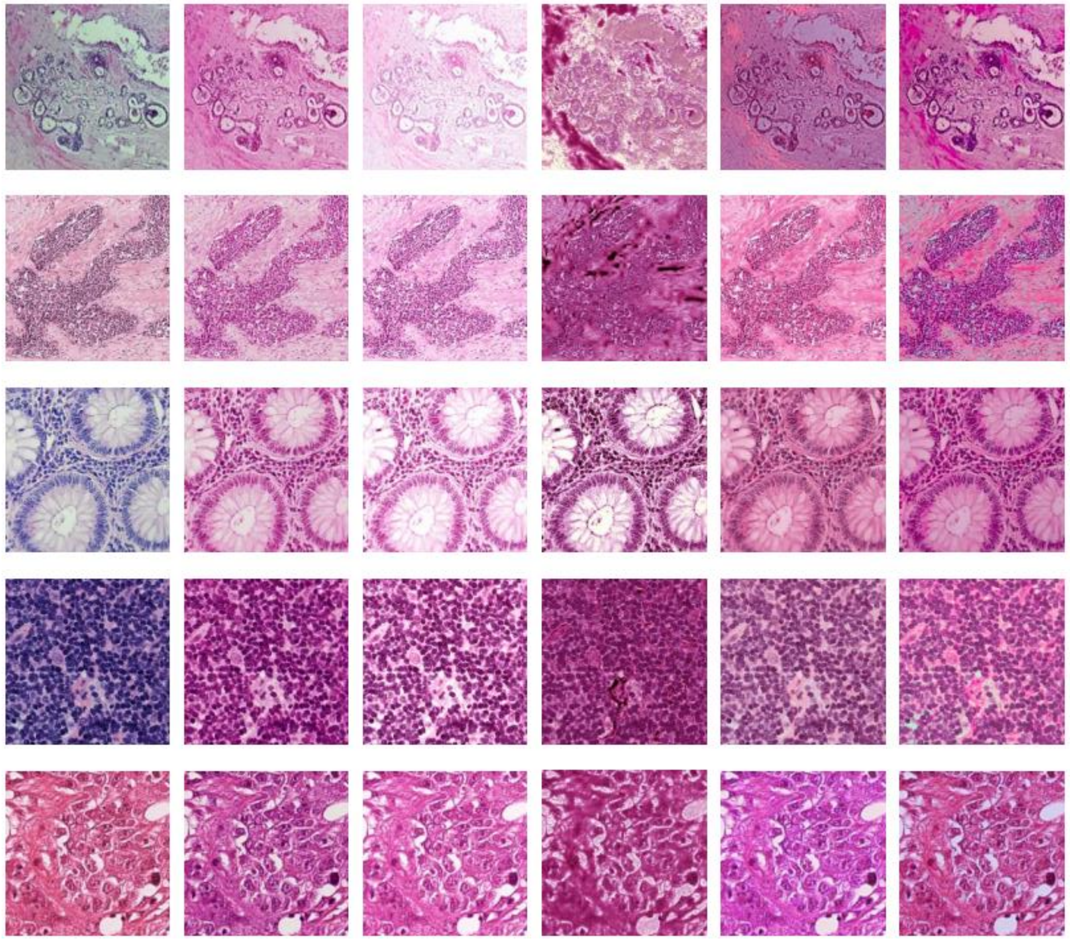
Model trained on color normalized data requirs color normalized images at test time

Random color augmentation during training also helps in achiving robust model



Preprocessing - Stain/Color Normalization



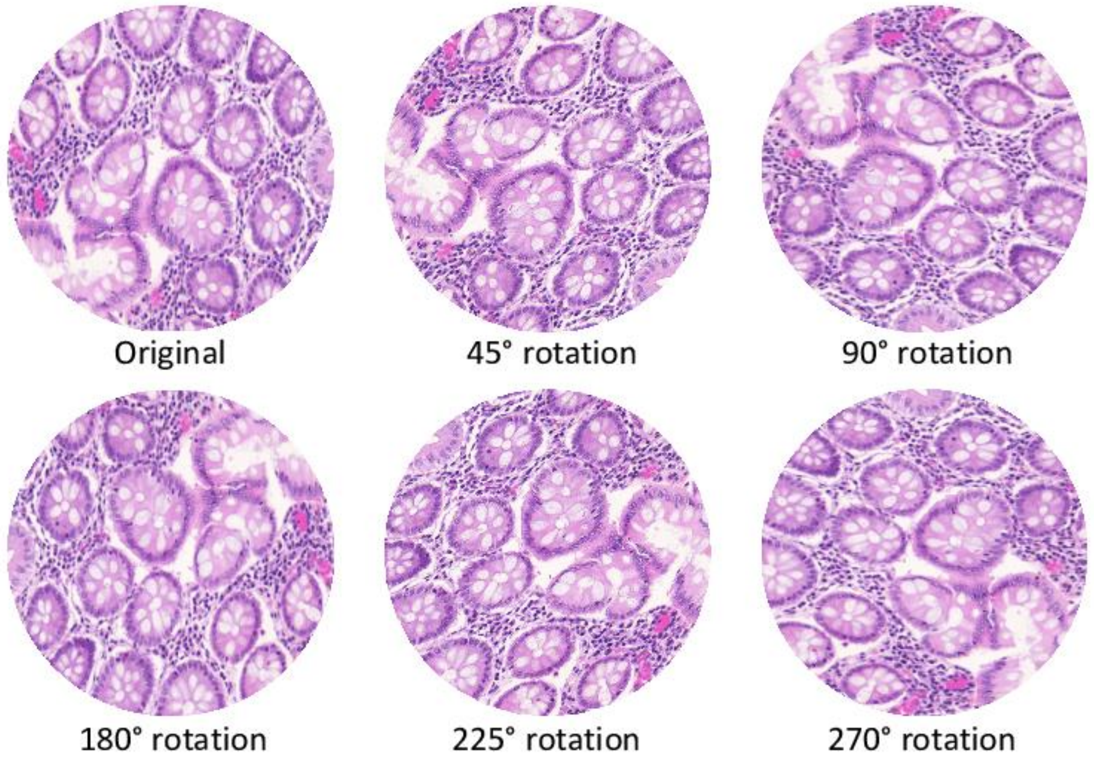


Preprocessing - Data Augmentation

A tissue in a WSI is placed without any consideration of its orientation

CP model should be invarint to ratation and flipping

Random rotation and flipping during model training is important to achive more robust model

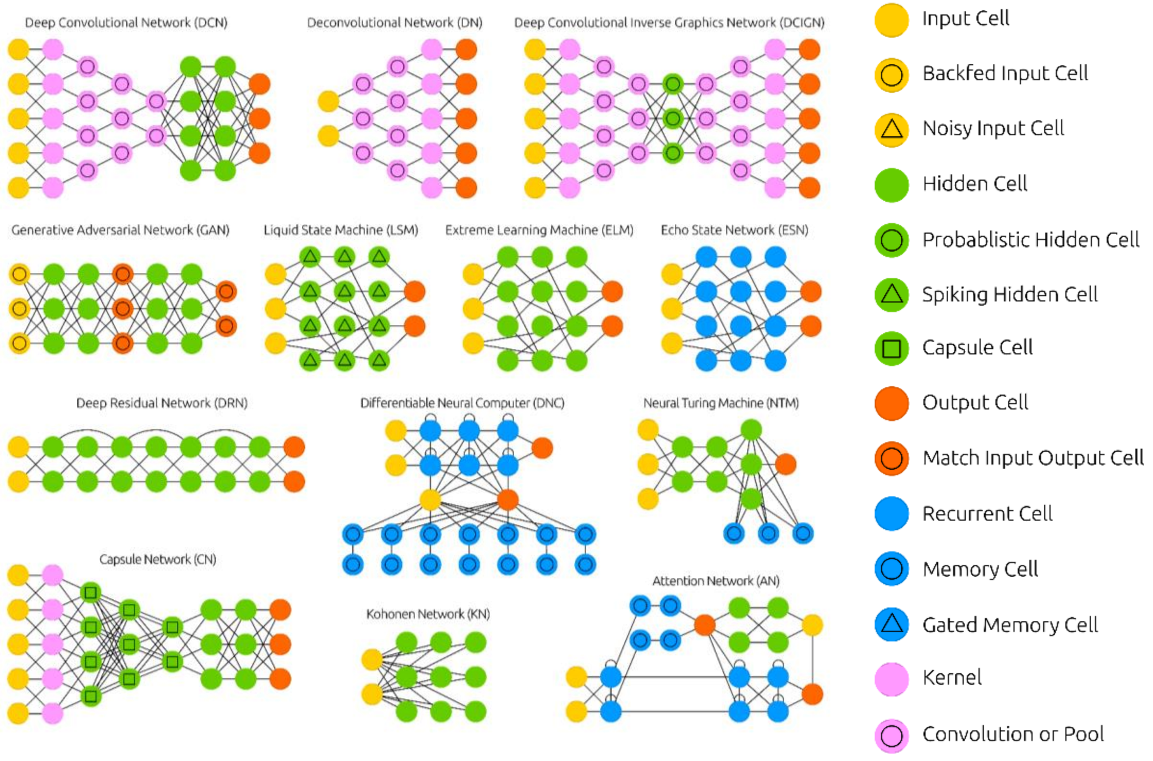


AI Model

Different CP problems require different types of network architecture

Standard networks proposed for natural images (Resnet50, Inception-v3, Deeplab, etc) work well for common CP problems e.g., cancer classification or segmentation

Some CP problems require technical novelity to incorporate domain knowledge to get better performance

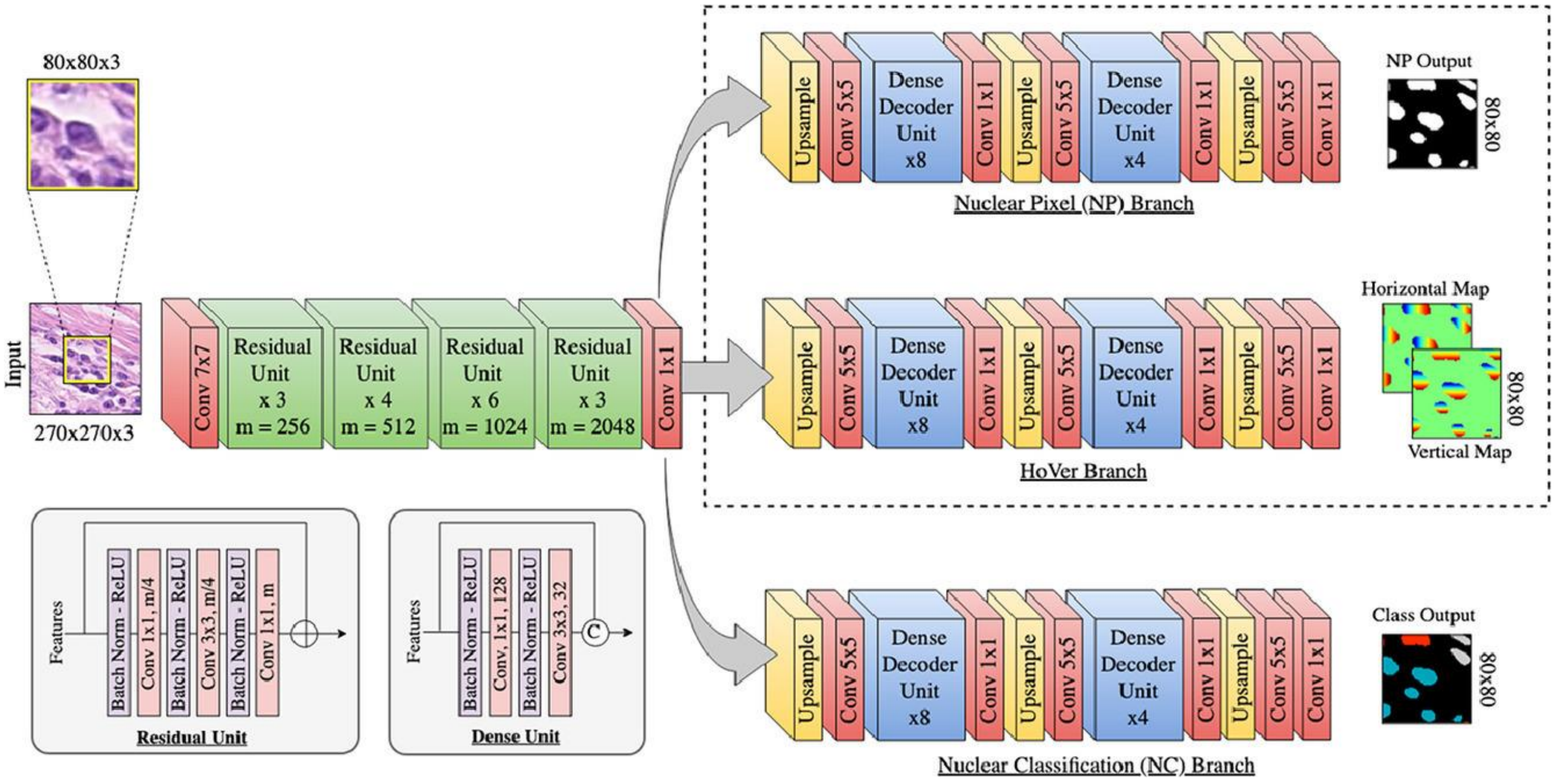


AI Model

Convolutional Neural Network	Classification and Segmentation
Auto-Encoder	Representation Learning Unsupervised Learning
Generative Adversarial Networks (GANs)	Synthetic dataset creation Domain adoption
Long Short-Term Memory (LSTM) Networks	Survival analysis
Weekly Supervised Learning	WSI level classification

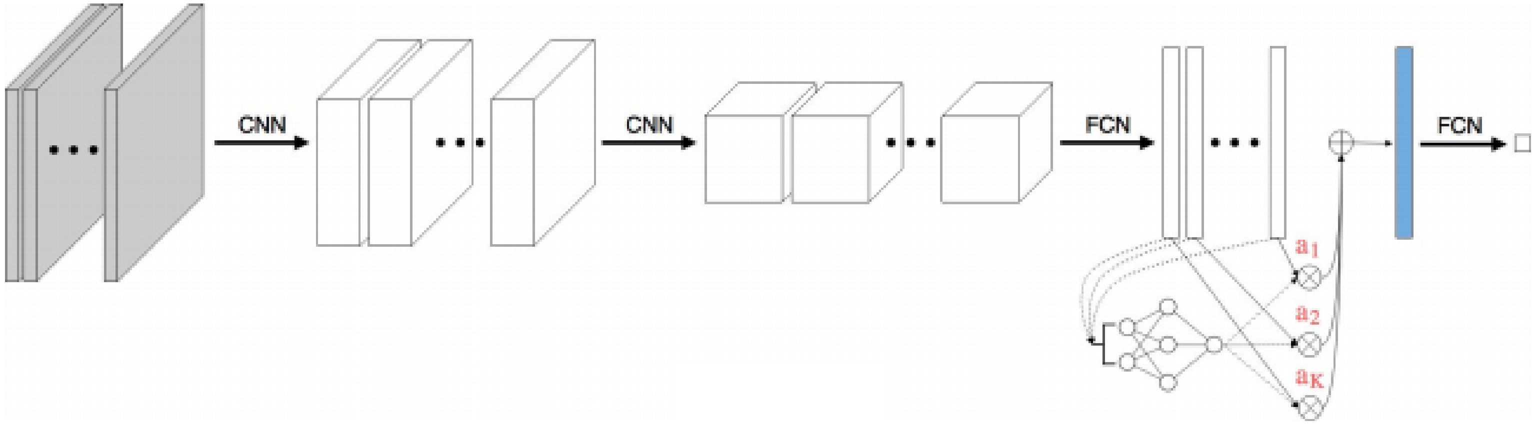
AI Model – Cell Segmentation and Classification

HoverNet: Simultaneous segmentation and classification of nuclei in multi-tissue histology images



AI Model – WSI Level Classification

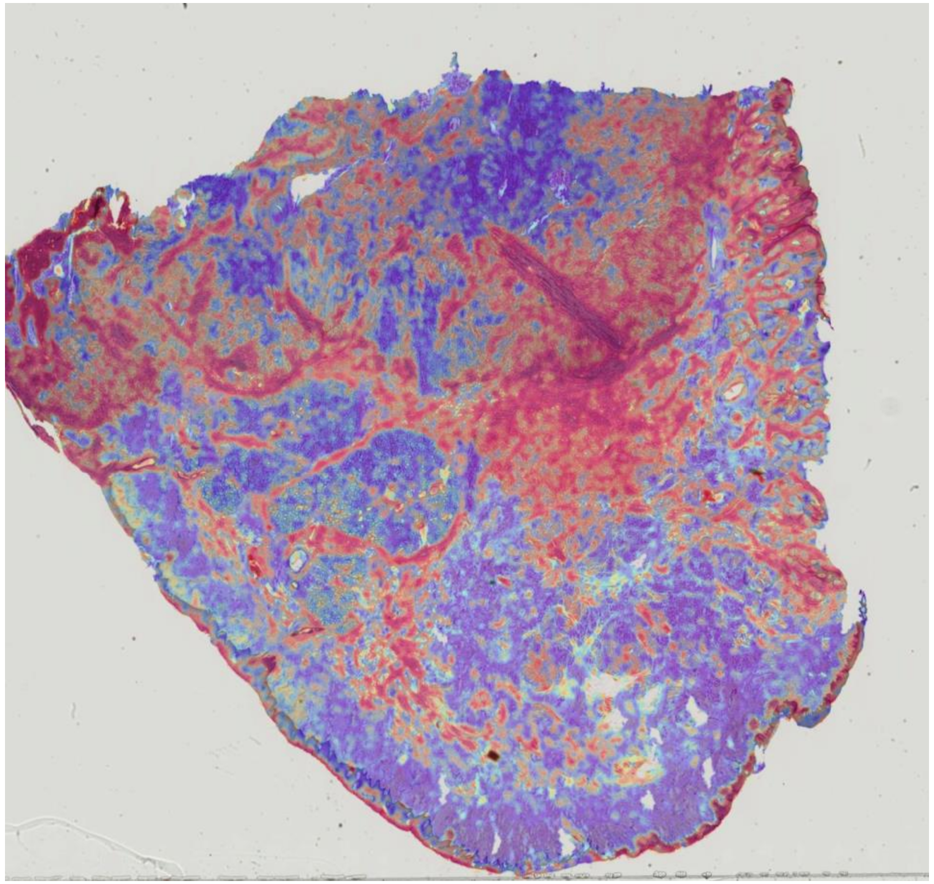
Attention-based Deep Multiple Instance Learning



Post Processing

Require stitching of patch level prediction into a WSI level prediction map

If CP problem requires a WSI level label then some integration strategy will be required to accumulate patch level prediction



Top Ranked Publication in CP

Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer

- Journal of the American Medical Association
- Impact Factor: 56.3

AI-based pathology predicts origins for cancers of unknown primary

- Journal: Nature
- Impact Factor: 54.6

Geospatial immune variability illuminates differential evolution of lung adenocarcinoma

- Journal: Nature Medicine
- Impact Factore: 49.2

Data-efficient and weakly supervised computational pathology on whole-slide images

- Journal: Nature Biomedical Engineering
- Impact Factor: 26.355

Top Computer Vision Jounal

- IEEE Transactions on Pattern Analysis and Machine Intelligence
- Impact Factor: 16.4

Challenges

- Availability of Imaging Datasets
- Ethical Approval
- Limited Ground Truth Labels/ Annotations
- Disk Storage
- Computational Resources

Challenges – Availability of Imaging Datasets

- There are many publicaly available histology datasets
- Dateset of rare type of cancers are quite small
- Datasets with balanced number of samples are not readily available
- Curation of clean and balanced dataset is challenging and expensive

Tissue	*n*	Tissue	*n*
Breast	1092	Ovary	376
Lung	1016	Liver	371
Kidney	885	Cervix	304
Brain	677	Soft tissue	259
Colorectal	623	Adrenal gland	258
Uterus	611	Pancreas	177
Thyroid	502	Esophagus	164
Head and Neck	501	Bone marrow	151
Prostate	495	Eye	80
Skin	468	Lymph nodes	48
Bladder	408	Bile duct	36
Stomach	380		

Challenges – Ethical Approval

Histology datasets consist of human tissues

One can only use these datasets after taking the **informed consent** of patients

Datasets should not contain any information which can be used to disclose the patient's identy

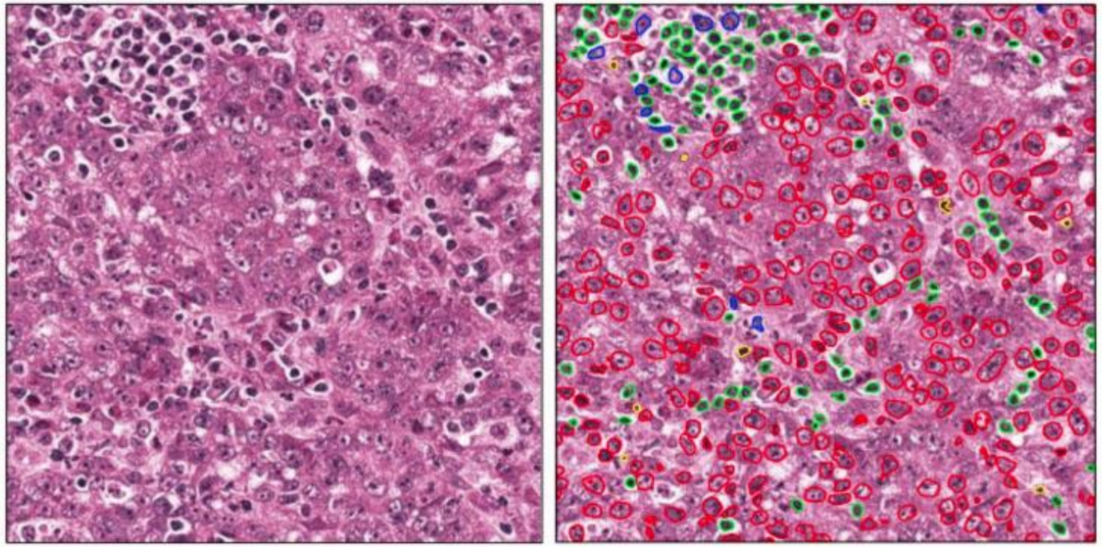
Challenges – Limited Ground Truth

Unlike natural images, only trained pathologist can mark ground truth (cancer) in histology images

Pixel/Cell level ground truth marking is the most labourios task as there are hundreds of thousands of pixels/cells in each WSI

There is known inter-oberserver varibility issue in ground truth marking

WSI level groud truth is the most readly available form of ground truth

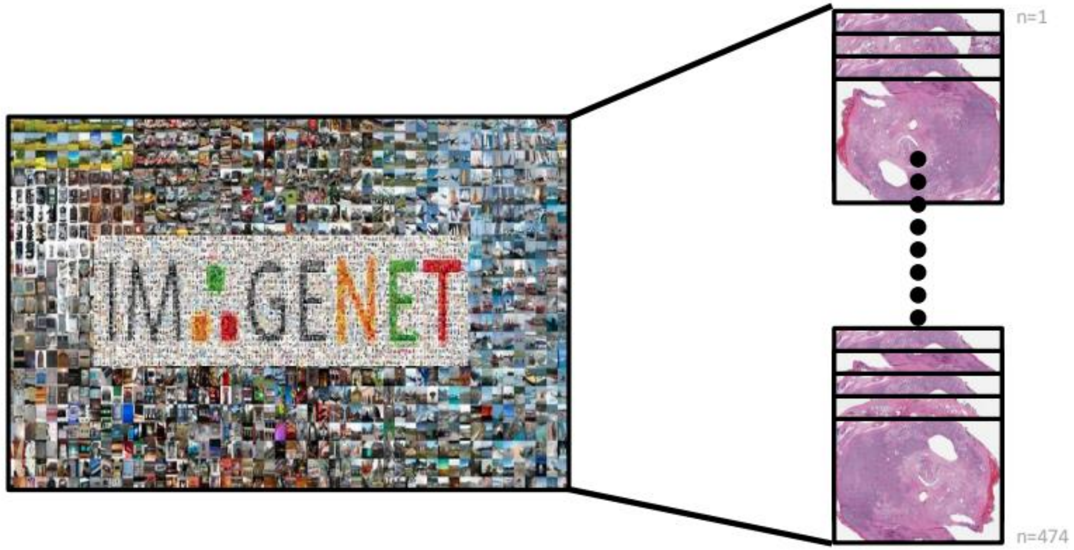


Challenges – Disk Storage

Storing a WSIs based dataset on a local machines required huge amount of disk storage

The disk storage required for ImageNet datest is less the storage required for 500 WSIs

One of the largest WSIs public dataset (TCGA) required **1.2 Petabyte** disk storage



All of ImageNet
482 x 415 * 14,197,122
= 2.8 trillion pixels

474 Whole Slides
100,000 x 60,000 *474
= 2.8 trillion pixels

Challenges – Computational Resources

AI based CP model also require huge computational resources both for model training and inference

Big AI based pathology labs and companys use super comptuers (Nvidia DGX2 or A100) for CP model training and inference

Small CP usually used pretrained features for model training of very large datasets

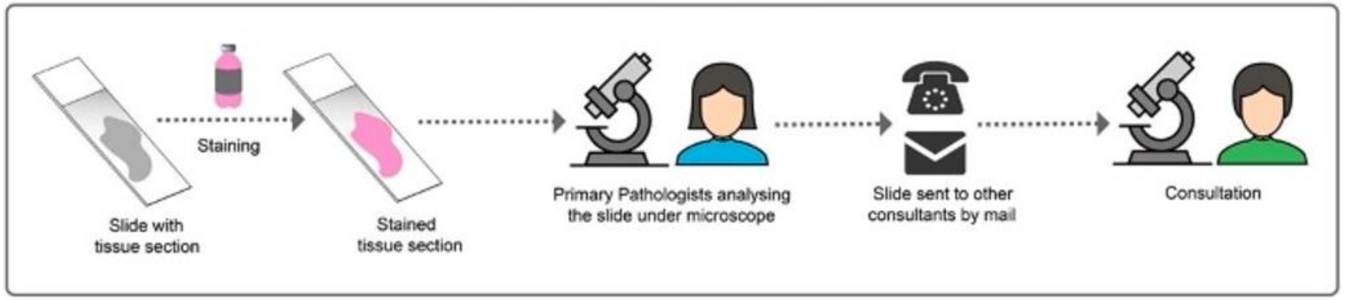
SYSTEM SPECIFICATIONS

GPUs	16X NVIDIA® Tesla® V100
GPU Memory	512GB total
Performance	2 petaFLOPS
NVIDIA CUDA® Cores	81920
NVIDIA Tensor Cores	10240
NVSwitches	12
Maximum Power Usage	10kW
CPU	Dual Intel Xeon Platinum 8168, 2.7 GHz, 24-cores
System Memory	1.5TB
Network	8X 100Gb/sec Infiniband/100GigE Dual 10/25/40/50/100GbE
Storage	OS: 2X 960GB NVME SSDs Internal Storage: 30TB (8X 3.84TB) NVME SSDs

State of CP in Pakistan

Almost all cancer hospital in Pakistan still follow the traditional pathway for cancer diagnosis

There is only whole slide image scanner somewhen in Sindh



State of CP in Pakistan

Computer vision labs (in NUST, LUMS, ITU) are working on CP problems using publicly available datasets

Some pathologists are also quite enthusiastic about CP and willing to mark ground truth on publicly available datasets

State of CP in Pakistan

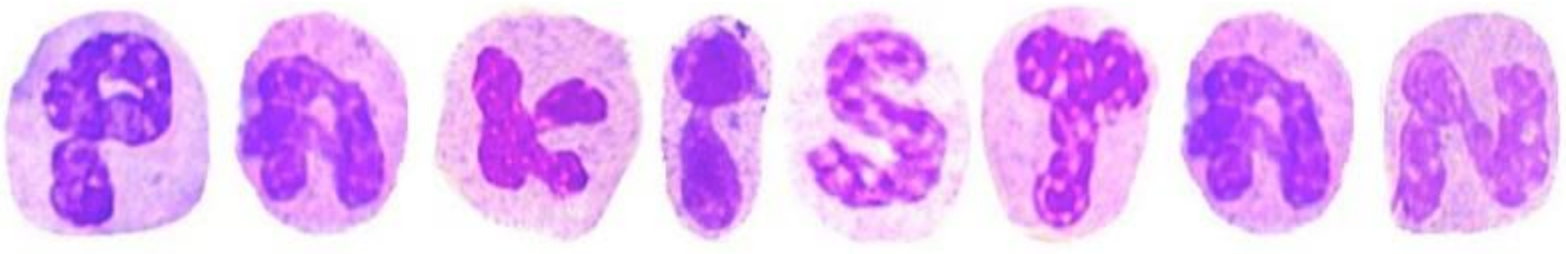
Computational resources in CV labs required for large scale CP studies are also not enough

This limitation can be considered as an opportunity to develop light-weight models for CP problems

International CP community has enough computational resources therefore they don't pay much attention to the development of light-weight methods for CP problems

State of CP in Pakistan

I have created a group (Digital Pathology - Pakistan) on LinkedIn to link pathologists and computer vision community to build a CP community in Pakistan



https://www.linkedin.com/groups/9059663/

Histopathology Datasets

Dataset or paper	Image size (px)	# images	Stain	Disease	Additional data	Potential usage
KIMIA960 [87,88]	308×168	960	H&E/IHC	various tissue		Disease classification
Bio-segmentation [89,90]	896×768, 768×512	58	H&E	Breast cancer		Disease classification
Bioimaging challenge 2015 [91,92]	2040×1536	269	H&E	Breast cancer		Disease classification
GlaS [23,93]	574–775×430–522	165	H&E	Colorectal cancer	Mask for gland area	Gland segmentation
BreakHis [15,94]	700×460	7909	H&E	Breast cancer		Disease classification
Jakob Nikolas et al. [88,95]	1000×1000	100	IHC	Colorectal cancer	Blood vessel count	Blood vessel detection
MITOS-ATYPIA-14 [96]	1539×1376, 1663×1485	4240	H&E	Breast cancer	Coordinates of mitosis with a confidence degree/six criteria to evaluate nuclear atypia	Mitosis detection, nuclear atypia classification

Histopathology Datasets

Kumar et al. [97,98]	1000×1000	30	H&E	Various cancer	Coordinates of annotated nuclear boundaries	Nuclear segmentation
MITOS 2012 [20,99]	2084×2084, 2252×2250	100	H&E	Breast cancer	Coordinates of mitosis	Mitosis detection
Janowczyk et al. [100,101]	1388×1040	374	H&E	Lymphoma	None	Disease classification
Janowczyk et al. [100,101]	2000×2000	311	H&E	Breast cancer	Coordinates of mitosis	Mitosis detection
Janowczyk et al. [100,101]	100×100	100	H&E	Breast cancer	Coordinates of lymphocyte	Lymphocyte detection
Janowczyk et al. [100,101]	1000×1000	42	H&E	Breast cancer	Mask for epithelium	Epithelium segmentation
Janowczyk et al. [100,101]	2000×2000	143	H&E	Breast cancer	Mask for nuclei	Nuclear segmentation
Janowczyk et al. [100,101]	775×522	85	H&E	Colorectal cancer	Mask for gland area	Gland segmentation
Janowczyk et al. [100,101]	50×50	277,524	H&E	Breast cancer	None	Tumor detection

https://www.sciencedirect.com/science/article/pii/S2001037017300867

Conclusion

CP is an emerging filed in medical image analysis domain

Working in CP does require development of technically novel methods

Knowledge of state-of-the-art AI methods is required to solve clinically important CP problems

Conclusion

AI based CP methods have potential to be published in journals with very high impact factor

High impact CP studies require huge computational, storage and human resources

Experience in AI based CP can help in getting a good PhD position/job

Thank You!