1. Computational Pathology

Computational Pathology is extraction of actionable knowledge from Digital Pathology datasets using different computer algorithms (usually AI/deep learning) in order to assist pathologists with precised diagnosis. A generic workflow is presented in following Figure-1.

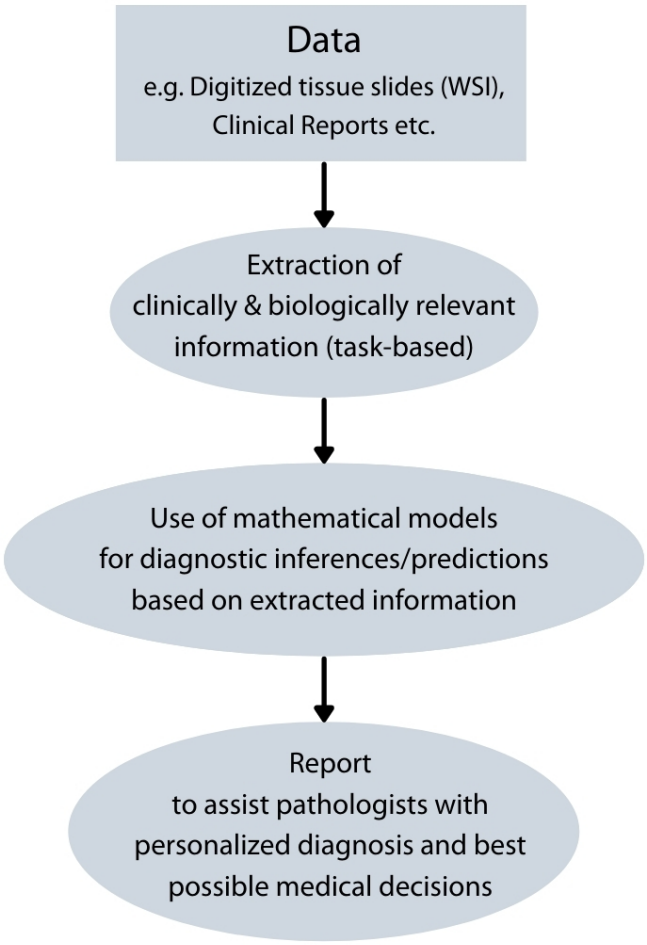

Figure 1: General Computational Pathology Workflow

2. Histology Images Preparation

Glass slides are scanned by a specialized scanner to create virtual histology slides. This digitization enables Telepathology and effective implementation of AI / Deep learning models in Digital Pathology for precision diagnosis. Histology slides are digitized in four steps [11]:

1. Image Acquisition using specialized slide scanners (e.g. PIPS)
2. Image Storage
3. Editing/Annotation
4. Display/Navigation/Management of virtual slides

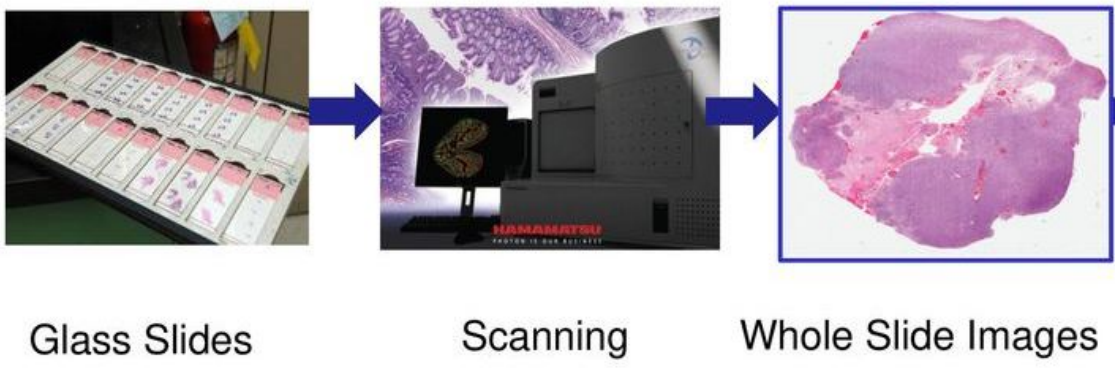

Figure 2: Virtual histology slides creation

3. Whole Slide Imaging and Virtual Magnification

Whole Slide Imaging is process of scanning conventional pathology glass slides in order to create their digital representations[11]. PIPS, one of the first whole slide scanner was approved by FDA in 2017, after ensuring that clinical interpretations made using virtual slides are comparatively same as those of made by using conventional microscopy [2].
Whole slide scanners use following scanning methods:

1. Tile-based Scanning: Squared patches of slide are focused, scanned, and then later virtually stitched seamlessly to generate highest-quality WSI.
2. Line-based Scanning: Scanning is done across a pre-selected axis to generate image strips.

A WSI multi-resolution pyramid is built with different magnification levels (40x, 20x, 10x), as specified by the task (Figure-3). Typically, seven to eight levels are enough [6]. Mostly, Resolution of WSI is determined by microns/pixel. Generally, the resolution of 0.5 mircon/pixel is considered sufficient for diagnosis. Digital magnification standards are following:

1. **10x**: 1 microns/pixel
2. **20x**: 0.5 microns/pixel
3. **40x**: 0.25 microns/pixel

WSI images are also referred as gigapixel images. A scanned area of 25mm x 50mm with 0.25microns/pixel resolution is a whole slide image of 20 gigapixels.

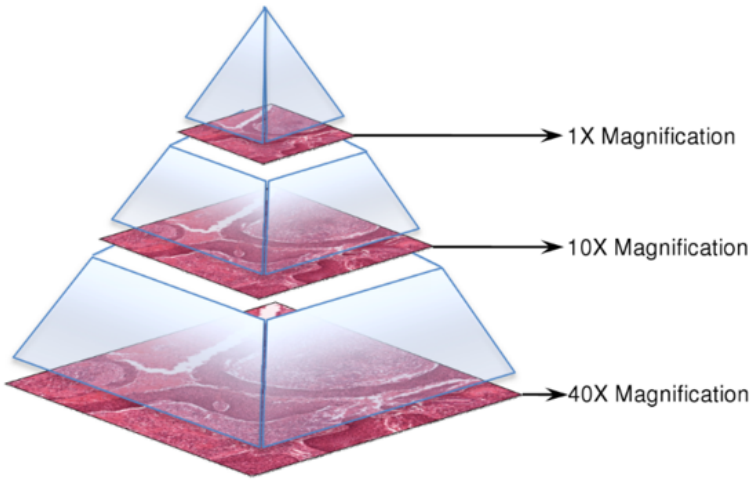

Figure 3: A general WSI Pyramid

4. Use of slide staining in Digital Pathology

Tissue samples on a glass slide are stained by different chemicals in order to highlight different tissue and cellular features, and enhance contrast among them. Staining also enables AI/Deep learning models for more accurate feature detection.
Hematocylin and Eosin(H&E) staining is one of the most preferred methods among pathologists. [3]. It stains cellular features as described in Table-1:

Some other special stains are also used in different tasks, to visualize elements unidentifiable by H&E staining. These methods are referred as IHC(Immunohistochemistry) methods as elaborated in Table-2.

H&E Staining	Color	Stain
H (Hematoxylin)	Purplish Blue	Cell Nucleus, Organelles containing RNA & DNA
E (Eosin)	Reddis/Pink	Cytoplasm, cell walls, extracellular fibers

Table 1: H&E Staining

IHC Stain Type	Histological Purpose
Masson Trichome	Differentiates Collagen & smooth muscle in tumors
Periodic Acid Schiff	Used to stain kidney & liver biopsies, stains areas with higher carbohydrate proportions
Perls' Prussian Blue Iron	Used to detect and identify abnormal amount of ferric iron in tissues, bone marrow.

Table 2: IHC Stains & their use cases

5. AI use cases in Computer Pathology

Most of the diagnosis tasks (detection & grading) in pathology require very precise and accurate analysis of tissue slides. AI & Deep learning has potential to overcome many of the workload challenges, improved detection and better prognosis by finding ROI from a malignant sample image [7]. The automated cancer detection algorithm for lymph nodes attained an accuracy of 92%, as compared to the accuracy of human pathologist as 96%. But, combination of both human pathologist and automated detection achieved an accuracy of 99% [1].

0.1 Use Cases:

1. **Metastasis Detection Classification**: Metastasis is measure of cancer spread from the primary site to other sites through blood or lymph nodes.

2. **Tumor Proliferation**: Tumor Proliferation is an important prognostic biomarker. It can be assessed through Mitosis detection & counting in WSI images [14]. End-to-end techniques for cancer prognostication are also being proposed [10].

3. **Nuclei Segmentation**: Tumor Proliferation is an important prognostic biomarker. It can be assessed through Mitosis detection & counting in WSI images [14]. End-to-end techniques for cancer grading through mitosis counting have also been proposed [10].

4. **Survival Prediction**: Whole slide images can also be used for cancer survival prediction by using MIL techniques as proposed by Jiawen Yao [16]. MIL techniques were also used for cancer staging [12]. Another method presented by Richard J. Chen is PatchGCN, a context-aware, spatially-resolved patch-based GCN that heiracichally

agregates instance-level histology features to model local and global topological structures in the tumor micro-environment.

5. Some of the use cases were suggested by Andrew Janowczyk as given in Table- [8].

6. Other use cases include:

   - Stain Normalization/ Computational Staining: Stain invariability across the laboratories is a challenge. So, stain normalization techniques (based on color deconvolution) reduces the color and intensity variations in stained images, that also impacts the prediction results significantly. [4]. Deep learning can also be used to computationally stain the images in an effective way. [13].
   - Effecient WSI processing learning: Ming Y. Lu proposed a high throughput WSI processing and learning method CLAM (Clustering-constrained attention MIL). This method uses attention-based learning to automatically identify sub-regions of high diagnostic value, and classifies the whole slide [9].
   - Virtual re-staining: H & E stained images can be re-stained virtually to IHC by using a GAN based technique [15].
   - Histology-Genomic Integration for better Prognosis: Most of the prognostic models use either Histology or Genomics alone. Richard J. addressed the integration of both modalities, and proposed a image-omic prognostic model using Multi-modal Deep learning [5].

Use Case	**Goal**
Segmentation	
Carcinoma Localization	Detection & Quantification of tumor for cancer grading
Nuclei Segmentation	Cancer grading, rate of proliferation
Ephithelium Segmentation	Identification of Tumor infiltrating lymphocytes
Tubule Segmentation	Area of tubule is clinically used for Breast Cancer grading
Detection	
Lymphocyte Detection	Linked to disease outcome, contributes in prognosis & survival prediction
Mitosis Detection	Cancer grading
Classification	
Lymphoma Typing	Diagnosis

Table 3: Deep Learning Use Cases

Table 4: Datasets

Datasets	**Details**
WSI Tissue Classification	
CRC-Tissue Phenotyping Dataset - link	Images: 280k patches from 20 WSIs Stain: H&E Two Settings: Fold1: 70-30 train-test split, no patient-level separation Fold2: Patient-level separation, 14 for train, 7 for test
WSI Segmentation (Metastases Detection)	
CAMELYON16 Dataset - link	Images: 400 WSIs of lymph nodes - tiff Stain: H&E Train: 170 (100 Normal, 70 with Metastases) 100 (60 Normal, 40 with Metastases) Test: 130 WSIs Annotations: XML, WSI bin masks
CAMELYON17 Dataset Challenge - DataLink	Images: 1000 WSIs from 100 patients -tiff Train: 500 WSIs (5 slides/patient) Test: 500 WSIs (5 slides/patient Stain: H&E Annotations: XML – (lesion level) Note: CAM16 data is also reused
BCSS(Breast Cancer Semantic Segmentation) Dataset - link	Images: 20,000 segmentations from tissue regions of breast cancer Resolution: 0.255MPP Annotation: JSON, masks are also provided

Table 4: Datasets

Datasets	**Details**
Glands Segmentation	
Warwick-QU - link	Disease: Colorectal Cancer Images: 165 – bmp format Resolution: 20x (0.625 microns/px) - Zeiss Stain: H&E Composition: 74 Benign, 91 Malignant
Nuclei Instance Segmentation	
MoNuSeg '18 Dataset - link	**Single-organ** Images: 30 (+7) images (30k nuclei annotations)– tiff format Magnification: 40x Stain: H&E Train: 22k nuclei annotation Test: 7k annotations
MoNuSec'20 Dataset - link	**Multi-organ** (4 different organs) Images: 31k nuclei annotations – tiff Magnification: 40x Stain: H&E
NuCLS Dataset - link	Images: 2,20,000 labeled nuclei FOV images (0.2MPP) Format: FOV images-png, Masks-png, Annotations-CSV Stain: H&E Single Rater: Nucleus Boundaries generated by image processing heuristics, refined by MaskRCNN & then corrected/uncorrected by Pathologist Multi Rater: Annotations were provided both by pathologists & Non-Pathologists (image processing heuristics)

Table 4: Datasets

Datasets	**Details**
Mitosis Detection	
MITOSTAPIA – link	Images(Frames): 1136 - tiff Resolution: 40X (0.244 microns/px) Stain: H&E Note: Same dataset can also be used for nuclei atypia scoring.
MIDOG '21 – link	Images: 200+ in tiff cropped ROIs of 200+ Individual tumor cases 1721 mitotic figures (MF) and 2714 hard examples Stain: H&E

6. Why Patch Extraction is necessary?

An entire gigapixel image can't fit in a GPU, neither it is efficient for training a neural network. So, a WSI is cropped into smaller patches at different resolution level and settings. A whole slide image typically contains multiresolution-pyramid for downsampling. Typically, 4 resolution levels are contained. Downsampling can be done by using a downsampling factor from 1,4,16 and 32.

The patch size varies according to the task specification. Substantial variance is present in different cellular figures i.e. nuclei. Image patches can be further compressed to smaller encodings. For example, tasks like mitosis detection(2000x2000) requires higher dimension patches than lymphocyte detection(100x100).

After generation, selection methods are following:

1. Random Selection
2. Identify ROI(tumor), then random generation
3. Patch Clustering

7. Computational Pathology Journals & Conferences

Journals	**Impact Factor**
Journal of Pathology Informatics	3.23
Scientific Reports	5.133
Pattern Recognition	7.740
Medical Image Analysis	8.545
Artificial Intelligence in Medicine	6.69
Nature Biomedical Engineering	18.952
JAMA	8.483
Conferences	
International Conference on Machine Vision and Image Processing (MVIP)	
International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)	
Medical Imaging with Deep Learning (MIDL)	
IEEE International Conference on Bioinformatics and Biomedicine, (BIBM)	
IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)	

Table 5: CP Journals & Conferences

8. A typical Computational Pathology pipeline for cancer grading

Typical cancer grading pipelines are depicted in figure 4 and 5. The pipeline includes following steps:

1. Patch Generation
2. Stain/Color normalization (if required)
3. Patch Selection (Random, ROI, Clustering)
4. Dataset split (positive & negative examples)

5. Model Selection (GoogleNet, ResNet, AlexNet) & Training
6. Testing Results

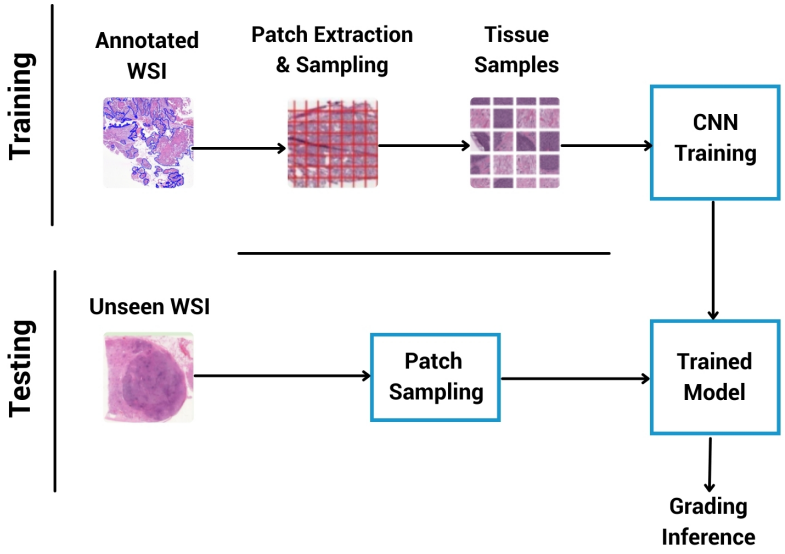

Figure 4: A generic Computational cancer grading pipeline

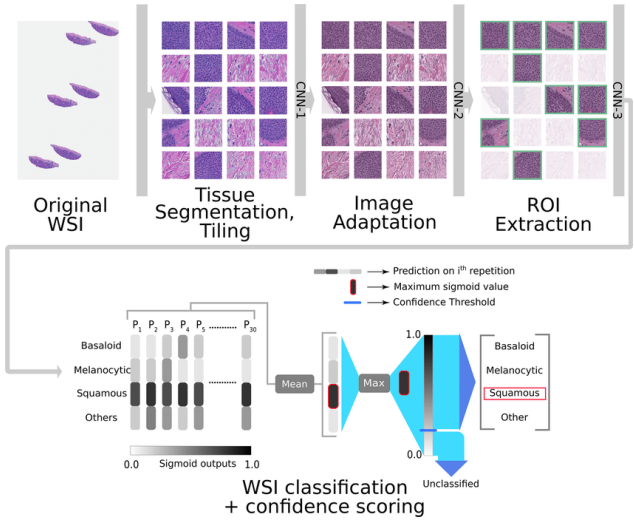

Figure 5: A generic Computational cancer classification pipeline

References

[1] Artificial intelligence diagnoses with high accuracy • healthcare-in-europe.com. https://healthcare-in-europe.com/en/news/artificial-intelligence-diagnoses-with-high-accuracy.html. (Accessed on 10/07/2021).

[2] Fda allows marketing of first whole slide imaging system for digital pathology — fda. https://www.fda.gov/news-events/press-announcements/fda-allows-marketing-first-whole-slide-imaging-system-digital-pathology. (Accessed on 10/02/2021).

[3] An intro to routine and special staining in histopathology : Leica biosystems. https://www.leicabiosystems.com/knowledge-pathway/an-introduction-to-routine-and-special-staining/. (Accessed on 10/06/2021).

[4] Andreea Anghel, Milos Stanisavljevic, Sonali Andani, Nikolaos Papandreou, Jan Hendrick Rüschoff, Peter Wild, Maria Gabrani, and Haralampos Pozidis. A high-performance system for robust stain normalization of whole-slide images in histopathology. *Frontiers in Medicine*, 6:193, 2019.

[5] Richard Chen, Ming Lu, Drew Williamson, Tiffany Chen, Jana Lipkova, Muhammad Shaban, Maha Shady, Mane Williams, Bumjin Joo, Zahra Noor, and Faisal Mahmood. Pan-cancer integrative histology-genomic analysis via interpretable multimodal deep learning. 08 2021.

[6] Zichao Guo, Hong Liu, Haomiao Ni, Xiangdong Wang, Mingming Su, Wei Guo, Kuansong Wang, and Taijiao Jiang. A fast and refined cancer regions segmentation framework in whole-slide breast pathological images. *Scientific Reports*, 9:882, 01 2019.

[7] Zilong Hu, Jinshan Tang, Ziming Wang, Kai Zhang, Ling Zhang, and Qingling Sun. Deep learning for image-based cancer detection and diagnosisa survey. *Pattern Recognition*, 83:134–149, 2018.

[8] Andrew Janowczyk and Anant Madabhushi. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *Journal of Pathology Informatics*, 7, 2016.

[9] Ming Lu, Drew F. K. Williamson, Tiffany Y Chen, Richard J. Chen, Matteo Barbieri, and Faisal Mahmood. Data efficient and weakly supervised computational pathology on whole slide images. *Nature biomedical engineering*, 2021.

[10] Ramin Nateghi, Habibollah Danyali, and Mohammad Sadegh Helfroush. A deep learning approach for mitosis detection: Application in tumor proliferation prediction from whole slide images. *Artificial Intelligence in Medicine*, 114:102048, 2021.

[11] Liron Pantanowitz. Digital images and the future of digital pathology. *Journal of pathology informatics*, 1, 08 2010.

[12] Ashwin Raju, Jiawen Yao, Mohammad MinHazul Haq, Jitendra Jonnagaddala, and Junzhou Huang. Graph attention multi-instance learning for accurate colorectal cancer staging. In Anne L. Martel, Purang Abolmaesumi, Danail Stoyanov, Diana Mateus, Maria A. Zuluaga, S. Kevin Zhou, Daniel Racoceanu, and Leo Joskowicz, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, pages 529–539, Cham, 2020. Springer International Publishing.

[13] Aman Rana, Alarice Lowe, Marie Lithgow, Katharine Horback, Tyler Janovitz, Annacarolina Da Silva, Harrison Tsai, Vignesh Shanmugam, Akram Bayat, and Pratik Shah. Use of Deep Learning to Develop and Analyze Computational Hematoxylin and Eosin Staining of Prostate Core Biopsy Images for Tumor Diagnosis. *JAMA Network Open*, 3(5):e205111–e205111, 05 2020.

[14] Mitko Veta, Yujing J. Heng, Nikolas Stathonikos, Babak Ehteshami Bejnordi, Francisco Beca, Thomas Wollmann, Karl Rohr, Manan A. Shah, Dayong Wang, Mikael Rousson, Martin Hedlund, David Tellez, Francesco Ciompi, Erwan Zerhouni, David Lanyi, Matheus Viana, Vassili Kovalev, Vitali Liauchuk, Hady Ahmady Phoulady, Talha Qaiser, Simon Graham, Nasir Rajpoot, Erik Sjöblom, Jesper Molin, Kyunghyun Paeng, Sangheum Hwang, Sunggyun Park, Zhipeng Jia, Eric I-Chao Chang, Yan Xu, Andrew H. Beck, Paul J. van Diest, and Josien P.W. Pluim. Predicting breast tumor proliferation from whole-slide images: The tupac16 challenge. *Medical Image Analysis*, 54:111–121, 2019.

[15] Zhaoyang Xu, Carlos Fernández Moro, Béla Bozóky, and Qianni Zhang. Gan-based virtual re-staining: A promising solution for whole slide image analysis. 01 2019.

[16] Jiawen Yao, Xinliang Zhu, Jitendra Jonnagaddala, Nicholas Hawkins, and Junzhou Huang. Whole slide images based cancer survival prediction using attention guided deep multiple instance learning networks. *Medical Image Analysis*, 65:101789, 07 2020.