Deliverable D3.1

WSI Preprocessing and Standardization Protocol

Project	**CLARIFY – Cloud ARtificial Intelligence For pathologY**
Grant Agreement ID:	**860627**
Consortium coordinator:	**UNIVERSITAT POLITECNICA DE VALENCIA**
Start and end date:	**1 November 2019 - 31 October 2023**
Funded under:	**H2020-EU.1.3.1.**
Date of issue:	**31-Aug-2021**
Due date:	**30-Apr-2021**
Leader in charge of deliverable:	**UNIVERSITETET I STAVANGER (UIS)**

Dissemination level	
X	PU = Public
	PP = Restricted to other programme participants (including the EC)
	RE = Restricted to a group specified by the consortium (including the EC)
	CO = Confidential, only for members of the consortium (including the EC)

CHANGE REGISTER

Version	Date	Author	Organization	Changes
A_DRAFT	06-Apr-2021	Neel Kanwal	UiS	Initial Draft
	30-Apr-2021	Kjersti Engan	UiS	Document restructuring
	21-May-2021	Arne Schmit	UGR	Addition of survey/figures
	24-May-2021	Neel Kanwal	UiS	Changes for internal review
	25-May-2021	Fernando Perez Bueno	UGR	Modification in the color section
	11-Jun-2021	Neel Kanwal	UiS	Draft modification
	28-Jun 2021	Kjersti Engan	UiS	Recommendation and conclusion etc.
	10-Jul-2021	Saul Fuster Navarro	UiS	Revision and modifications
	10-Jul-2021	Javier Mateos	UGR	Review and suggestions
	10-Jul-2021	Rafael Molina	UGR	Review and suggestions
	10-Jul-2021	Laëtitia Mariana Launet	UPV	Review and suggestions
	20-Jul-2021	Valery Naranjo Ornedo	UPV	Review and suggestions
	20-Jul-2021	Sandra Morales Martinez	UPV	Review and suggestions
	13-Aug-2021	Neel Kanwal	UiS	Reviews addressed
	20-Aug-2021	Kjersti Engan	UiS	Final version
A	31-Aug-2021	Valery Naranjo	UPV	Formatting only

Statement of independence

The work described in this document is genuinely a result of efforts pertaining to the CLARIFY project: any external source is appropriately referenced.

Confirmation by Authors: Neel Kanwal   UiS
                                        Kjersti Engan   UiS

Abbreviations

AI	Artificial Intelligence
CAD	Computer Aided Diagnostics
CNN	Convolution Neural Network
DICOM	Digital Imaging and Communication in Medicine

DL	Deep Learning
DNN	Deep Neural Network
DoA	Description of Action
DP	Digital Pathology
EMC	Erasmus Medical Centre
HR-NMBC	High-Risk Non-Muscle Invasive Bladder Cancer
HSM	Hue Saturation Mixture
HSV	Hue Saturation Value
H&E	Hematoxylin and Eosin
ICA	Independent Component Analysis
INCLIVA	Fundación para la Investigación del Hospital Clínico de la Comunistas Valenciana
ML	Machine Learning
NMF	Non-Negative Matrix Factorization
NN	Neural Network
OD	Optical Density
RGB	Red Green Blue
ROI	Region of Interest
SML	Spitzoid Melanocytic Lesions
SUH	Stavanger University Hospital
SVD	Single Value Decomposition
SVM	Support Vector Machine
TNBC	Triple-Negative Breast Cancer
TURBT	Transurethral Resection of Bladder Tumor
WP	Work Package
WSI	Whole Slide Image(s)

Table of Contents

List of Figures

List of Tables

1 Executive Summary

The overall goal of the CLARIFY - CLoud ARtificial Intelligence For pathologY - project is to develop novel AI-based methods for diagnosis and prognosis and content-based image retrieval using secure data handling for Whole Slide Images (WSI) from three different example diseases. This deliverable presents techniques useful for preprocessing and standardization of WSI, focusing on the three example diseases in CLARIFY.

Preprocessing can boost the performance of machine learning (ML) models trained on heterogeneous data collected from different institutions. This document describes challenges arising from processing datasets with a collection of different image formats, slide aging, and batch level effects from different scanners since three different pathology laboratories with different scanners are involved in data collection. Some common artifacts and their image processing solution to assess the quality of WSI and filter uninformative patches are reported. Example images collected from our datasets highlight artifacts, the differences in contrast, and brightness level. A general preprocessing pipeline is proposed to exclude non-relevant patches from the gigapixel images. It may not be imperative to propose a single stratified pipeline since requirements may vary for different diseases and different subsequent analyses. Preprocessing is a computationally expensive task; the selection of suitable methods may be a subjective choice. We outline the results of an internal survey of approaches currently used in different CLARIFY research groups and present a methodological overview of possible steps in the preprocessing pipeline.

Deliverable D3.1 is included in WP3 "*Automatic WSI interpretation through artificial intelligence and machine learning*" within Description of Action (DoA) of CLARIFY Project task T3.1 "*Preprocessing and standardization of WSI through advance image processing.*" This deliverable defines methods to process images that ensure a minimum quality level, read different microscopy scanner formats, color normalization, and identify relevant areas to facilitate subsequent analysis. It summarizes various aspects of image processing techniques applied to Hematoxylin and Eosin (H&E) patches.

2 Introduction

The overall goal of the CLARIFY - CLoud ARtificial Intelligence For pathologY - project is to develop novel AI-based methods for diagnosis and prognosis and content-based image retrieval using secure data handling for histological images using computational pathology. In the CLARIFY project, three cancer types are considered, namely, Non-muscle-invasive bladder cancer (NMBC), Spitzoid melanocytic lesions (SML), and Triple-negative breast cancer (TNBC). The available images are Whole Slide Images (WSI) stained with Hematoxylin and Eosin (H&E). WSI are high-resolution images commonly known as gigapixel images resulting from digital scanning of histological slides.

This document aims to present the preprocessing and standardization recommendations for the CLARIFY project. The deliverable document is a part of the work package (WP3), titled "Automatic WSI interpretation through artificial intelligence and machine learning." WP3 addresses:

1. Automatic interpretation of WSI based on Artificial intelligence (AI) for diagnosis and image retrieval.
2. How to efficiently preprocess and standardize histological WSI?
3. How to extract the most significant features?
4. How to find the best model that uses features to diagnose through different crowdsourcing strategies?
5. How to find similar cases based on feature similarity?

The deliverable D3.1 "*WSI preprocessing and standardization protocol*" is the output of actions of the task T3.1 "*Preprocessing and standardization of WSI through advance image processing*," with the following objectives:

1. Research and development of image processing and machine learning algorithms for WSI preprocessing and standardization.
2. Ensuring a minimum quality level of the images.
3. Recognition and reading of most used scanner file formats.
4. Identification of color variation and relevant areas in WSI for subsequent analysis.

Here is the description of partners in T3.1.
1. University of Stavanger (UiS) (Lead)
2. Universidad de Granada (UGR)
3. Universitat Politecnica de Valencia (UPV)
4. TYRIS Software (TYRIS)
5. Helse Stavanger HF (SUH)
6. ERASMUS Medical Centre (EMC)
7. Fundación para la Investigación del Hospital Clínico de la Comunitat Valenciana, Fundación INCLIVA (INCLIVA)

2.1 Image Processing for Histopathology

Pathology has been a front-face as a first critical step in diagnosing cancer over the last century. Histopathology is the examination of diseased tissue with a microscope, and it is an integral part of cancer diagnostics. The process starts with abnormal findings in routine clinical examinations. In order to investigate further, physicians order a tissue sample of the suspicious area [1]. The tissue is extracted in a biopsy, surgery, or autopsy. The tissue will thereafter be fixated in formalin, placed in wax, and cut in thin slices before being placed on a glass slide in the laboratory [2]. The potential handling of tissue at biopsy, fixation, microtomy, and scanning steps may result in artifacts. These slides can be stained differently with color dyes based on diagnostic analysis such as H&E or immunohistochemical staining (CD25, CD8, etc.). Hematoxylin stains acidic structures such as DNA and RNA with a purple tone. Eosin stains basic structures such as most cytoplasm proteins with a pink tone. The combination of H&E is the most used staining for tissue analysis [3]. A pathologist examines these slides and performs a microscopic analysis to find slides with abnormal cells. Digitization of this histopathological process helps to accumulate a large number of slides and create a digital archive to collect old samples for prognostic analysis [4].

Digital pathology (DP) refers to using a virtual microscope by digital scanning of tissue slides, producing WSI. WSI can be used to conduct broad computer-aided diagnosis (CAD) [5]. The CAD systems cater better image interpretation in conjunction with artificial intelligence algorithms [6]. The manual process of making a stained slide can cause some artifacts and variations in the appearance of the slides [2] [3]. In addition, different scanner hardwares will behave differently and might cause various artifacts based on its technology. There are dozens of microscopy scanners that can be categorized broadly as fluorescence and brightfield scanners. These digital slide scanners differ in many aspects, such as color sampling rate, scanning method, and output formats [7] [8].
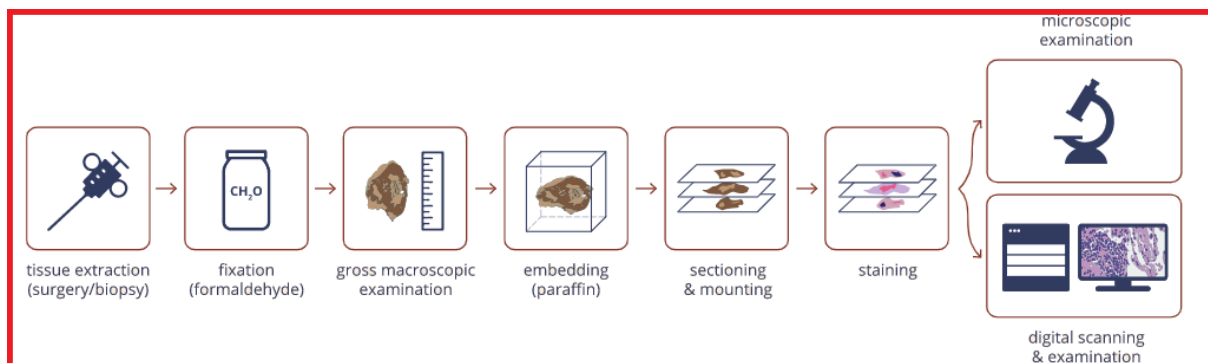
Figure 1: A standard workflow of histopathological laboratory for creating digital slides [1] (Photo taken from AnalogIntelligence.com with reprint permission).

The different diseases and problems we address in the CLARIFY project [9] might need different types and degrees of preprocessing; thus, this will not be proposed as a standardized system. CLARIFY objectives include developing AI models and Machine Learning (ML) algorithms to classify patients' disease based on diagnosis, such as grading and staging a cancer type [10]. Among other objectives, significant goals include prediction and prognosis in terms of the risk of recurrence and progression and to extract features that may contain other relevant prognostic information concerning the status of a patient, and extracting features to be used with a content-based image retrieval (CBIR) system [11].

As a step in getting there, we would need to extract Region of Interest (ROI) in the large WSI; for example, in NMBC, the urothelial tissue and its variants will include the most important diagnostic information. These ROIs may vary based on subsequent analysis in every cancer type and will be used to perform slide-level annotation. CLARIFY mainly deals with H&E-stained images for three cancer diseases. These images are collected at the hospitals of the medical partners from many surgeries and biopsies over many years. Since a retrospective prognostic analysis requires several years with follow-up information, we need to use slides prepared in different years [7].

Different laboratories and scanners can have different protocols and color variations, even from a fixed lab and scanner. The H&E staining can vary from day to day or due to the manual processes of making and staining slides. DP lacks a golden standard of single format and metadata structure [12]. Clinical institutions participating in CLARIFY project are using hardware from different vendors such as Hamamatsu, Leica, and Roche, resulting in different output formats. A critical analysis of available data is carried to investigate metadata and artifacts.
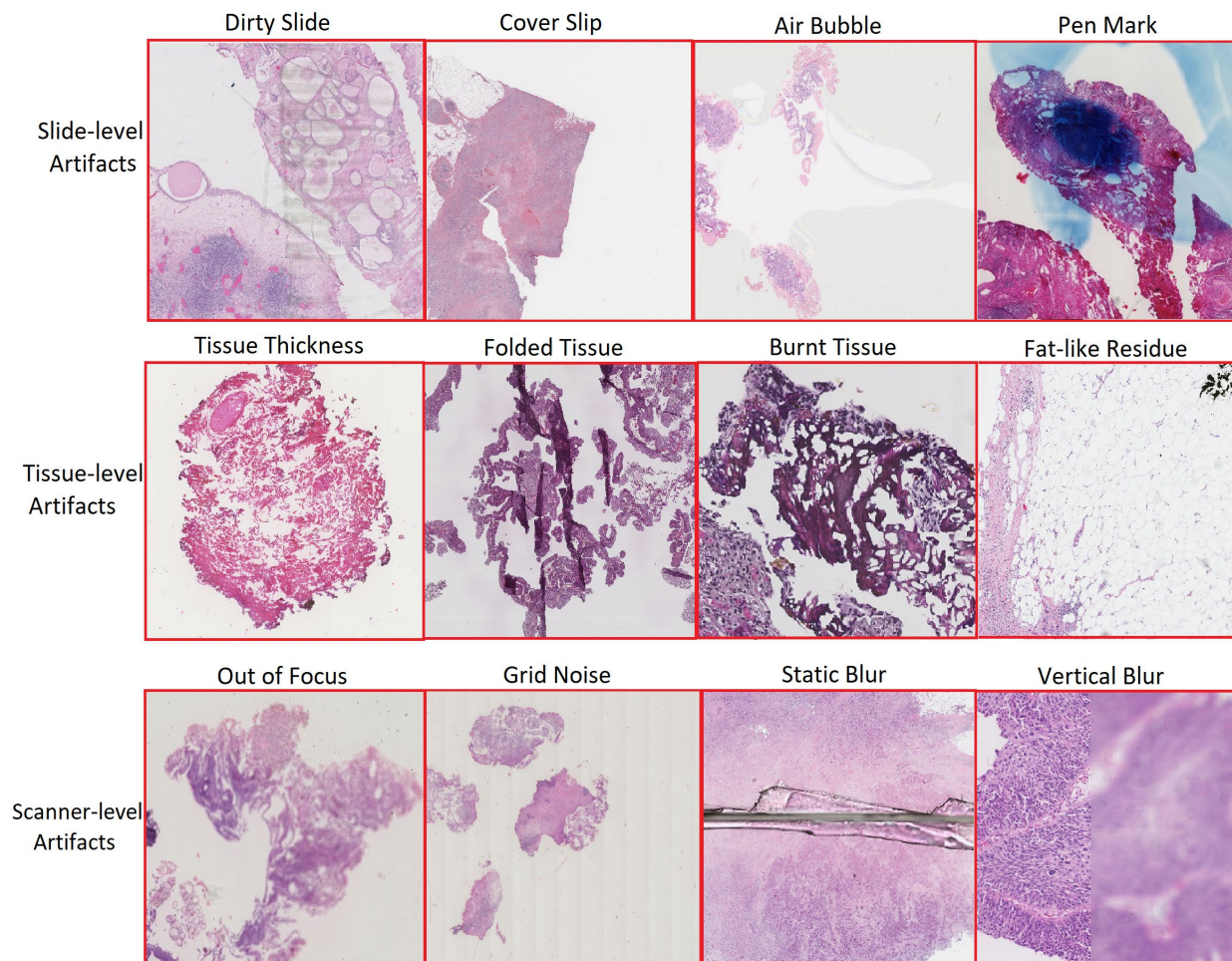
Figure 2: A collection of artifacts from different WSI that will be used in CLARIFY Project.

Handling images is critical in developing an intelligent computer vision algorithm [13] [14]. Noticeable artifacts are observed in some WSI. Some minor abnormalities involve a smaller portion of the specimen. The microdata in WSI is vast and contains many artifacts. These abnormalities can be broadly categorized into three streams 1) Slide-level artifacts, 2) Tissue-level artifacts, and 3) Scanner-level artifacts. These artifacts result from improper fixation, freezing, impregnation, and paraffin wax. Figure 2 demonstrates H&E patches with different artifacts.

We plan to provide a collection of possible approaches to deal with these artifacts. A preprocessing pipeline can be designed to cover all necessary filters to eliminate images that may deteriorate model performance. The selection of image filters and their thresholds plays a significant role in their computation. This parametric assessment may be tuned for different datasets based on lightness, cracks, pen marks, tissue fats, and other tissue anomalies. Before proceeding with this sequence of processing steps, quality checks are necessary. Image quality can be determined based on basic information such as magnification, format, and other derived information such as blur ratio and pixel-to-use. The grid patching method is suitable to perform image analysis in a further step.

Minor variations in dye ratios and tissue sectioning result in different colors from every batch of slides. For example, slides produced in a different laboratory and/or scanned by a different scanner or from different diseases. Training models using samples from various datasets results in more generalized models, and to do so, it would be beneficial to form a standardized training dataset format. Figure 3 shows images from three different host institutions of the CLARIFY project. It illustrates changes in their color, tissue orientation, and other scanning metadata parameters.

Color variation is a primary issue for the automatic analysis of images from different centers. Therefore, stain normalization is the most popular step in preprocessing. When it comes to color variations, we will go for various solutions such as i). Template-based color normalization uses global statics of template to modify pixels in the original image [15]. The template matching can be carried by various methods. ii) Color deconvolution works on separating stains and computing color in separate color-space [16]. Lastly, iii) Color augmentation methods can be used to iterative and recreate datasets to train a robust model for numerous stain variations [17].
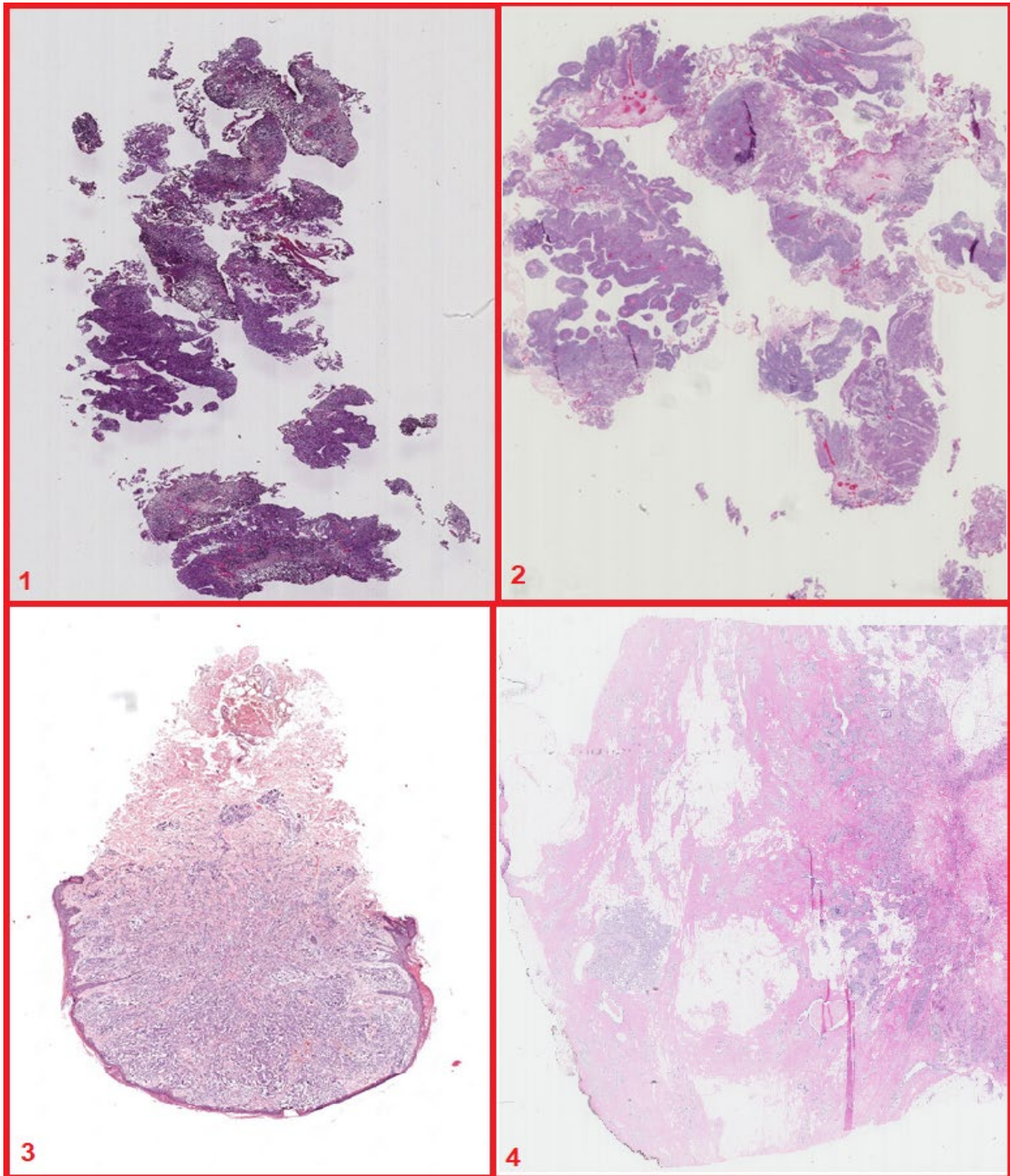
Figure 3: Images from different sources in the consortium belonging to different cancer types; 1) Non-Muscle Invasive Bladder Cancer (NMBC) from SUH; 2) HR- NMBC from EMC; 3) SML from INCLIVA; 4) TNBC from SUH

2.2 Deliverable Objectives

Data variation may output wrong predictions in the deployment phase of ML models [18]. DL algorithms designed for a specific dataset or type may not generalize well to another unseen dataset of a similar but yet different type. Numerous image processing libraries can read different histopathological WSI formats. The annotation tool we will use in the CLARIFY project is developed by the CLARIFY partner UPV. It requires a deep zoom image pyramid (DZI), so the first step for all images that will undergo manual annotations is to convert from the different scanner formats to DZI. However, for further processing, DZI might not fit well for processing compared to the original WSI. We may need to use the original images to create a preprocessing protocol.

The following questions will be addressed.

1. How can we assess the quality of the WSI?

2. What are different automatic methods to find various artifacts in WSI?

3. Which metadata is essential to collect for preprocessing?

4. How to define stages for a preprocessing pipeline?

5. What are common patching strategies?

6. Do we need a patches-on-the-fly approach while training/development?

7. What are the advantages and disadvantages of extracting patches before preprocessing?

8. Which magnification level is suitable to create patches?

9. Can we use multiple resolutions to train the network?

10. Can ROI-based patches help find positive and negative classes?

11. Before training and prediction, do we need to detect and exclude specific images, like a blur, folded regions, cauterized areas?

12. Is deblurring of patches useful in our large datasets?

13. What is a suitable patch size to train accurate models?

14. Should color normalization be part of the preprocessing?

15. Which color matching methods are appropriate in image processing?

16. What (local/global) parameters to evaluate for normalization outcomes?

17. Do we need to perform color augmentation of the whole dataset?

18. Is color augmentation while training an alternative to color normalization?

19. How to overcome the class imbalance problem?

2.3 Data Availability

The CLARIFY project aims to use histopathological images of NMBC, SML, and TNBC. The WSI datasets are collected from all the partner hospitals. SUH will contribute to breast cancer and bladder cancer diseases, EMC for HR-NMBC data, and INCLIVA for SML images only.

Thus, the data material is heterogeneous in multiple ways; i) medical nature due to different diseases, ii) staining variations coming from multiple labs, iii) data format and quality variations due to different scanning hardware, and iv) further diagnostic analysis might be relevant for the different diseases. Different scanners in use at the different partner hospitals include

A) <u>Leica SCN400</u> by Stavanger University Hospital.
B) <u>Hamamatsu NanoZoomer 2.0HT</u> by Erasmus Medical Centre Rotterdam.
C) <u>Ventana, HTi Scan by Fundación</u> INCLIVA Valencia.

The digital pathology market has a handful of vendors with their proprietary image acquisition technology [19]. The scanner performs either fluorescence or brightfield illumination. Scanner manufacturers have different light interpolation algorithms for (stitching, stacking), which manipulate images in different batches. WSI images are typically huge and are known as gigapixel images. The images can be approximated to size around 100K x 100K pixels, and typically the full resolution level corresponds to 40x magnification of the tissue. WSI is significant because they have many images at different resolutions to preserve different optical zooms. These gigapixel images can be viewed on their propriety image viewers or ordinary web-based viewers such as

A) Aperio Image Scope for Leica Scanner with image format "scn."
B) NDP.view2 for Hamamatsu Scanner with image format "ndpi."
C) QuPath for Roche Ventana Scanner with image formats "tif" or "bif."

These viewers also work with other standard digital microscopic images but may lack some interpolative features due to variation in image compression methods. A DICOM viewer (<u>Sedeen Viewer</u>) can be used for nearly all formats. It shows separately micro and macro data with all properties. Viewers are helpful for the pathologist in observing images and their properties.

Color variation in histological images has a considerable influence on ML models. It is crucial to minimize the computational impact of different staining to diagnose cancer effectively [15]. Figure 4 shows three samples of NMBC from SUH, and all of them show stain variation due to different tissue thicknesses. Along with it, when tumors are surgically removed by transurethral resection of bladder tumor (TURBT) for slide making, some parts of the tissue are damaged/burnt due to hot sectioning apparatus. Although the slides were scanned using the same hardware, there is a noticeable color difference in the images. This can be due to day-to-day variations in the staining at the lab, different ages of slides before scanning, different degrees of cauterized areas after TURBT, etc.
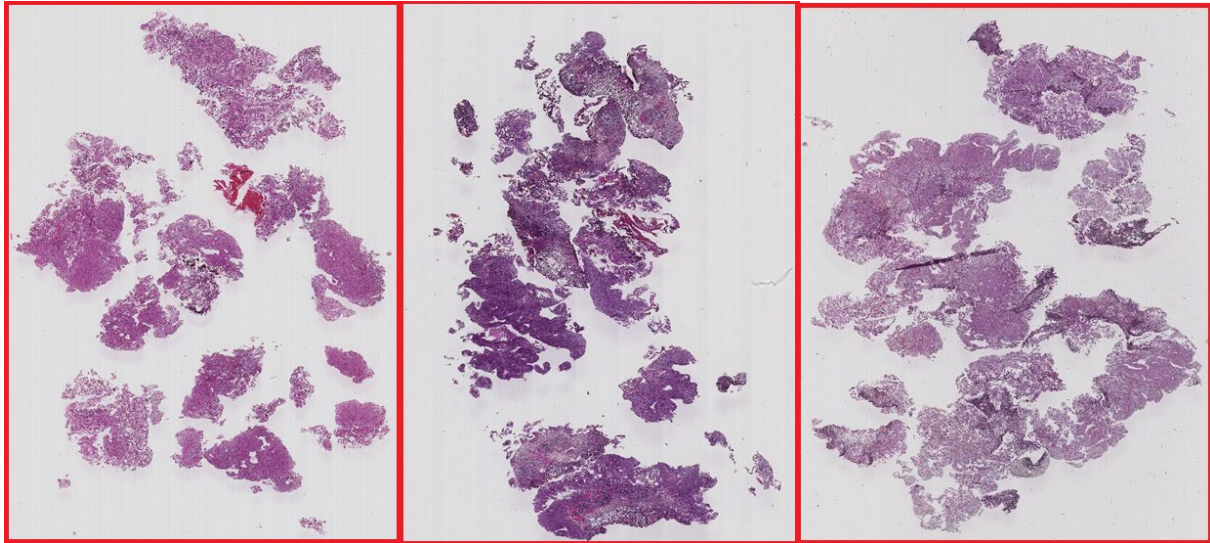
Figure 4: WSI of NMBC from SUH

Figure 5 highlights another similar example of stain variability in WSI of NMBC from EMC, where slides were stained in different years. Moreover, the difference between figure 4 and figure 5 are both examples of the same cancer type (NMBC). Figure 6 shows three examples of WSI of SML from INCLIVA, and Figure 7 shows TNBC samples from SUH. Here we can evaluate noticeable color differences even when the WSI is from the same laboratory and disease.
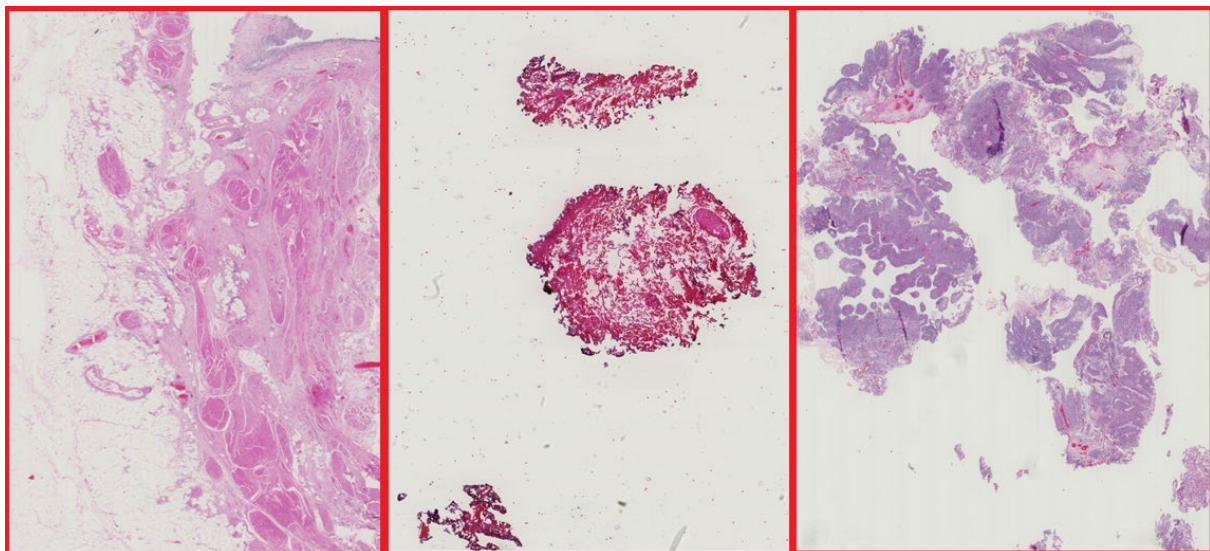
Figure 5: WSI of HR-NMBC from EMC
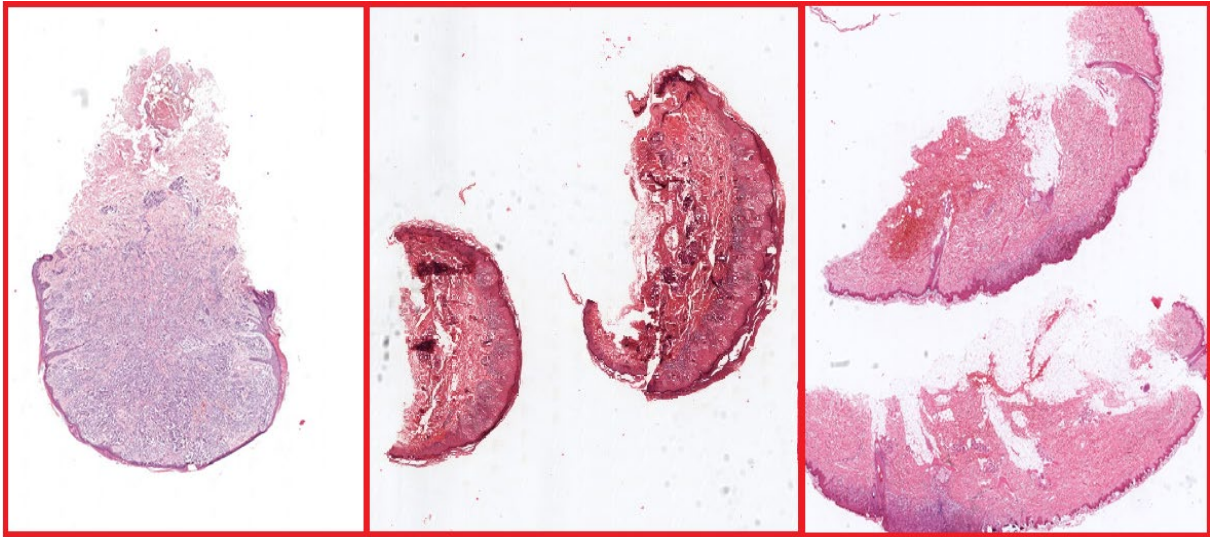
Figure 6: WSI of SML from INCLIVA
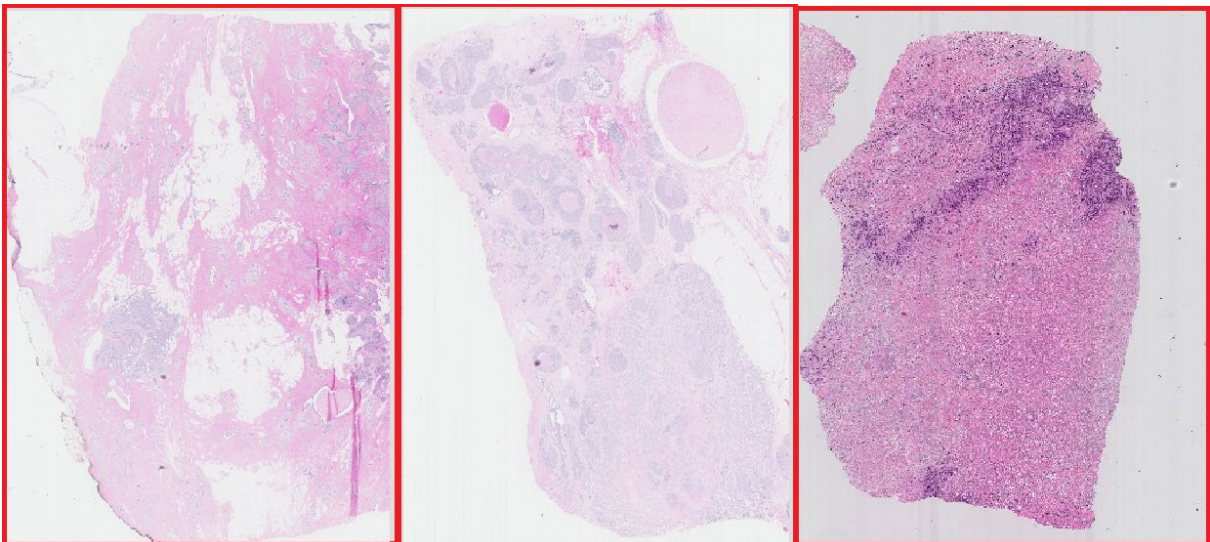
Figure 7: WSI of TNBC from SUH

Similar variation can be observed in other publicly available datasets, as shown in table 1.

Dataset Name	Dataset Size	Format	Annotation Labels	Class Balance	Access URL
BACH	400 H&E Micros copy (2048 x 1536) 400 WSI	.tiff .svs	Normal, Benign, Insitu Carcinoma, Invasive Carcinoma	YES	https://iciar2018-challenge.grand-challenge.org/Dataset/
TUPAC16	500 WSI 73 WSI for Mitosis 148 WSI for ROI	.svs .tiff .svs	Tumor Proliferation Score, Molecular Proliferation Score, Mitotic Figures, Regions for Mitosis Counting	N/A	https://tupac.tue-image.nl/node/3 https://arxiv.org/ftp/arxiv/papers/1807/1807.08284.pdf
CAMELYON`16	400 H&E WSI	.tif	Binary lymph node metastases	No	https://drive.google.com/drive/folders/0BzsdkU4jWx9Bb19WNndQTlUwb2M
CAMELYON`17	1000 WSI	.tiff	pN-stage label, Detail Metases	N/A	https://drive.google.com/drive/folders/0BzsdkU4jWx9BaXVHSXRJTnpLZU0
BreakHis	9109 H&E (700*460)	.png	Malignant, Benign	No	https://www.kaggle.com/ambarish/breakhis
MITOSIS (2014)	284 H&E at 20x 1136 H&E at 40x	.tiff	Nuclear atypia score [1-3] True Mitosis, Probably Mitosis,	No	https://mitos-atypia-14.grand-challenge.org/dataset/

			Nota Mitosis		
MITOSIS (2012)	50 H&E at 40x (2084*2084) and (2252*2250)	.tiff	Prophase, Metaphase, Telophase	N/A	http://ludo17.free.fr/mitos_2012/dataset.html
Sentinel Lymph Node (Pcam)	327680 (96*96) fro 400 H&E stained WSI	HDF5	a binary label for presence of metastic tissue	YES	https://github.com/basveeling/pcam https://www.kaggle.com/c/histopathologic-cancer-detection/data
TCIA	96 H&E WSI at 20x Aperio	.svs	Low, Medium and High Tumor Cellularity, TIL-E,TIL-S, Epithelial Normal,UDH,ADH Malignant Epithelia, IDC, ILC, MuC, DCIS ,MC	No	https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=52758117
TCGA-BRCA	1134 H&E WSI at 40x Aperio	.svs	N/A	N/A	https://cancer.digitalslidearchive.org/#!/CDSA/brca/TCGA-3C-AALI/TCGA-3C-AALI-01Z-00-DX2?slide=TCGA-3C-AALI-01Z-00-DX2.CF4496E0-AB52-4F3E-BDF5-C34833B91B7C.svs

Table 1: Publicly available datasets

2.4 Metadata Handling

Metadata can be defined as information that describes the data. It is generally known as data of the data [20]. Metadata can be used to understand the intrinsic difference within images in terms of basic statistics such as magnification, pixel size, levels, format, and other file-related information. It helps retrieve data from storage or compare sources of the images. Metadata is useful for cataloging and contextualizing visual information, and it is different for every hardware vendor.

2.4.1 Image Metadata

Metadata can be generally classified into three categories; i) Technical metadata: It gives an overview about width, height, objective focus, z-spacing, etc. ii) Descriptive metadata: It includes information related to file name, scan date, image description, etc. and lastly iii) Administrative metadata: This mainly provides vendor related information such model type, software version, etc. Figure 8 shows a combination of image macro and image properties with Sedeen viewer. Figure 9 shows a small range of metadata for file formats ".scn," ".ndpi," and ".tif."
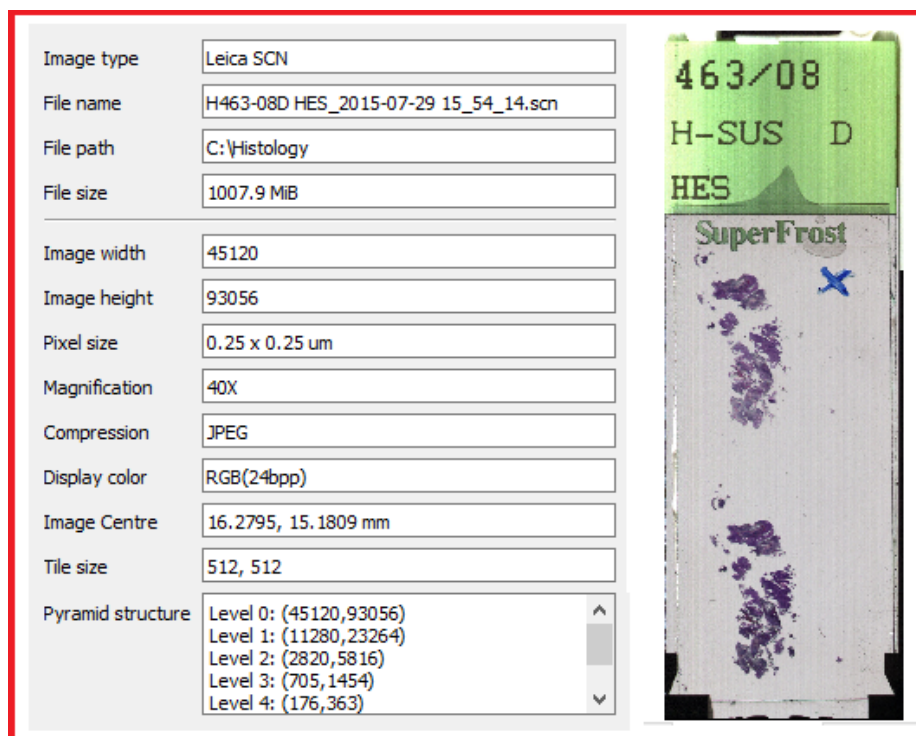
Figure 8: WSI metadata and macro from a SUH data sample (Image properties output by Sedeen Viewer)

```
'openslide.level[2].downsample',        ['width',                                ['width',
'openslide.level[2].height',             'height',                                'height',
'openslide.level[2].tile-height',        'bands',                                 'bands',
'openslide.level[2].tile-width',         'format',                                'format',
'openslide.level[2].width',              'coding',                                'coding',
'openslide.level[3].downsample',         'interpretation',                        'interpretation',
'openslide.level[3].height',             'xoffset',                               'xoffset',
'openslide.level[3].tile-height',        'yoffset',                               'yoffset',
'openslide.level[3].tile-width',         'xres',                                  'xres',
'openslide.level[3].width',              'yres',                                  'yres',
'openslide.level[4].downsample',         'filename',                              'filename',
'openslide.level[4].height',             'vips-loader',                           'vips-loader',
'openslide.level[4].tile-height',        'slide-level',                           'slide-level',
'openslide.level[4].tile-width',         'hamamatsu.AHEX[0]',                     'leica.aperture',
'openslide.level[4].width',              'hamamatsu.AHEX[0].fluorescence',        'leica.creation-date',
'openslide.level[5].downsample',         'hamamatsu.AHEX[0].ploidy',              'leica.device-model',
'openslide.level[5].height',             'hamamatsu.AHEX[1]',                     'leica.device-version',
'openslide.level[5].tile-height',        'hamamatsu.AHEX[1].fluorescence',        'leica.illumination-source',
'openslide.level[5].tile-width',         'hamamatsu.AHEX[1].ploidy',              'leica.objective',
'openslide.level[5].width',              'hamamatsu.AHEX[2]',                     'openslide.bounds-height',
'openslide.level[6].downsample',         'hamamatsu.AHEX[2].fluorescence',        'openslide.bounds-width',
'openslide.level[6].height',             'hamamatsu.AHEX[2].ploidy',              'openslide.bounds-x',
'openslide.level[6].tile-height',        'hamamatsu.Created',                     'openslide.bounds-y',
'openslide.level[6].tile-width',         'hamamatsu.MHLN[0].24bit',               'openslide.level-count',
'openslide.level[6].width',              'hamamatsu.MHLN[0].36bit',               'openslide.level[0].downsample',
'openslide.level[7].downsample',         'hamamatsu.MHLN[1].24bit',               'openslide.level[0].height',
'openslide.level[7].height',             'hamamatsu.MHLN[1].36bit',               'openslide.level[0].width',
'openslide.level[7].tile-height',        'hamamatsu.MHLN[2].24bit',               'openslide.level[1].downsample',
'openslide.level[7].tile-width',         'hamamatsu.MHLN[2].36bit',               'openslide.level[1].height',
'openslide.level[7].width',              'hamamatsu.Macro.S/N',                   'openslide.level[1].width',
'openslide.mpp-x',                       'hamamatsu.NDP.S/N',                     'openslide.level[2].downsample',
'openslide.mpp-y',                       'hamamatsu.Objective.Lens.Magnificant', 'openslide.level[2].height',
'openslide.objective-power',             'hamamatsu.PSHV',                        'openslide.level[2].width',
'openslide.quickhash-1',                 'hamamatsu.PSHV.10x',                    'openslide.level[3].downsample',
'openslide.vendor',                      'hamamatsu.PSHV.40x',                    'openslide.level[3].height',
'tiff.DateTime',                         'hamamatsu.PSHV.ploidy',                 'openslide.level[3].width',
'tiff.ImageDescription',                 'hamamatsu.PSHV.ploidy.10x',             'openslide.level[4].downsample',
'tiff.Make',                             'hamamatsu.Product',                     'openslide.level[4].height',
'tiff.Model',                            'hamamatsu.SourceLens',                  'openslide.level[4].width',
'tiff.ResolutionUnit',                   'hamamatsu.Updated',                     'openslide.mpp-x',
'tiff.Software',                         'hamamatsu.XOffsetFromSlideCentre',      'openslide.mpp-y',
'ventana.BuildDate',                     'hamamatsu.YOffsetFromSlideCentre',      'openslide.objective-power',
'ventana.BuildVersion',                  'hamamatsu.YRNP[0]',                     'openslide.quickhash-1',
'ventana.FocusMode',                     'hamamatsu.YRNP[1]',                     'openslide.region[0].height',
'ventana.FocusQuality',                  'hamamatsu.YRNP[2]',                     'openslide.region[0].width',
'ventana.LabelBoundary',                 'hamamatsu.calibration.version',         'openslide.region[0].x',
'ventana.Magnification',                 'hamamatsu.ccd.height',                  'openslide.region[0].y',
'ventana.Mode',                          'hamamatsu.ccd.width',                   'openslide.vendor',
'ventana.ScanDate',                      'hamamatsu.ccd.width.ploidy',            'tiff.ResolutionUnit',
'ventana.ScanMode',                      'hamamatsu.coarse.focus.pitch',          'tiff.XResolution',
'ventana.ScanRes',                       'hamamatsu.colorfilterID',               'tiff.YResolution',
'ventana.ScannerModel',                  'hamamatsu.cube.kind',                   'slide-associated-images']
'ventana.ShowLabel',                     'hamamatsu.exposure.barcode.macro',
'ventana.UnitNumber',                    'hamamatsu.exposure.slide.darkfield.macro',
'ventana.Z-layers',                      'hamamatsu.exposure.slide.macro',
'ventana.Z-spacing',                     'hamamatsu.fine.focus.pitch',
'slide-associated-images']               'hamamatsu.focalplane.leftbottom',
                                         'hamamatsu.focalplane.lefttop',
                                         'hamamatsu.focalplane.rightbottom',
```

Figure 9: Metadata properties obtained for three data formats using OpenSlide. The figure does not show all parameters but overviews technical, descriptive, and administrative metadata.

2.4.2 Annotation Metadata

In CLARIFY, experts' annotations are performed by uropathologists, breast pathologists, and dermo pathologists. It is considered essential metadata to identify ROI in the H&E slide. Detailed manual annotations can be pretty valuable for supervised learning. In this project, we have protocols with particular labels and variant histology types for every cancer type. On the other hand, annotation is performed on a local tool that saves labeling metadata in a schema in a relational database. The web-based annotation tool offers strong support for pixel-level and slide-level annotation. Pixel-level annotation helps identify regions and locate mitosis or lymphocyte subtype classification, whereas slide-level annotation may include placing image-level information such as cancerous or staging data. Table 2 shows a summary of all three annotation protocols for NMBC, SML, and TNBC.

Cancer Kind	Annotation Label
HR-NMIBC (High-Risk Non-Muscle Invasive Bladder Cancer)	**1. Urothelium tissue:**
	(1) without immune cell infiltration
	(2) with immune cell infiltration
	(3) normal
	(4) cancerous
	(5) variant histology
	(6) flat lesion
	2. Stroma tissue:
	(1) without immune cell infiltration
	(2) with immune cell infiltration
	3. Muscle
	4. Blood
	5. Damaged tissue:
	(1) Cauterized
	(2) Blurry
	(3) Folded
	(4) Others
	6. Cancerous invasive areas

	7. Mitosis
	8. WHO04/16:
	(1) High grade
	(2) Low grade
	9. WHO73:
	(1) Grade 1
	(2) Grade 2
	(3) Grade 3
	10. Flat lesions
	(1) Dysplasia
	(2) CIS
	11. Staging
	1. pTa
	2. pT1
	3. pTis
SML (Spitzoid Melanocytic Lesions)	**Global regions:**
	1. Spitzoid nevus (benign)
	2. Spitzoid tumor of uncertain malignant potential (STUMP)
	3. Spitzoid melanoma (malignant)
	Small patterns:
	Typical mitosis
	Atypical mitosis
	Ulcers
	Necrosis
TNBC (Triple Negative Breast Cancer)	**Histopathological subtype labels:**
	1. infiltrating duct carcinoma
	2. lobular carcinoma

	3. mixed ductal-lobular carcinoma
	4. Others
	Interesting factor labels:
	1. Tumor infiltrating lymphocytes (TILs)
	2. Typical mitosis
	3. Atypical mitosis
	4. Fibrotic focus
	5. Necrosis
	6. Adipocytes

Table 2: Description of Annotation Metadata for MicroDraw

2.4.3 Clinical Metadata

Clinical data is also considered as metadata for diagnosis and prognosis. Every health institution in this consortium has its specific format of book-keeping for hospital records. It can also be considered as patient-level information, which is valuable for developing recurrence and survival models. The aggregated approach is also used by combining annotations and clinical information, which is generally pseudonymized before research.

Tables 3,4 and 5 provide detailed clinical metadata layout for NMBC (by SUH), TNBC (by SUH), and SML (by INCLIVA), respectively.

Identifier	Description	Field value
WHO 73	Grade of the primary tumor after WHO -73. This is done by an experienced uropathologist when going through all cases for validation.	1= grade 1
		2= grade 2
		3= grade 3
WHO 04	Grade of the primary tumor after WHO -04. This is done by an experienced uropathologist when going through all cases for validation.	0= PUNLMP
		1=low grade
		2=high grade
Stage	T-stage for primary tumor	1=Ta
		2=T1

		3=T2
		4=T3
		5=Tis
CK20	CK20 Immunohistochemistry	1= negative (≤3)
	Immuno Reactive Score, threshold 3	2=positive (>3)
MAI SUH protocol	Mitotic activity index, number of mitosis in tumor cells at 1,59 mm^2.	0=Low ≤15
	Threshold 15	1=High >15
Ki67	Ki67 measured by image analysis. Threshold 39%	0=Low ≤39
		1=High >39
Recurrence	Recurrence of urothelial carcinoma/ carcinoma in situ locally in the bladder (not registered after cystectomy).	0=No
		1=Yes
Time to recurrence	Time from a date at diagnosis to local recurrence in the bladder (in months). Those without recurrence are given time from diagnosis to last known control with cystoscopy or cystectomy.	
Stage Progression	Any progression in TNM (also after cystectomy).	0=No
		1=Yes
Time to progression	Time from the date at diagnosis to any progression in TNM (in months). Those without progression are given time from diagnosis to death, to last known contact if they moved, or until 30.6.2016 for the rest.	
Metastasis	Metastasis to nodes (N) or distant metastasis (M) from urothelial carcinoma/ carcinoma in situ in the bladder. The evaluation is based on histology, radiology, and/ or clinical information in the patient journal.	0=No
		1=Yes
Follow-up recurrence	Time from diagnosis (in months) to cystectomy or last control with cystoscopy (might be most relevant for recurrences).	
Follow-up progression	Time from diagnosis (in months) to death or until 30.6.2016 (might be most relevant for progression).	
Sex	Sex	1=male
		2=female
Age		
Multifoc	Multifocal tumor	0=No

		1=Yes
Size	Tumor size (mm)	
CIS	Carcinoma in situ	0=No
		1=Yes
PPH3	Phospho HistoneH3 Immunohistochemistry in 1.59 mm2. Threshold 38	1= low ($\leq$38)
		2=high (>38)
CD25	CD25 Immunohistochemistry. Threshold 1%.	1= low (<1)
		2=high ($\geq$1)
P53	P53 immunohistochemistry. Threshold 15%	1= low (<15)
		2=high ($\geq$15)
BCG	BCG treatment	0=No
		1=Yes
Chemo	Received chemotherapy	0=No
		1=Yes
Immunotherapy	Received immunotherapy	0=No
		1=Yes
NGS	Next-Generation sequencing performed Oncomine Focus Panel (DNA only)	0=No
		1=Yes
FGFR3-alteration	FGFR3 hotspot mutation?	0=No
		1=Yes

Table 3: Description of Clinical Metadata for NMBC by SUH

Identifier	Description	Field Value
eofus	End of Follow Up Status	1: Alive and well
		2: Alive with Distant Metastasis
		3: Alive with local recurrences
		4: Dead of other causes
		5: Dead of Disease with local recurrences
		6: Dead of Disease with Distant Metastasis
		7: Lost to follow-up
FUT	Follow Up Time in Months	Date last control – date breast cancer diagnosis in days /30.5
cur_rela	Current relapse status, at the date of last control	1: no metastasis
		2: locoregional recurrences
		3: distant metastasis
		4: Lost from follow-up
		9: Unknown
meta life	Any Meta or not in a lifetime, including locoregional recurrences	1: No meta
		2: Meta
		9: Unknown or LFTU
meta1e	First metastasis	1: local regional
		2: other breasts
		3: bone
		4: liver or another organ
		5: brain
		6: multiple
		8: not applicable
		9: unknown
MAI	Mitotic Activity Index (mitosis/1.59 mm2)	

PPH3	PPH3 evaluation (positive cells/1.59mm2)	
Ki67	Ki67 percentage hotspot/cold spot in 500 tumor cells	
MAI_10	MAI 9 or lower vs. MAI 10 and higher	1: MAI<10
		2: MAI>9
		9: unknown
PPH3_13	H3 with threshold 13	0: PPH#<13
		1: PPH3>=13
		9: unknown
Age	Age at time of diagnosis	Time of the first diagnosis – birthday in days /365
origrade	Original grade	1: grade 1
		2: grade 2
		3: grade 3
tub_form	Tubular formation	1: >75%
		2: 10-75%
		3: >10%
		9: unknown
nuc_atyp	Nuclear Atypia	1: mild
		2: moderate
		3: marked
		9: unknown
mit_imp	Mitotic Impression	1: 0-5
		2: 5-10
		3: >10
		9: unknown
sumgrad	Sum of tub, Nuc At, Mit Imp.	

Nottgrade	Nottingham grade	1: 3-5
		2: 6-7
		3: 8-9
		99: unknown
tum_size	Tumor size [cm]	From the original pathology report
tsize2		1: Tsize<= 2.0 cm
		2: Tsize>2.0 cm
		9: unknown
oestr_re	Estrogen Receptor based upon IHC >1%	1: positive
		2: dubious (1-10%)
		3: negative
		9: unknown
prog_rec	Progesteron Receptor based upon IHC >10%	1: positive
		2: dubious (1-10%)
		3: negative
		9: unknown
htype	Histologic type Tumor	5: Ductal
		7: other
l_status	Lymph Node Status	1: positive
		2: negative
		3: no lymph nodes found
		9: unknown/not performed
n_nodes	Number of nodes	
np_nodes	Number positive nodes	
HER2	HER2 based upon IHC, 10% tumor cells	0: negative =score 0
		1: negative = score 1
		2: positive = score 2
		3: positive = score 3

FUT_recur	Time to the first recurrence	
LeucoInfiltration	Leukocyte infiltration= TILs	1: No
		2: Yes
Siteleuco	Where are the leucocytes located	1: infiltrating
		2: tumor border
		3: spot like
		4: everywhere
		5: focally without a typical pattern
FF	Fibrotic focus	0: absent
		1: present
FFsize	Fibrotic focus size	1: < 1/3 of the tumor
		2: > 1/3 of the tumor
Necrosis	Necrosis in fibrotic focus	0: absent
		1: present

Table 4: Description of Clinical Metadata for TNBC by SUH

Identifier	Description	Field Value
Age (years)	Age at the time of diagnosis	1. 0-11
		2. 12-18
		3. 19-29
		4. 30-49
		5. >50
Gender	Type of gender	1. Female
		2. Male
Tumor location	Location of the primary tumor	1. Head and neck
		2. Trunk
		3. Upper limb and shoulder
		4. Lower limb and hip

		5. Not recorded or specified
Date of Diagnosis	Year and month of the histopathological diagnosis	year/month
Local recurrence	New tumor growth at the primary site after the surgical excision	1. No
		2. Yes
Date of local recurrence	Year/month of recurrence diagnosis	year/month
Regional relapse		1. No
		2. Satellite/in transit metastasis
		3. Lymph node metastasis
		4. Lymph node AND satellite/in transit metastasis
Date of regional relapse	Year/month of regional relapse diagnosis	year/month
Distant metastasis	Tumor dissemination to distant sites	1. No
		2. Distant metastasis to skin, soft tissue and/or nonregional lymph node
		3. Distant metastasis to lung
		4. Distant metastasis to non-Central Nervous System visceral sites
		5. Distant metastasis to Central Nervous System
Date of distant metastasis	Year/month when the distant metastasis was detected	year/month
Follow up		1: Alive and well
		2: Alive with local recurrence
		3: Alive with regional relapse
		4. Alive with distant metastasis
		5: Dead of Disease
Last date of follow up	Last control	year/month

Table 5: Description of Clinical Metadata for SML by INCLIVA.

2.5 CLARIFY Survey

In order to empirically assess the different approaches used by CLARIFY research teams and provide an overview of common preprocessing strategies, a survey was conducted on this topic. There were three sections with a total of twenty-two questions. The questions focused on collecting feedback and experiences with multiple choices and short answers in cancer types, preprocessing methods, image formats, and data handling. A total of 17 CLARIFY researchers have anonymously answered the questionnaire. Participants had a track of working with different cancer types and data formats.

On average, researchers have hands-on experience working more on skin cancer followed by breast and bladder cancer WSI. A major of research was done on H&E WSI rather than other staining types due to its common diagnostic importance. Figure 10 and Figure 11 illustrate the responses. The results of the survey will be included in the relevant chapters in the next section.
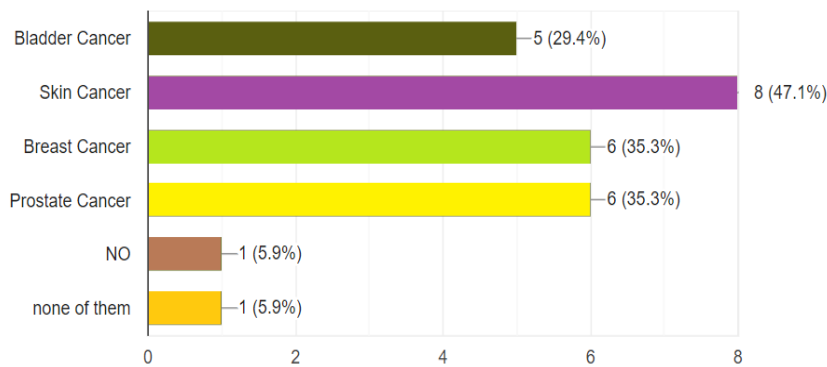
Figure 10: An answer to the question based on research in a cancer type.
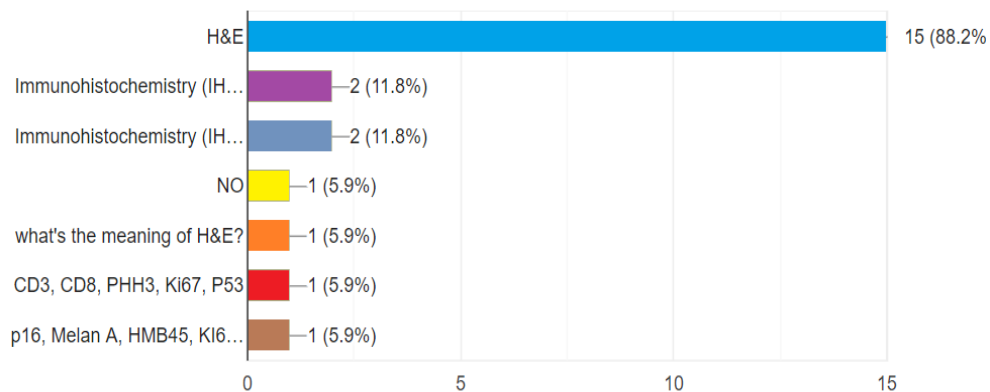
Figure 11: An answer from a questionnaire for handling WSI stains.

3 Preprocessing Pipeline

WSI is produced by scanning a histological tissue slide with a digital microscope scanner of high resolution. It requires suitable parametric calibration for a detailed analysis [21]. Thus, WSI is typically massive, varying from a few hundred megabytes to several gigabytes in size. The large size makes it intractable to design systems to load the entire image into the computer memory during training of DL algorithms etc. Theoretically, there could be several choices in setting up ML pipelines. The selection of processing stages depends on relevant artifacts, data-generating strategy, and batching method for task-specific classifiers [22].

Different workflows can be adapted to save a preprocessed WSI dataset or to deal with the necessity of patching the gigapixel images before processing:

A) Patch on-the-fly: One possible workflow is to work with a list of coordinates pointing to patch corners ROI to extract the patches on the fly. Different coordinates can be tagged with multiple classes, for example, both tissue type and grade. For further processing and training of models, multiple coordinate lists can be used as input, and different resolutions and sizes of tiles are formed and processed on the fly. Optional preprocessing steps can be performed on the fly. Alternatively, preprocessing steps as color normalization can be performed once, and the relevant patches can be stitched together and stored again as larger images.

B) Pre-patch: A different workflow patches up the images based on the annotation input or other preprocessing steps. These patches are stored for further processing. After that, these fixed patches are used in further processing and training of models.

These workflows will be used in CLARIFY depending on what is best suited for the specific problem and the partner's hardware facilities. Figure 12 illustrates (when-to-patch) processing.
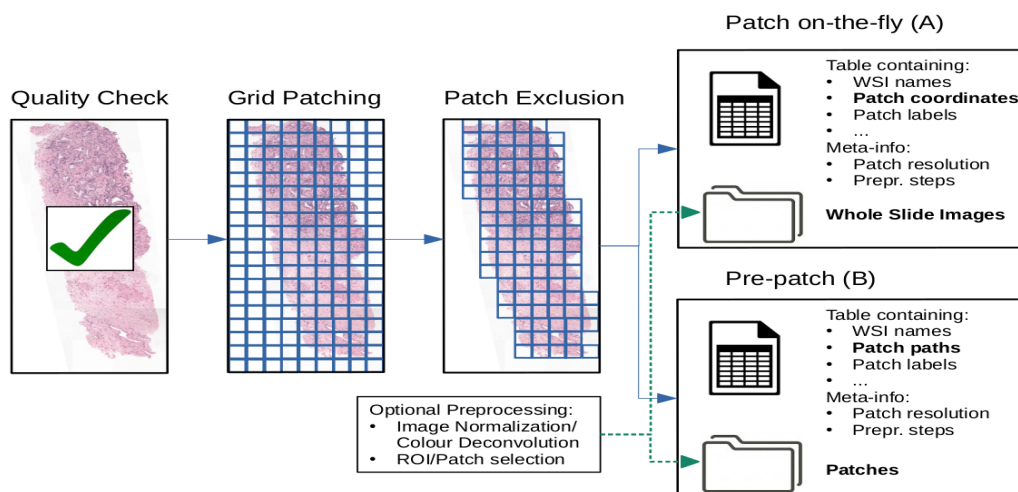
Figure 12: Common Preprocessing Pipeline. Some preprocessing steps are optional and can vary depending on the cancer type, dataset, and deep learning methodology. To save the preprocessed dtaset, either WSI-based Saving or Patch-based Saving can be used. Illustration for (When-to-patch) patch where patching on-the-fly and pre-patch are explained.

Here is a conceivable flow diagram that proposes a five-stage preprocessing pipeline to deal with several aspects of data processing. It is not imperative to select a single method due to variation in artifacts and metadata, as described in section 2. Moreover, our internal survey shows that most researchers in CLARIFY network work with the pre-patch approach (B) instead of the patch on-the-fly approach (A). 20% say they use a combination of strategy A and B, depending on the current problem, see Figure 13.

Regardless of the decision of *when to patch*, i.e., pre-patch or on-the-fly, more or less the same pipeline can be thought of for the preprocessing steps.

Figure 14 depicts two possible preprocessing pipelines with the same main steps or challenges: 1) Image quality check 2) Patching method (*what to patch*) 3) Irrelevant patches and artifact detection (patch exclusion) 4) color variation challenge (normalization or augmentation) 5) Patch selection (task-dependent choice of subsets of patches and balancing of classes). The order of steps 4 and 5 might change depending on if relevant tissue is segmented simultaneously as artifacts. We will cover each of these steps in the following subsections.
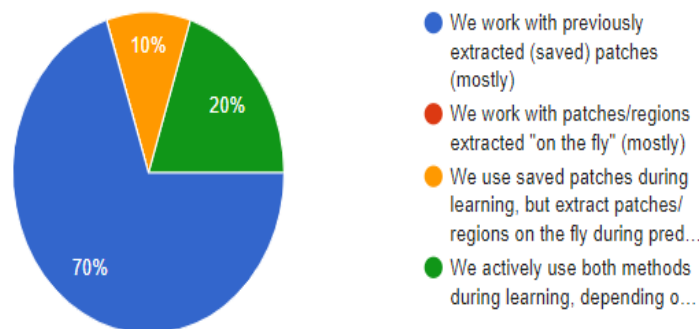
Figure 13: A survey response to questions on strategies on when to patch.
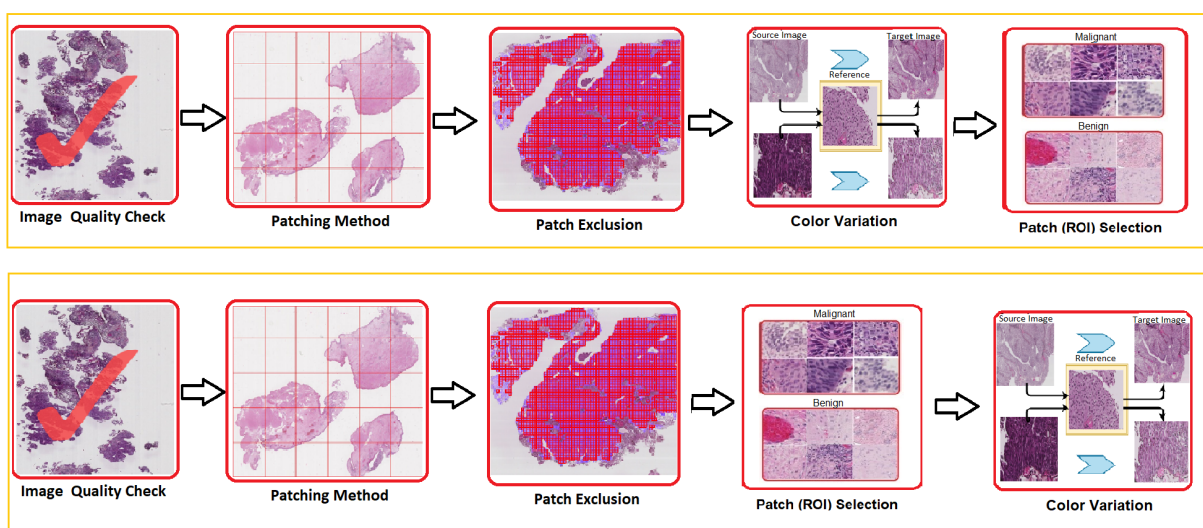
Figure 14: Propositions of two possible five-stage pipelines to process WSI.

3.1 Image Quality Assessment

A significant challenge in DP is the presence of artifacts and abnormalities in the dataset. In smaller datasets, researchers can go through all the data to find any inappropriate samples. This task becomes more demanding and labor-intensive when the dataset is large and the image requires significant attention at different zoom levels. In addition, describing the image quality could be subjective to intra and inter-reader variability [23] [24]. Clinical research involves a manual process that is error-prone and laborious with statistical process control. Automated approaches can help identify unintentional defects in slide preparation such as blurriness, tissue folding, stain mismatching, etc. [25]. A reproducible method can overturn the question of assessing the quality of a previously unseen slide during the prediction phase when manual intervention is unpractical.

It is possible to identify inappropriate WSI slides in duplicates, predetermined file format, resolution, size, and magnification level primarily based on metadata. Digital slides failing to pass the quality check will be excluded from training datasets and the prediction of previously unseen slides. In the current practice, manual assessment is the most common practice used by 87.5% of the participating CLARIFY researchers, as seen in Figure 15. Apart from that, many respondents considered image resolution a critical factor, while automatic artifacts detection is less common.
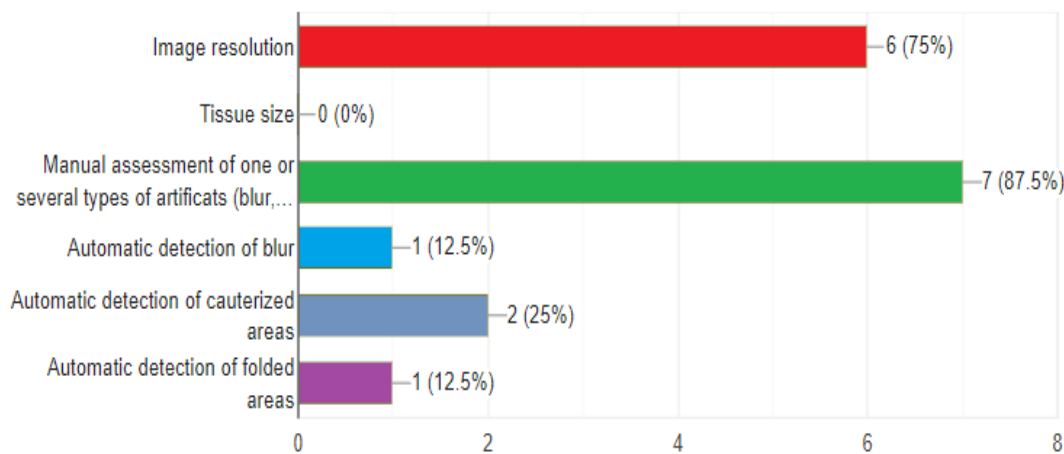
Figure 15: Practices of quality assessment of WSI in CLARIFY research teams.

It would also be desirable to assess the *quality* of the image after checking the technical metadata parameter. The quality might be measured based on contrast, colors, degree of blur. Due to the size of the WSI, straightforward measures for quality might be too computationally expensive to calculate on the entire image. Some measures could be based on global processing on the lower resolution images, and others will have to be done on patches or regions of the image. There might be large parts of the WSI that are diagnostically relevant even if some parts of the WSI have folds, blur, or other artifacts [26]. Therefore, we do not propose removing a WSI based on detected artifacts; instead, we propose removing patches at the patch exclusion stage.

When a batch of WSI is assessed at once, preparing a dataset, it is possible to detect outliers in the dataset to potentially exclude these. HistoQC [24] [27] is an open-source tool for performing rapid quality control over a batch. It helps to delineate and identify batch-level outliers such as stains obscurity in the dataset. It enables users to employ various metrics such as color histogram, edge detector, and pen detectors. This process results in numerical values for each filter which can be used for thresholding for different parameters like improper reagents, poor microtomy and tissue placement [14]. We have performed a test on all three datasets from SUH, EMC and INCLIVA. Table 6 shows the deployed configuration for these experiments.

Module name	Targeted artifact or metric
Basic Module	WSI magnification, file pyramid levels, and microns per pixel
Light Dark Module	Identify tissue location and folded tissue
Classification Module	Identify pen marking, coverslip, and cracks
Bubble Module	Demarcate contours of air bubbles on WSI
Bright Contrast Module	Overall and per channel tissue brightness, indicating stain/scan variations
Blur Detection Module	Identifies out of focus WSI regions

Table 6: Testing modules used in quality control tool HistoQC

Figure 16 shows a modular and user-friendly view of the HistoQC tool with an output of image masks. HistoQC successfully identified pen markings and tissue debris scattered around. It requires more parametric selection for blurry regions and air bubbles in slides, especially for stromal-rich regions. The tool may not distinguish patterns among cohorts if all slides have similar artifacts, such as a dirty slide. At the same time, it perfectly gives clues about heterogeneity in digital slides. Figures 17, 18 & 19 show experimental outputs for SML, NMBC respectively. The quality assessed images will be used for the patching module, which is described in section 3.2.
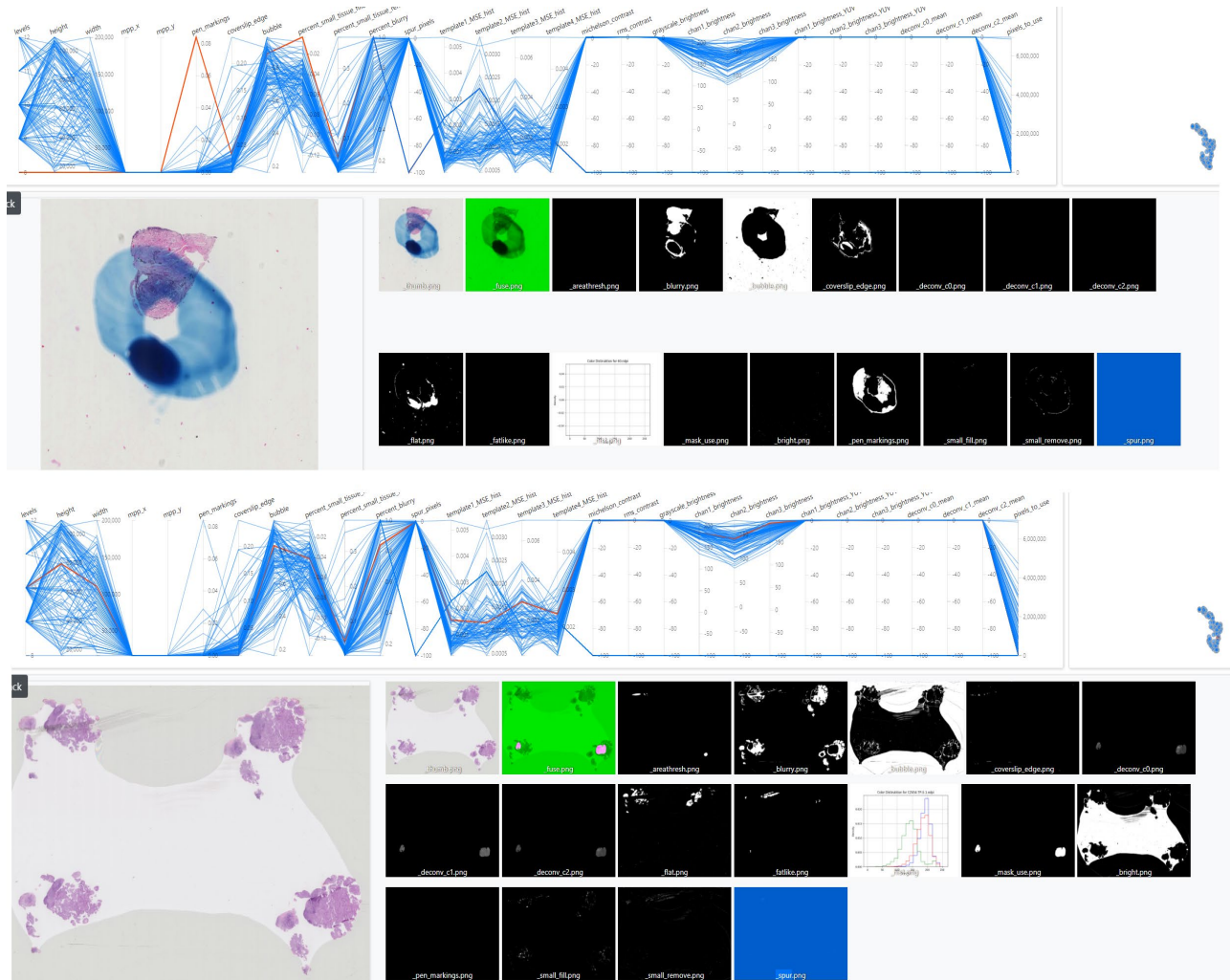
Figure 16: HistoQC output for selected modules in configuration; image shows different filter outcomes as binary masks to identify usability of selected thresholds on a particular dataset. Line plots show outliers for different image operations. Vertical lines show output values for every WSI in the dataset and a red line showing the output of varying WSI concerning others in the batch. WSI can be sorted based on a particular metric such as contrast, resolution or blur ratio, etc. WSI are part of the NMBC dataset provided by EMC.

Figure 17: The output of the HistoQC tool for the INCLIVA dataset. The euclidean plot helps in identifying outliers with a variation in the dataset. WSI can be sort by "flat_areas" to find higher tissue presence on the slide. Pen markings can be quickly evaluated by sorting with higher values or with "tenenGrad" to find faded slides.
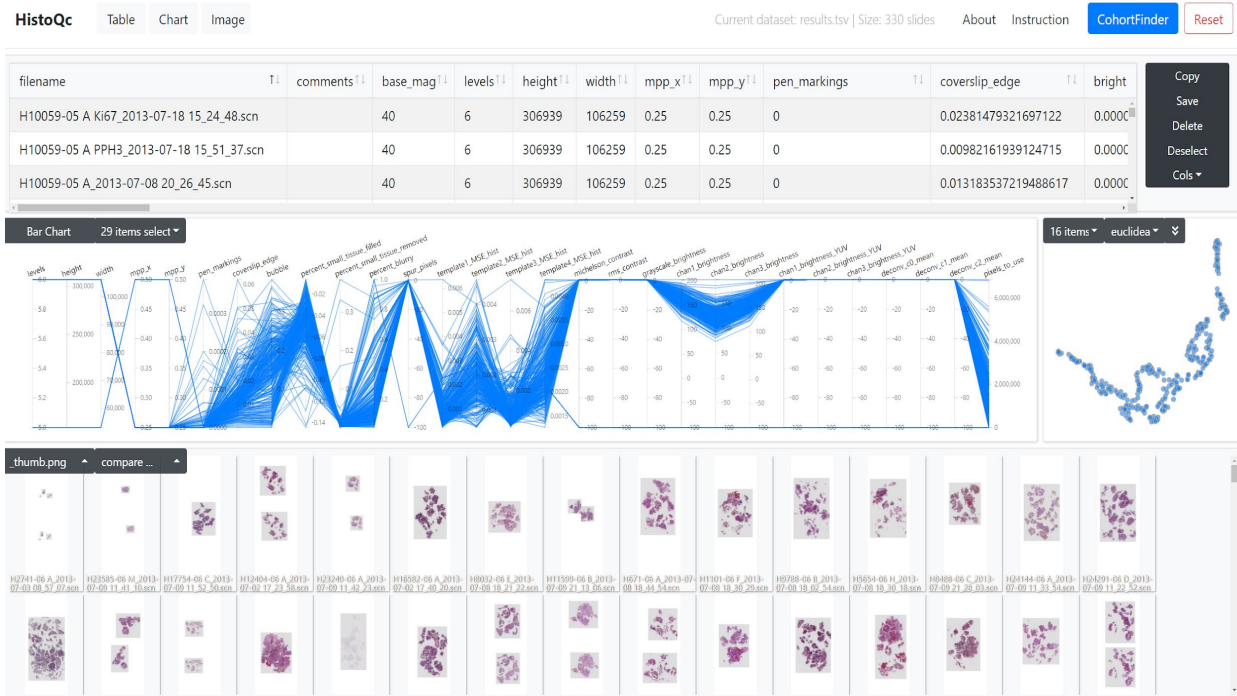
Figure 18: The output of the HistoQC tool for the SUH dataset. Dataset can be sorted with different metrics to find relevant slides for artifacts such as "Percent of small tissue removed" finds faded stains.
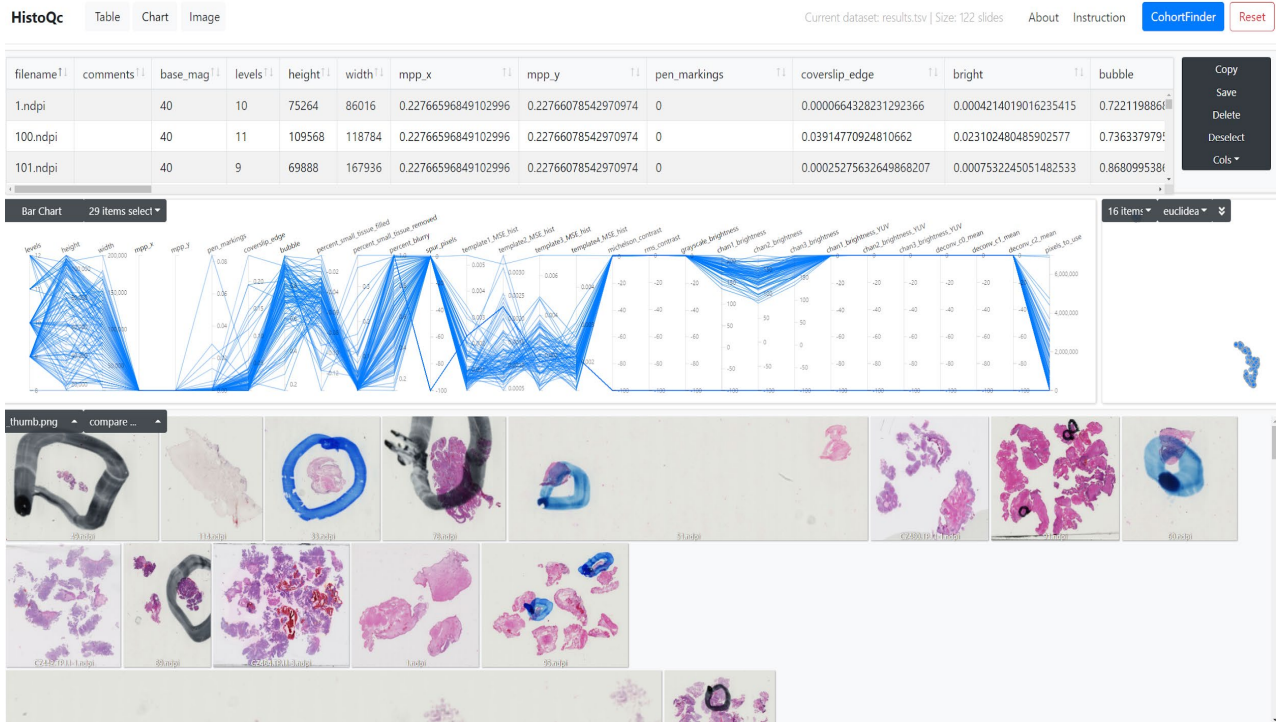
Figure 19: The output of HistoQC tool for EMC dataset. "Bubble" field can sort WSI by the bubble present in the background. WSI that are containing pen marking results in a positive number. WSI with a particular change in the cohort can be easily detected by field in a vertical line chart and sorting it to pick a particular image.

3.2 Patching Methods

WSI contains a lot of information and has a high potential for automated analysis [28]. However, feeding the whole slide image to a neural network all at once is a memory-consuming and inefficient task. It is also usually uncommon to have extensive processing resources available. Some prevalent practices comprise splitting bigger images into smaller images, called patches [29]–[31]. These patches can be extracted with many different strategies and can be attached with related labels from annotations.

The selection of patch sizes is dependent on many factors for the supervised task. Often used sizes are square patches of 64*64, 128*128, 256*256, or 512*512 and may range higher for low dimensional WSI. One reason to choose smaller patch sizes can be increasing the size of the dataset. Smaller patches are suitable when memory resources are limited and the model takes longer to train. Opposite to more extensive patches, we have an overall smaller dataset suitable to more considerable computational resources [32].

These patches can be extracted at different magnification levels to train multiscale models for segmentation or classification tasks [33] [34]. Another reason to choose different patch sizes is the required size when using transfer learning to use patches with an existing DNN model. Furthermore, patching can be considered the field-of-view and detail level desired for a specific task.

The CLARIFY survey results show that most researchers (45.5%) consider all the patch sizes mentioned earlier suitable for feeding an AI model. The other answers show that the used patch sizes are quite diverse: 27.3% consider the patch size 512x512 suitable, 9.1% 224x224, and 9.1% mention the patch sizes 256x256 and 128x128. Figure 20 shows the response to the question.
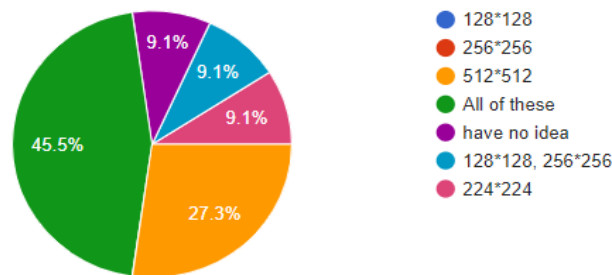
Figure 20: An outcome of internal survey for patch sizes.

3.2.1 Grid Patching

There are practical limitations for obtaining detailed annotations. Finding the ground truth label for every patch is a time-consuming job. Usually, pathologist performs rough labeling on patient-level or slide-level. In a clinical study, biopsies from a single patient may show multiple cancerous WSI, which can be used for unsupervised or weakly supervised tasks [35]. This lower graduality of labels leaves a choice to patching complete WSI by grid patching and image filters to automatically localize relevant patches. It may result in redundant background patches that may require storage for non-informative patches. Various operations can be employed to identify tissue and other regions based on Otsu thresholding or hysteresis. A foreground mask can be used to discard the patches only, including the background. A common practice in grid patching includes square patches that sometimes overlap when augmenting training dataset size. The patches can be pre-stored and later reused or extracted during the training. Figures 21 and 22 demonstrate the method.
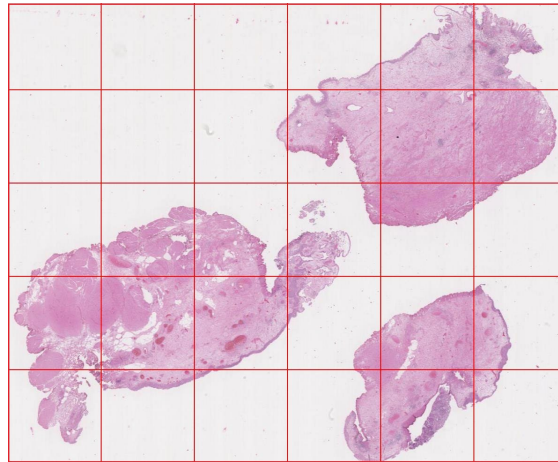
Figure 21: Demonstration of grid patching over WSI (Example using NMBC sample from EMC)
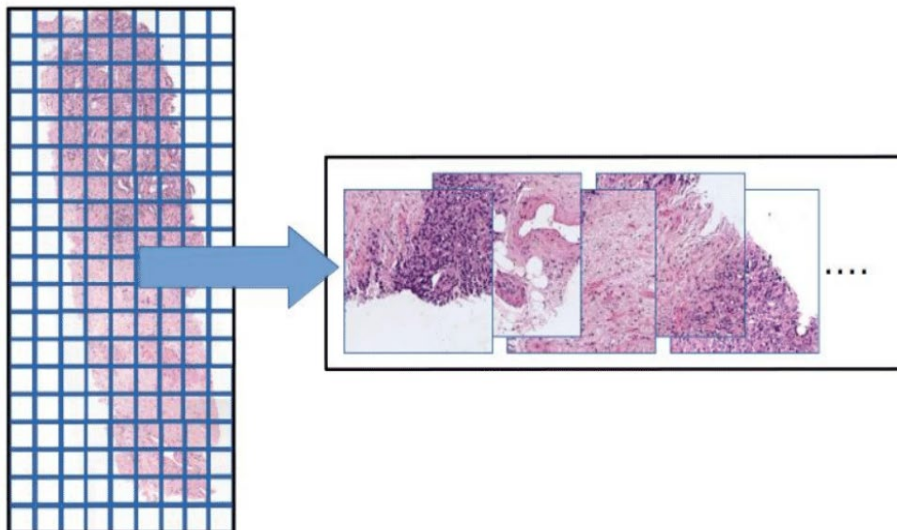
Figure 22: Outcome of grid patches where we have non-overlapping patches for every part of WSI regardless of pixel-level annotation.

3.2.2 Guided Patch Extraction based on Expert Annotations

Many histopathological tasks take place at a lower level to estimate cellular features such as lymphocytes. Pixel-level annotation requires extensive field expertise for every cancer type [4] [36]. This approach can help in developing patch-based supervised models by extracting predefined size patches only from the ROI. This section describes what-to-patch efficiently differs from selecting patches for subsampling from a pool of data defined in section 3.5 of this document. The drawback here is that a guided-patching strategy can lead to a class imbalance problem, but random sampling may even induce a much higher-class imbalance issue. All extracted patches using this method are non-overlapping and have class labels based on expert annotations. This step may reduce complexity by outputting fewer patches than the one in grid patching.

For example, in the case of NMBC disease, a nuclei segmentation model will require urothelium patches in its training set compared to other tissue types since they contain most cellular figures. Wetteland et al. [37] used a guided patch strategy to create a smaller patch dataset based on the expert annotation of five tissue types. The dataset was later used in training segmentation models to identify six different classes of tissue in NMBC. This step helped directly in avoiding patch selection by going through patches required by the task.
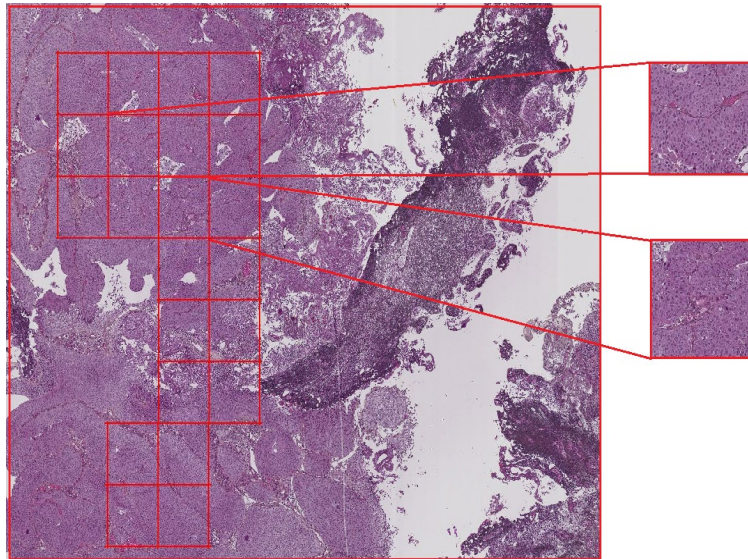
Figure 23: A demonstration of a guided patching strategy based on pixel-level annotation.

3.2.3 Patch Sizes

Patch size is a critical component in automated tasks. A task-based model-building paradigm can be explained combined with patch sizes and regions of interest in different cancer types [38]. Segmentation tasks delineate the boundaries of primitives (e.g., tubules, duct, urothelium in NMBC) to outline morphological features. In contrast, detection tasks typically identify the center of primitives (e.g., mitosis detection, lymphocyte detection). Classification tasks look for specific features of primitives (e.g., nuclei, lymphocyte) [39]. Segmentation is typically considered more challenging than detection in cases where there are multiple small regions [40].

Nuclei segmentation aims to find pleomorphism, a critical factor in some clinical cancer grading schemes [41]. It looks for variability in the shape and size of cells. Non-nuclei regions such as stroma will absorb hematoxylin and can easily be found using color deconvolution schemes. Patches at higher resolution must be significantly bigger than nuclei, but patches as small as 128*128 can be suitable for training a network [42]. Figure 24 shows the process of creating patches and masks for nuclei segmentation.
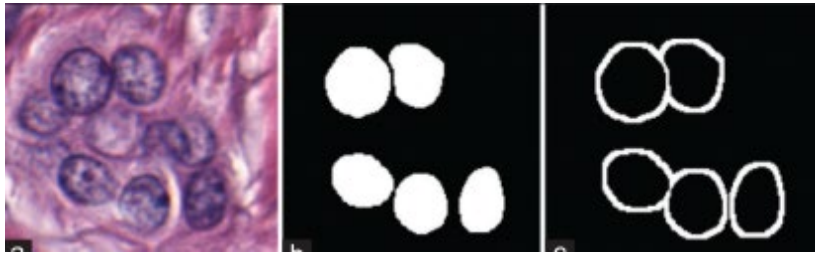
Figure 24: Input image patch size 128*128 with a few annotated nuclei and dilated mask. It was reprinted from [22] under Creative Common Attribution-Noncommercial-Share Alike 3.0 License.

Tubule Segmentation helps in estimating critical areas in the BCa grading scheme by scoring its aggressiveness in breast cancer [43]. Tubules have a complex structure, unlike nuclear morphology. Similarly, epithelium segmentation becomes critical in the prediction spread of disease in the epithelium-stroma joint of tissue. Here, patching aims to reduce uninteresting regions from the image to provide crisp boundaries to the model. Input patches of 64*64 or more can be considered appropriate for this task. Figure 25 and 26 shows examples of epithelium segmentation and tubule segmentation respectively.
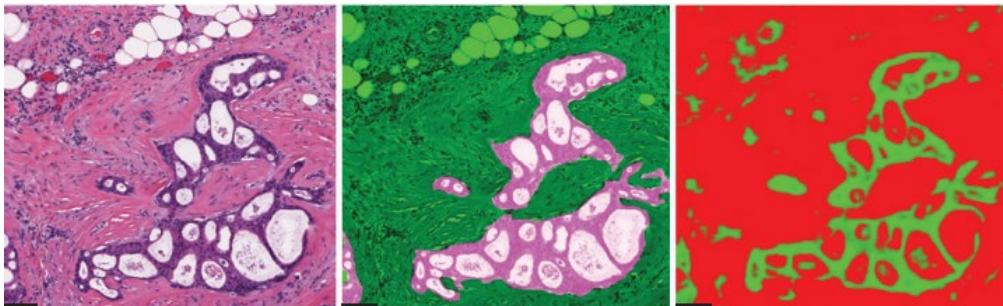
Figure 25: Input patch with a size of 256*256 for epithelium segmentation with annotations; Green region shows epithelium, and red shows stroma. It is reprinted from [22] under Creative Common Attribution-Noncommercial-Share Alike 3.0 License.
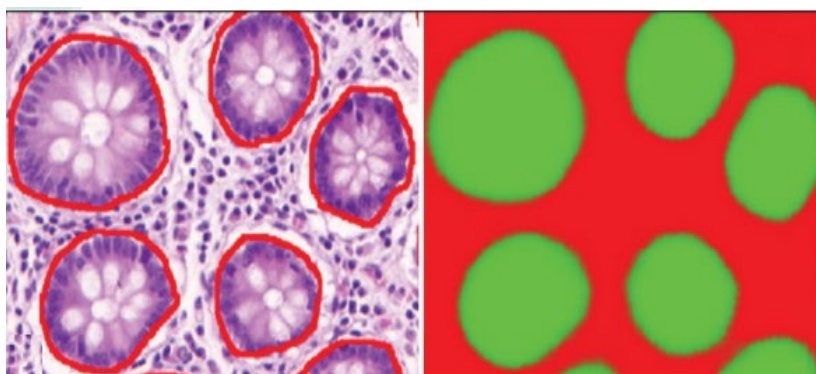
Figure 26: Input patch with the size of 128*128 for tubule segmentation with pixel-level annotation for benign tubule in the red line. It is reprinted from [22] under Creative Common Attribution-Noncommercial-Share Alike 3.0 License.

A lymphocyte is a subtype of white blood cells, an integral part of the immune system [19]. Lymphocyte detection is performed to find infiltration to evaluate the strength of immune response against foreign bodies. Lymphocytes can be better tested with immunohistochemical staining, but they also absorb hematoxylin and appear blue-like nuclei [44]. Different staining such as CD25 and CD138 helps find lymphocyte subclass, which is hardly distinguished from a normal cell. Lymphocytes are chromatically darker and are approximately 10 pixels in diameter (~0.25um/pixel). A large patch size would reduce the discriminative ability of the classifier by introducing surrounding tissue in relevant features. Likewise, mitosis detection is an essential aspect in breast cancer grading where the mitotic figures help identify the rate of cell division process [45]. Patching at a magnification of x40 can capture lymphocytes and mitotic figures with smaller patch sizes. The figure 26 and 27 exhibits lymphocytes and mitotic figure respectively.
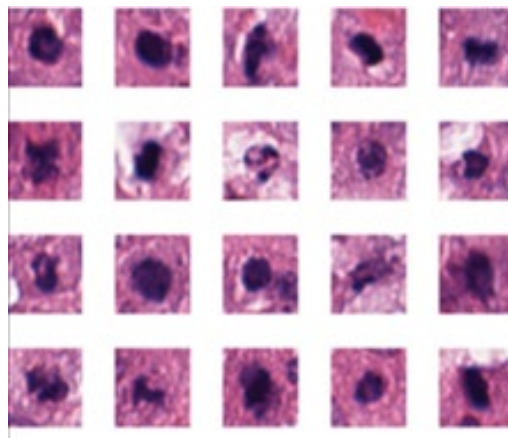
Figure 27: Input patches of 32*32 for lymphocyte. It is reprinted from [22] under Creative Common Attribution-Noncommercial-Share Alike 3.0 License.
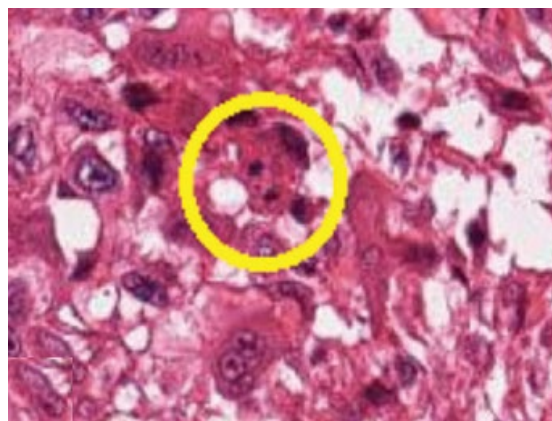
Figure 28: Input patch of 64*64 with mitosis in telophase stage. The yellow circle shows a more diminutive size where the blue ratio segmentation approach is applied. It is reprinted from [22] under Creative Common Attribution-Noncommercial-Share Alike 3.0 License.

3.2.4 Multiresolution Patches

Different researchers [6] [37] [46] have shown that it can be beneficial to use multiresolution models in computational pathology. It will somewhat correspond to the multiresolution information a pathologist would use when a tissue section is assessed by zooming in to get cellular details and zooming out to get the morphological tissue context. The multiresolution approach can feed an AI model with patches centered around the corresponding point over different resolutions. The different resolutions can be combined with an early fusion or a late fusion approach. Figure 29 shows tiles of each class at different magnification levels.
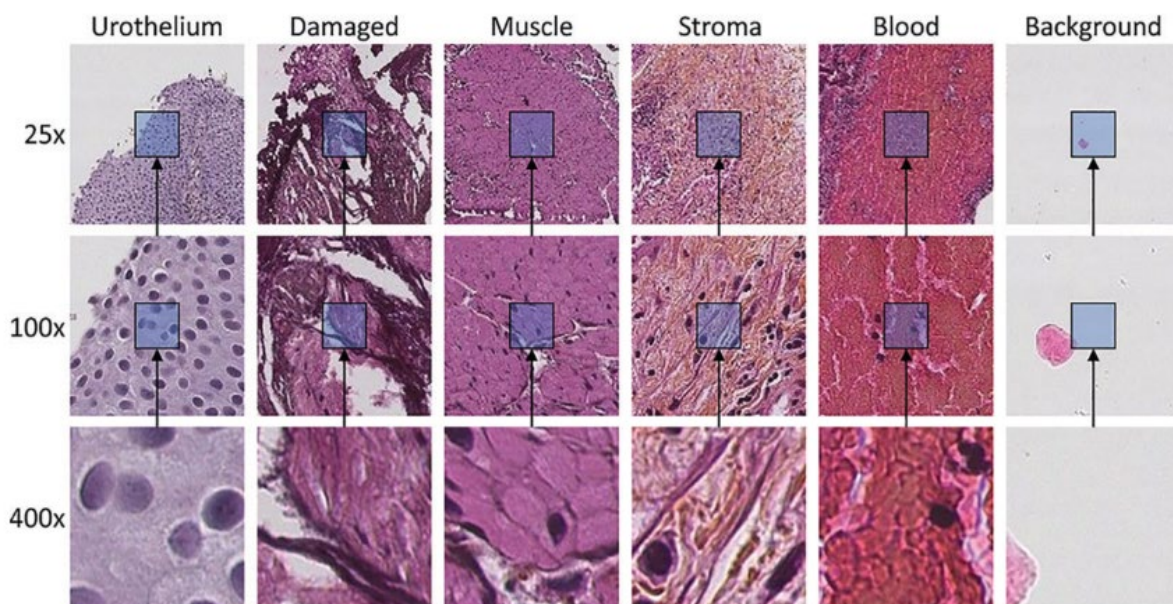
Figure 29: Multiresolution patches for NMBC from SUH. The figure shows 128*128 patches at different magnification levels. A demonstration of six tissue classes in a WSI. It is reprinted from [34] under Creative Commons Attribution 4.0 License.

3.3 Patch Exclusion

The preparation of digital H&E slides is highly dependent on the competence of the histotechnologist [13][47]. Many slides come with artifacts that can introduce unnecessary patterns in ML models and reduce the prediction accuracy. There are various artifacts in our datasets due to the condition of tissue and hardware, such as folded tissue, focal blur, or air bubbles. It is vital to identify patches that should be discarded based on blurry regions, blood sections, burnt tissue, and folded tissue. If it is desired to stitch patches back into a larger image, to be used with a patch on-the-fly strategy, excluded patches can be replaced by a whole black or white patch. These patches are not going to be used further in preprocessing pipeline. Pathologists do not use them for diagnostic purposes in clinical practice. Patch exclusion is useful both for global level tasks like grading at slide level as well as local tasks such as tissue classification or localizing regions of worst grade etc. Examples of some common artifacts are shown in Figure 30.
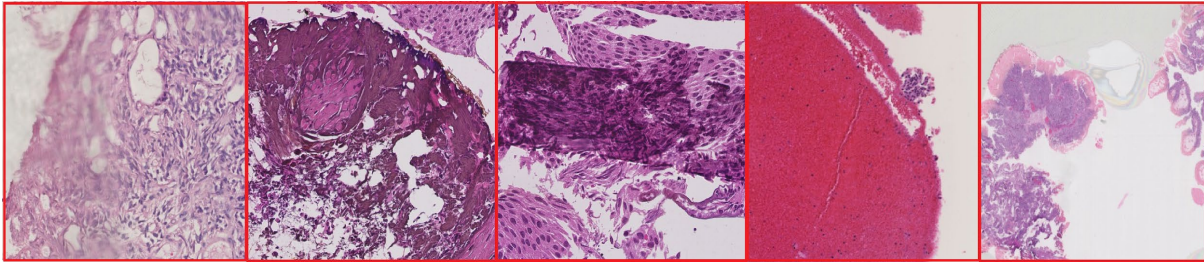
Figure 30: A few examples of patches that can be excluded. Blur tissue, burnt tissue, folded tissue, blood, and air bubble from left to right.

The survey shows that various attributes are used to exclude patches during preprocessing, see Figure 31. While 44.4% of the researchers pay attention to all the mentioned attributes (blur, burnt tissue, folded tissue, air bubble), blur is considered the most crucial anomaly among other patch artifacts.
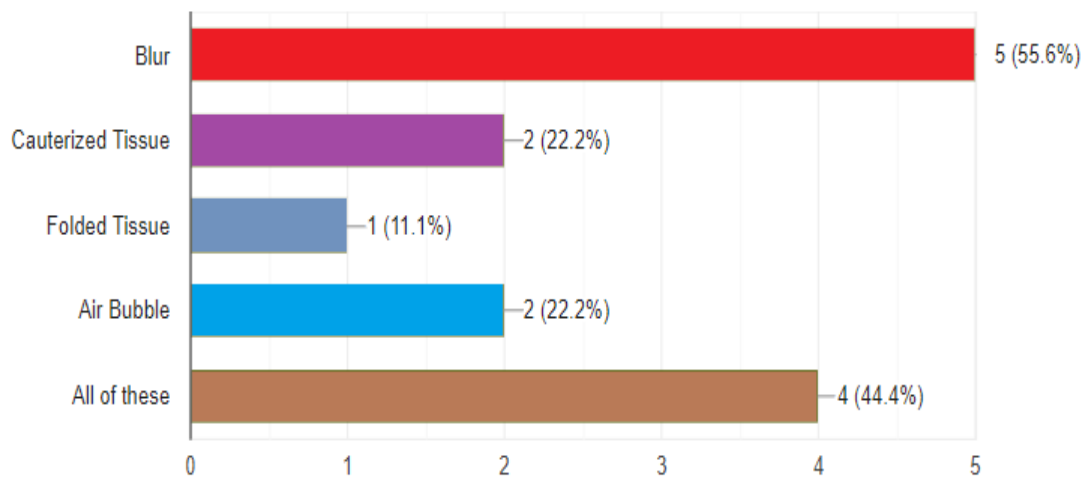
Figure 31: A response in the survey for standard practices in patch exclusion

We will discuss the following topics based on assessment for patch exclusion:

1) Foreground segmentation in large WSI (discarding background areas).
2) Detection of blurred regions.
3) Assessment of folded regions.
4) Detection of air bubbles.
5) Detection of blood and burnt areas.

3.3.1 Foreground/Background Segmentation

All WSI includes large background areas, and it is necessary to mask out the tissue areas from the background before further processing. To simplify the task, it is possible to utilize a low-resolution version of the WSI to detect the foreground mask; this can later be interpolated to be used with the full resolution image. Some methods of identifying background include HSM (Hue, Saturation, Mixture) color space, which was defined to understand the color distribution based on the human visual system [48]. Magenta channels intersect RGB (Red, Green, Blue) channels at a weight level which help in performing binary decomposition mask [49]. Optical density (OD) space, on the other hand, estimates corresponding stains to make a linear combination.

DP datasets sometimes contain WSI with a tiny tissue sample. These slides do not qualify for specific tasks where pixel annotations are impossible due to the absence of primitive areas. It could be subjective to include patches from these slides as a part of a negative class. The size of tissue on the slide also varies for different cancer types due to organic biopsy. Figure 32 shows an example of a minimal tissue on the slide compared to other images in the dataset. Moreover, Figure 33 shows an alternative example where image size was significant but underlying tissue was not sufficiently diagnosable.
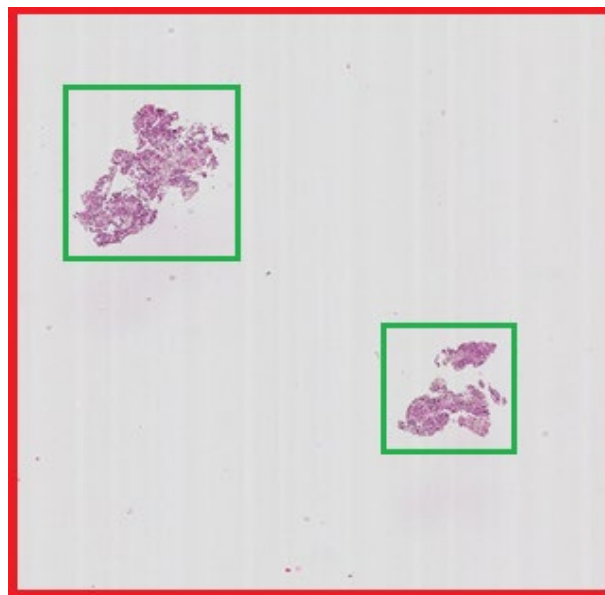
Figure 32: WSI for NMBC from the SUH database has less tissue placed than other slides.
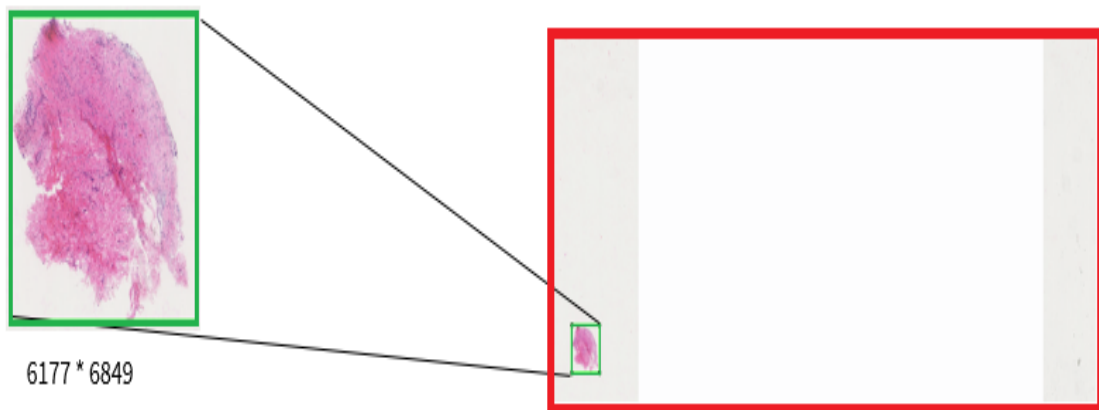
Figure 33: WSI for NMBC from EMC database with a small tissue sample where image has higher scanning resolution.

Entropy is one of the simple mathematical tools that can be used to measure average information in a random variable. This concept can be applied to find background and foreground ratios in a slide from its histogram. The discrete entropy of a histogram can measure the amount of tissue area over the scanned slide by measuring average surprise between bins. Assuming x is a random variable with probability density function f with support of a set of n, then differential entropy of a histogram can be described as [50]:

$$H(X) = -\sum_{i=1}^{n} f(x_i) \, log \left(\frac{f(x_i)}{w(x_i)}\right)$$

Histogram estimate can be interpreted like the maximum-likelihood estimate of discretized distribution. This approach is quick and straightforward, and biased towards bins with higher weights. Other multidimensional approaches such as gaussian mixture models [51] and expectation maximization (EM) [52] algorithm can provide better estimates. WSI with abnormal structure had a lower entropy value compared to other digital slides with higher tissue spread. The calculation was faster on coarse resolution. A few experiments were conducted to test the validity of this method and are reported in Table 7. Here we can observe that values for both WSI are smaller and fall in a similar range compared to other WSI with the thick and fair placement of tissue; also, computation at higher level results in smaller numbers than lower levels for the same image.

Figure #	Resolution	Level	Differential Entropy Value	Computational Time
Figure 32	306939*106259	0	3.34	0.01s
Figure 32	308*107	5	2.16	0.011s
Figure 33	51456*126976	0	4.72	0.007s
Figure 33	201*496	9	3.2	0.01s

Table 7: Experimental results for differential entropy calculation, for example, images.

Otsu threshold is another common computer vision technique that provides a single threshold value that separates foreground and background. This threshold value is calculated by the objective function of minimizing intra-class variance and maximizing inter-class variance. This method computes the histogram and probabilities of each intensity level; set up initial weight and mean; computes $\sigma_b^2(t)$ for all bins and updates weights and means. A maximum value is obtained as a threshold among all the bins. It can be defined as

$$\sigma_w^2(t) = w_0(t)\sigma_0^2(t) + w_1(t)\sigma_1^2(t)$$

$$w_0(t) = \sum_{i=0}^{t-1} p(i) \quad and \quad w_1(t) = \sum_{i=t}^{L-1} p(i)$$

Here weights w are probabilities from both classes (background and foreground) calculated from all L histogram bins. $\sigma_{0,1}^2(t)$ is the variance of both classes, which is evaluated using mean values. This method has limitations in forms of performance. It provides a reasonable threshold when histogram distribution is binomial. A significantly smaller object area compared to background drops bimodality and results in segmentation error.

A popular variation of this method is two-dimensional Otsu which yields better results in noisy images. Here intensity value of a pixel can be compared with the average intensity of the immediate neighbor (i,j). Let us divide each grey pixel into L levels and divide every occurrence with total pixels in Image N to find f_{ij} A two-dimensional Otsu can be developed on a two-dimensional histogram. So, weights can be redescribed in terms of joint probability as:

$$P_{ij} = \frac{f_{ij}}{N} \quad and \sum_{i=0}^{L-1}\sum_{j=0}^{L-1} P_{ij} = 1$$

$$w_0 = \sum_{i=0}^{s}\sum_{j=0}^{t} P_{ij} \quad and \; w_1 = \sum_{i=s+1}^{L-1}\sum_{j=t+1}^{L-1} P_{ij}$$

$$w_0 + w_1 = 1$$

$$u_0 = \left(u_{0i}, u_{0j}\right)^T = \left(\sum_{i=0}^{s}\sum_{j=0}^{t}\frac{ip_{ij}}{w_0}, \sum_{i=0}^{s}\sum_{j=0}^{t}\frac{jp_{ij}}{w_0}\right)$$

$$u_1 = \left(u_{1i}, u_{1j}\right)^T = \left(\sum_{i=s+1}^{L-1}\sum_{j=t+1}^{L-1}\frac{ip_{ij}}{w_1}, \sum_{i=0}^{s}\sum_{j=0}^{t}\frac{jp_{ij}}{w_1}\right)$$

$$u_T = w_0 u_0 + w_1 u_1$$

The optimal threshold is obtained by calculating the trace $tr(\alpha_B)$ of the resultant inter-class discrete matrix where optimal threshold pair $(s*, t*)$ is defined by maximizing class variance in trace [53].

$$tf(\alpha_B) = tr\left(\sum_{k=0}^{1} w_k[(u_k - u_T)(u_k - u_T)^T]\right)$$

$$(s*, t*) = \underset{(0 < s < L, 0 < t < L)}{argmax}\; tf(\alpha_B)$$

A pseudo-algorithm can be written as;

Pseudoalgorithm for 2-d Otsu [taken from [43]]

```
Max s,t = 0;
for ss: 0 to L-1 do
    for tt: 0 to L-1 do
        evaluate tr(S_b);
        if tr(S_b) > max
            max = tr(S,b);
            s = ss;
            t = tt;
        end if
    end for
end for
return s,t;
```

Table 8 shows the result of one-dimensional Otsu implementation. The method provides higher threshold values for WSI with the smaller tissue sample. We can see that figure 32 has a more reasonable outcome than figure 33. The first figure results take a significant time at higher resolution due to higher weights in both classes. The latter figure has lower computational time because class imbalance rises for background and foreground. It is due to the smaller objectivity of WSI based on a resolution that makes bins unequal.

Figure#	Resolution	Level	Otsu Threshold Value	Computational Time
Figure 32	306939*106259	0	254	689s
Figure 32	308*107	5	178	0.01s
Figure 33	51456*126976	0	254	0.05s
Figure 33	201*496	9	254	0.07s

Table 8: Experimental results for Otsu thresholding for selected examples.

Several other approaches to perform tissue segmentation for the efficient diagnosis, such as the hysteresis thresholding strategy. A Gaussian filter can be used to estimate noise combined with global thresholding methods to separate the background. Some researchers have implemented fixed thresholding methods by combing image color transformation to Hue-Saturation-Value (HSV) and OD color space.

3.3.2 Detection of Blurred Regions

Uneven tissue sample on a slide is hard to avoid owing to its manual cutting nature. It results in a non-uniform focal plane for microscopy. Image focus is of utmost importance for a pathologist to observe the clinical criticality of the patient. These slide imperfections are hard to find at a coarse level in gigapixel images. Image blur is the result of non-attentive high spatial frequencies. These approaches can be categorized in several ways: no-reference, reduced-reference, and fully reference methods. Fully and partially referenced approaches require a reference template to compare image smoothing filters with a tolerance value [24]. No reference metrics are easier to use but have disadvantages in ignoring outer parts of the image [54].

Some classical approaches can be used as global methods to evaluate sharpness. Wavelet-based transforms calculate edge compositions and sharpness using Haar wavelet transform before and after blurring. A Haar metric may range from zero to one. Other local blur detection methods include the kurtosis measure, which characterizes the skewness of a distribution. Histogram features may help to understand sharpness or blur in these patches as well. A few other methods exist in the literature, such as pixel-level metrics by Wu et al. [54] or Laplace and Gaussian filters.

Power spectrum calculates the average power of frequencies in an image. Image blur attenuates these frequencies quickly, which can be identified with a gradient. Laplace operator finds divergence of a function that finds local maxima or slope of steepest ascent. An experiment was developed on the idea of shape-from-focus [55], which combined Laplace and Gaussian to develop a mask to identify blur areas. The size of the Laplace kernel plays an important role here. Figure 34 shows a blurry slide which was significantly detected as shown in the mask. On the other hand, a significantly smaller blur region in a bigger WSI may not be well detected with a smaller kernel.
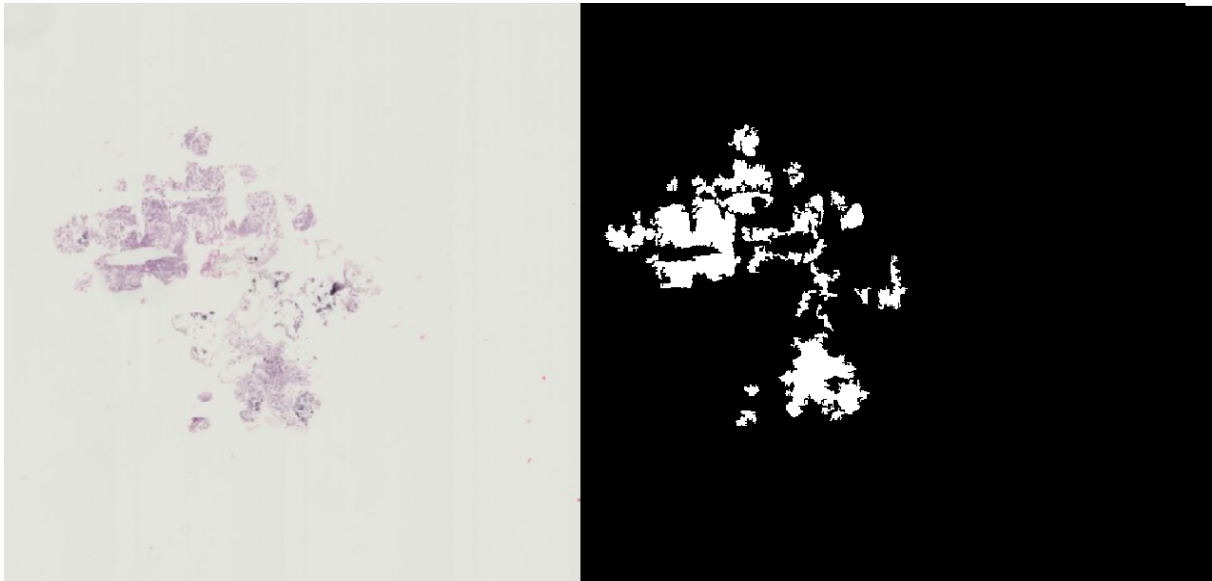
Figure 34: The output of a Laplace filter for detection of blur in a mostly blurry WSI.

3.3.3 Detection of Folded Regions

Some artifacts are an interplay between hardware and tissue placement on slides. Wrinkles and folds appear when paraffin sections are not stretched evenly or cut with a blunt microtome [47]. Finding folds can help eliminate them from being part of training datasets since they have saturated features. Color spaces provide a better abstractive model to look at color in different ways. Image enhancement methods facilitate the visualization of folded areas by transforming the image from RGB to HSV [56]. Thresholds for colorimetric features may vary in different sliding conditions. It is hard to establish a single threshold for all H&E slides. Otsu thresholding is sometimes implemented to get the intensities of light and dark tissues separately. It does not work well where tissue has higher thickness leaving a perceptual image of darker stain [57].

Figure 35 shows the outcome from Otsu thresholding, where results may compromise in case of very thick staining. Figure 36 shows the output of the image where the surrounding folded area is also identified as fold due to high saturation. Color space transformation works quite well at high magnification. Several studies have proposed training a classifier to find luminance and saturation using vector representation [58]. The straight lines of the folds might be detected with line detectors and combined with saturation and value channel thresholding.
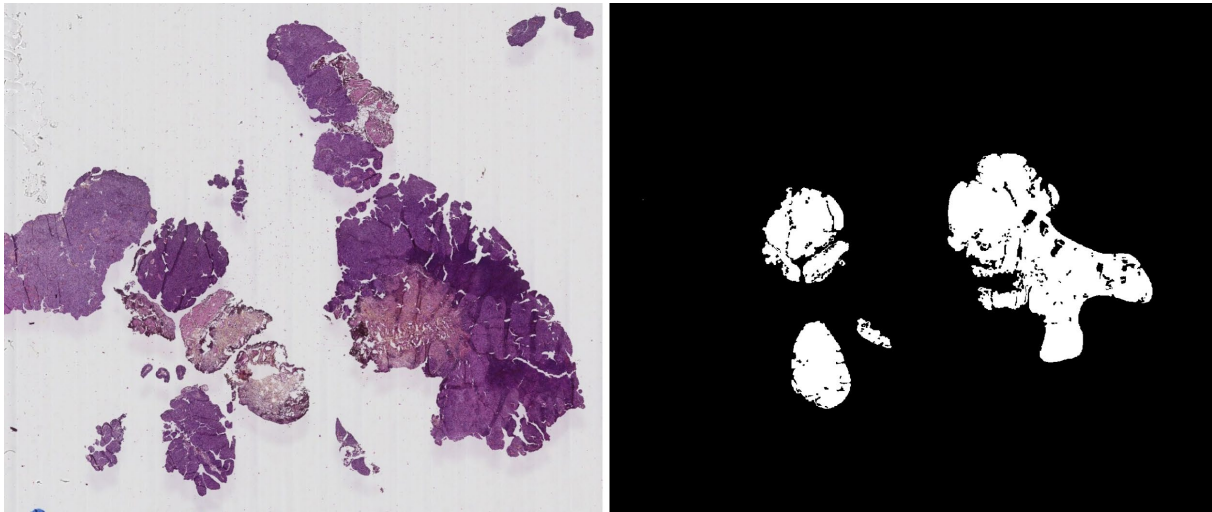
Figure 35: A WSI with the folded area and higher hematoxylin concentration (left). Image mask for folded areas with Otsu thresholding (right).
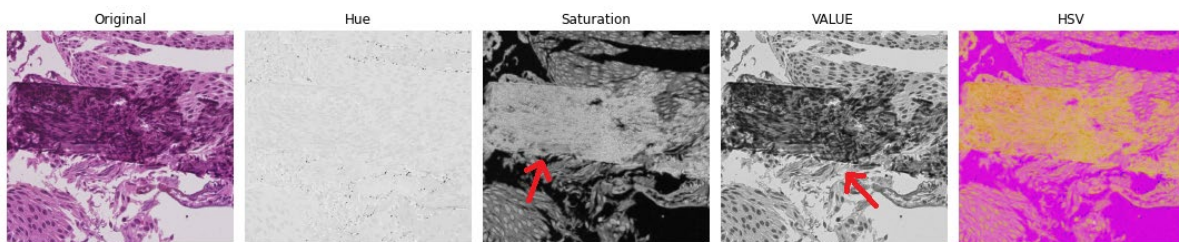
Figure 36: Tissue fold patch and its color-space transformation from RGB to HSV for thresholding. HSV image is plotted as RGB for visualization of the difference. Thresholding can be used on the value or saturation since they represent staining density apparent than the hue channel.

3.3.4 Detection of Air Bubbles

An air bubble is formed in a digital slide under the coverslip when thin tissue dries without proper mounting. This effect is reduced sometimes when the entrapped bubble collapses and leaves cracked areas [47]. This artifact can be identified by transforming the image into several color spaces. Frangi filters are commonly used in diabetic retinopathy for vessel detection but can also be utilized for this purpose. This filter helps to find tissue spurs and straight edges. Air bubbles in digital slides are detected by establishing a base experiment, as shown in figure 37.

A histogram can compute the masked output to find the blur ratio in the patch, as shown in figure 38. Color transformation also works at some degrees, as shown in figure 39, where Value and Hue channels separate bubble and background, which can be a good reference for thresholding. Our dataset contains few slides with air bubbles in the middle of the slide in a tube shape. Most of the air bubbles are present in the outer region of the slide. The research includes training SVM classifiers to evaluate quality scores based on the Pearson correlation coefficient [25].
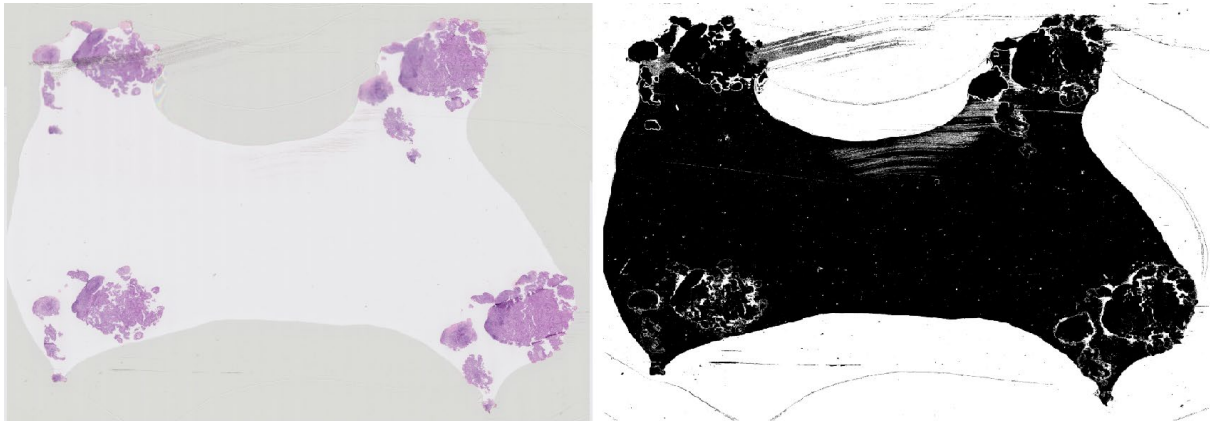
Figure 37: An example WSI surrounded with a significant bubble on the borders of slides (left). Image mask obtained using Frangi filer (right). It results in a suitable identification of the bubble with the white area.
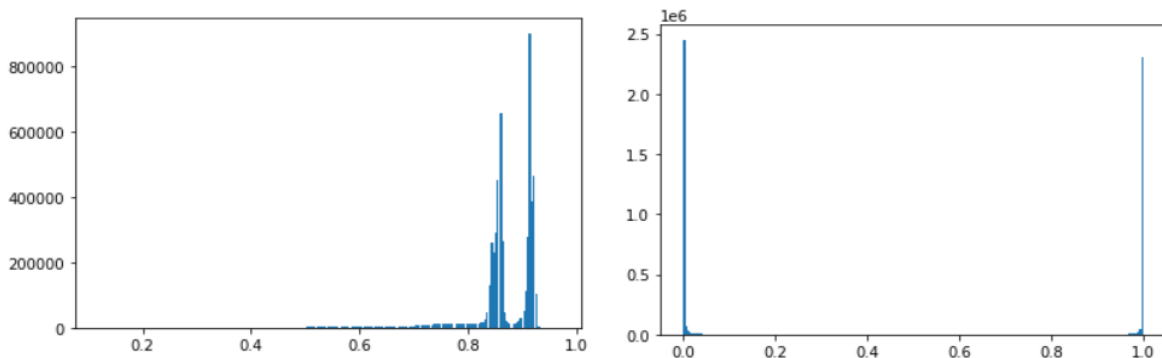
Figure 38: A histogram presentation of an image showing bimodal distribution for image and mask. Mask histogram can be used to find the percentage of a white and black pixel to calculate the bubble ratio.
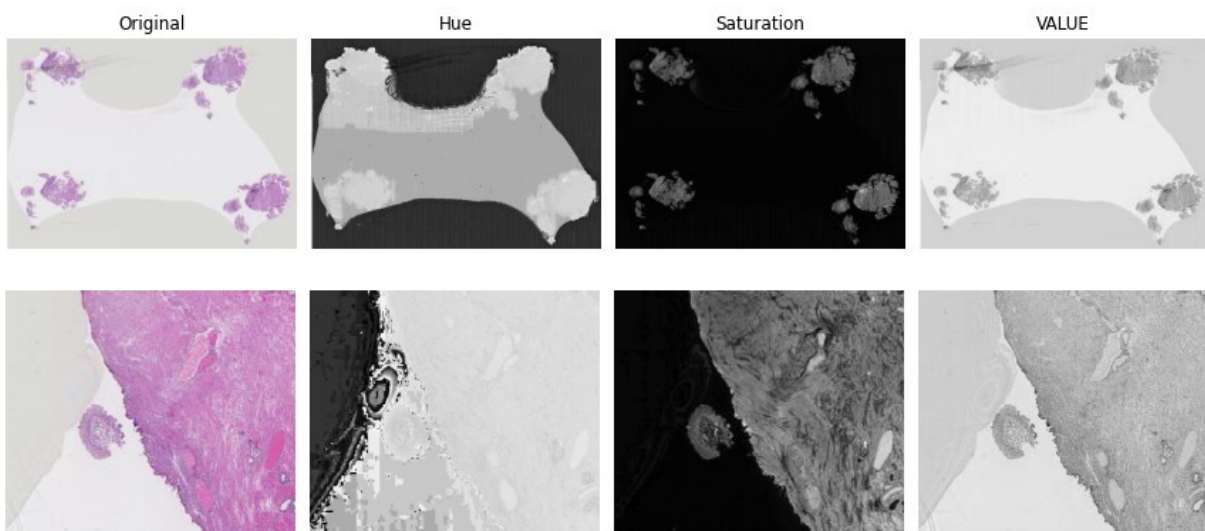
Figure 39: Two examples of air bubbles from the NMBC dataset. Conversion from RGB to HSV helps in predicting bubbles in thresholding from Value and Hue channels. It may be hard to distinguish background and bubble in the Saturation channel. The hue channel signifies the air bubble as a darker region compared to tissue or background

3.3.5 Detection of Blood and Burnt Areas

WSI are information-rich with improved objectivity, but this also comes with a cost of artifacts in the forms of noise [59]. Various types of noise are hard to detect and require a specific pipeline to find inconsistencies. Burnt tissues are the result of laser cauterization. These areas are darker and non-informative for diagnosis. Similarly, blood regions areas are almost non-informative regions for pathologists because they do not provide information related to cell shape and polarity.

It is possible to deconvolve the patches to find these areas, but the methods related to blood and burnt areas are not discussed much in literature [60]. Xiao et al. [61] have proposed intensity-based detection where a patch with a certain intensity level of one color can be flagged and excluded. This intensity threshold needs to be determined for every dataset based on the thickness of staining. A common constraint can be evaluation values from either RGB or transferred HSV. Figures 41 and 42 show how color transformation can help in creating thresholds. The function can be used to choose estimation using all channels, but this can be set relatively differently in blood and burnt tissue. For instance, blood samples at low resolution can easily be omitted by thresholding the Hue and Saturation channel, whereas, at fine resolution patches, the hue channel may prove to be less significant than saturation, as shown in figure 42. On the contrary, the Hue channel is irrelevant in burnt areas, but the value channel can find a steep gradient to discard patches.

$$E(x) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{h_{ij} + s_{ij} + v_{ij}}{3} \geq TH$$

Where N is patch size, h, s, v are hue, saturation, and value intensities [61].

A multiscale CNN-based classifier was trained to segment blood and damaged tissues from the NMBC database provided by SUH. A fixed patch size at different magnification scales was used to get a broader field of view. The method segments the entire WSI at patch resolution, classifying urothelium, stroma, muscle, damaged tissue, blood, and background at once. Figure 40 shows the output segmentation map for the research work of Wetteland et al. [37]. It demonstrated the ability of multiscale patches to extract regions which can also be used in patch exclusion.
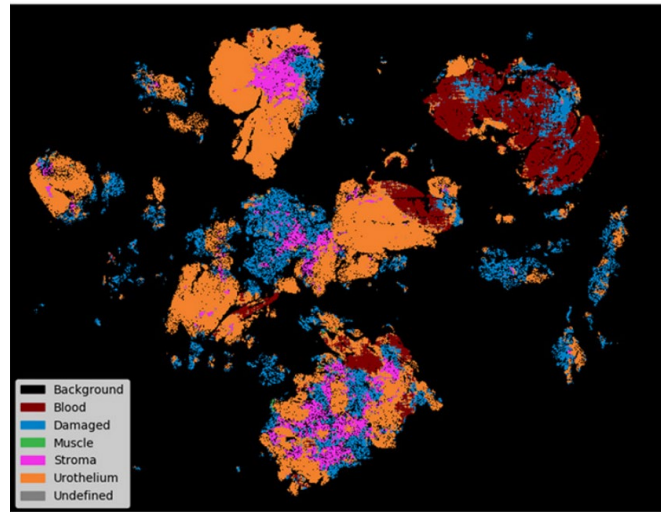
Figure 40: Segmentation map output from Wetteland et al. [37]. It was reprinted under the author's creative commons license.
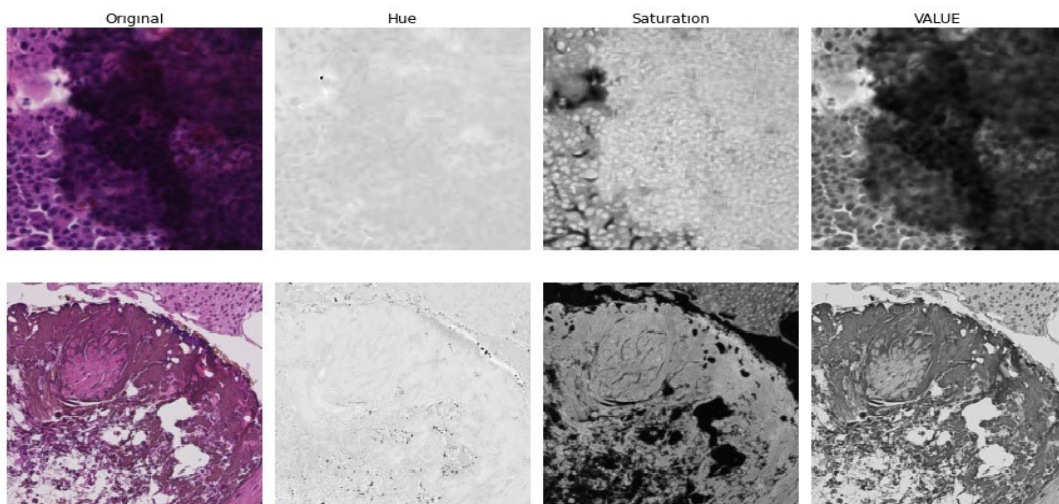
Figure 41: Color-space transformation applied on burnt and over-stained areas where cellular features are not observable.
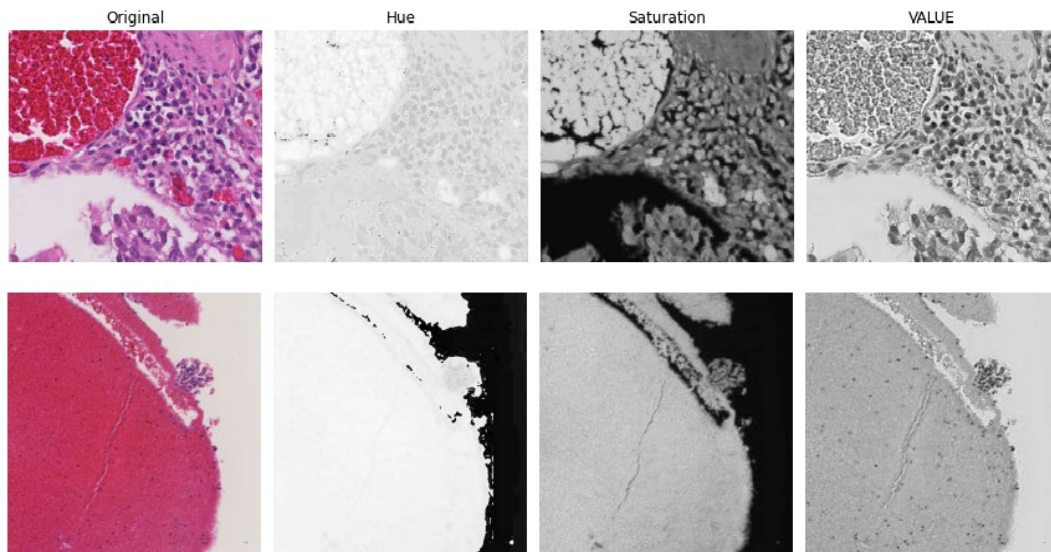
Figure 42: Color-space transformation applied on two blood image patches at 20x and 40x.

3.4 Staining and Color Variation

The variations in tissue and stain appearance complicate the quantitative tissue analysis [42]. Color variation has a significant influence on computer-aided diagnosis systems [17]. Therefore, dealing with color variation is an integral part of preprocessing to generate robust and generalized models, i.e., models that can work with data from different labs and scanners. Several approaches appear in the literature to deal with color variation. The most relevant ones are presented below:

i. Large training sets aim to diversify enough to cover all possible color variations for enabling robust models. However, this can be difficult to achieve in medical applications, where the datasets are limited due to the cost of acquisition.

ii. The use of blind color deconvolution techniques separates both H and E stains. This procedure isolates color information from the concentration of each stain. The concentrations can then be used to perform computer-aided diagnosis, reducing the impact of color variations. The separated values can also be used to perform color normalization.

iii. Color normalization matches the color properties of all the H&E slides, usually using a template image. The colors are modified to mimic the ones of the template image, while the structure of the image should not be modified.

iv. Color augmentation while training models is an alternative solution to develop stain-resistant classifiers. Color augmentation is a data augmentation method where values are randomly added to the image, such as saturation, hue, brightness, and contrast. It can approximate the large and diverse pieces of training etc., as in point i). Augmentation is done during training and does not have to be done during prediction.
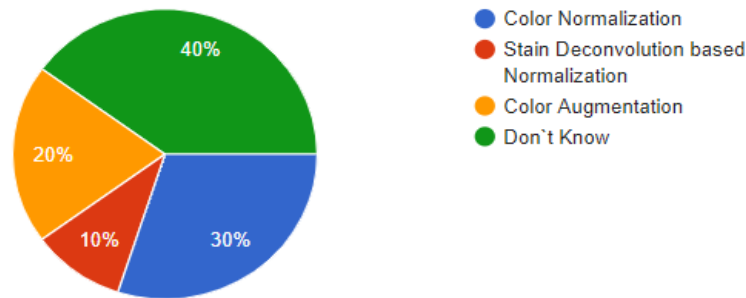
Figure 43: Pie diagram representing survey result for the preferred method to control color variability

In the conducted survey, we found that a large part of the CLARIFY researchers does not have a strong opinion about the most effective method to control color variability for their use-cases. Figure 43 shows that the researchers with a clear opinion on this are more inclined to favor color normalization (nearly 30%) over color augmentation (about 20%).

3.4.1 Stain Separation using color deconvolution

Considering the staining procedure, it is possible to separate both stains from the observed H&E image. Blind Color Deconvolution (BCD) estimates the representation of each dye in the RGB color channels (stain-color vectors) and the concentrations of the dye in each image pixel. CAD systems can then use the separation or perform color normalization. This approach stands out due to its potential histological structure preservation [33], but it is usually computationally heavy.

A wide range of solutions has been proposed to find the stain-color vector in the images. BCD methods usually apply the Beer-Lambert law [62] to transform slide images to the OD space. Let I be a RGB image represented as a 3xMN matrix where MN is the number of pixels. The observed OD image is obtained following.

$$y_c = -log\frac{i_c}{i_c^0}$$

Where i_c is the transmitted light on the color band c and i_c^0 the incident light (typically 255 for RGB images). Then the stained slide can be expressed as

$$Y = MC + N$$

Where Y (3xMN) is the observed OD image, M (3xns) is the color-vector matrix, C (nsxMN) is the concentration matrix, and N is a random noise matrix. The size of the matrices is determined by the number ns of stains in the image. H&E images use two stains, but many methods consider a third residual stain.

The initial color deconvolution was proposed by Ruifrok et al. [49] with a given color-vector matrix include Non-Negative Matrix Factorization (NMF) [63], Independent Component Analysis (ICA) [64] or Singular Value Decomposition (SVD) [16] (See [33] for a complete review). The CLARIFY members in the University of Granada (UGR) have presented several Bayesian approaches using prior distributions on the color vectors and stain concentrations. A similarity before the color vectors and a smoothness simultaneous autoregressive prior model on each stain concentration was used to achieve the deconvolution task [65]. This work was extended using a total variation before the paper [66] and sparse general super-Gaussian priors in research work [67]. Figures 44 and 45 represent the idea of deconvolution from two methods.
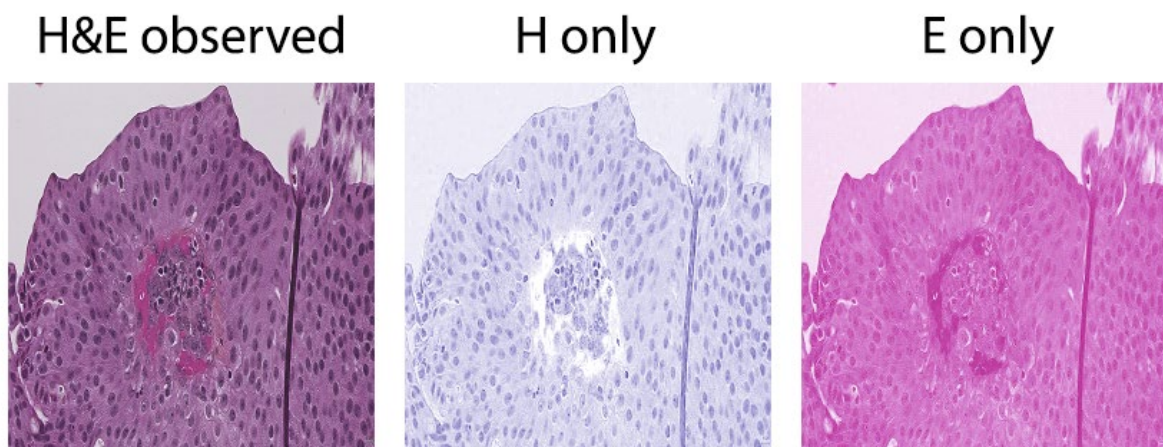
Figure 44: An experimental outcome from the Ruifrok deconvolution method. The given color vector does not faithfully represent the stains in the image, and the information is mixed in both channels.
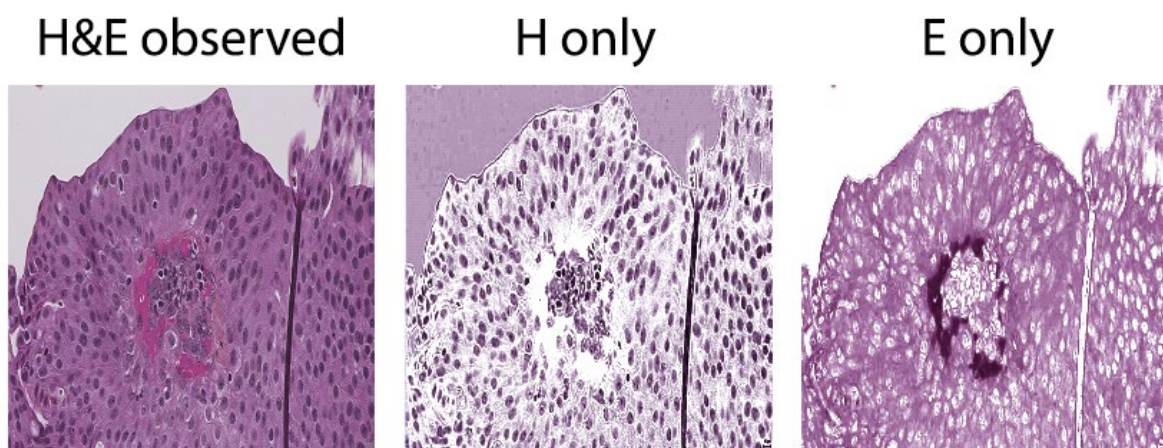
Figure 45: An experimental outcome from the Bayesian method in [67]. The color vector is correctly estimated, and the H- and E-channels information is correctly separated.

3.4.2 Stain Color Normalization

Color Normalization is the most popular approach for dealing with color variation [68]. Dealing with color variation can impact the latency of the pipeline, especially for large-size high-resolution images. It works by adjusting values in the image to map the source image to a new target image based on a template image. There are numerous approaches used in literature to reduce stain variability with traditional and ML methods. It is arguable to prioritize one method over the other based on complexity and effectiveness. However, the success of the color variation step is mainly measured based on the improved accuracy of computer-aided diagnosis systems. The authors of the paper [42] explain several approaches for color normalization, two common choices of template-based color normalization are shown in Figure 46. The following sections will discuss some aspects of the different approaches with details about use cases and implications.
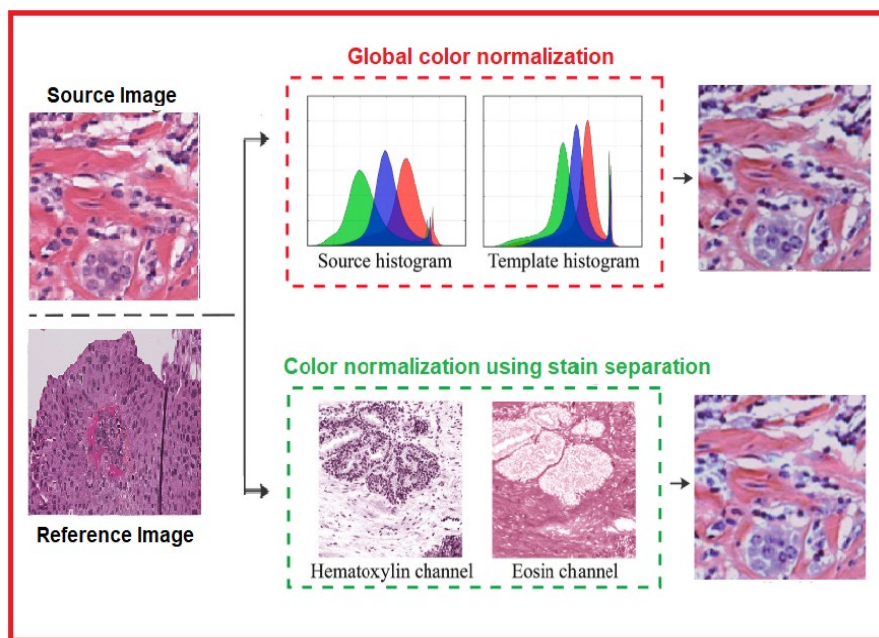
Figure 46: Different approaches to adjust color variability in H&E patches using source and reference patch. Color Normalization using global and color deconvolution.

3.4.2.1 Global Color Normalization

Global Color Normalization methods apply techniques previously designed for general imaging. They do not use stain separation and usually assume that both stains are equally distributed in the images. Global Color Normalization is the most naive approach and usually the fastest. However, it can lead to artifacts when the target and source images are very different [42] [67]. Histogram matching techniques adjust images colors in the RGB space through information represented in their histograms [33]. Color maps sort intensity levels of each RGB color channel of the original images to the sorted intensity levels of the corresponding channels of the template image. See figure 47 for histogram matching results.

Reinhard et al. [15] proposed a method based on color-space perception where perceptual lightness (L) and a pair of (α, β) parameters were used. They convert an RGB image into $l\alpha\beta$ color space and transfer the mean and variance of each axis from target image to source image. Then, the image is transformed back to RGB. The Reinhard method preserves all source intensity variations and structure but relies more on the target image. The output patch has approximately the same contrast as the target image. Although the method in [57] was designed and tested for general imaging like sunsets or paintings, it can be used for histopathological images. However, the $l\alpha\beta$ transformation does not separate stains properly. Figure 48 depicts the results from an experiment performed on an example image from one of the Clarify datasets. We have used the mean and standard deviation of perceptual parameters from fixed weights. This method is slightly more computationally expensive than histogram equalization.
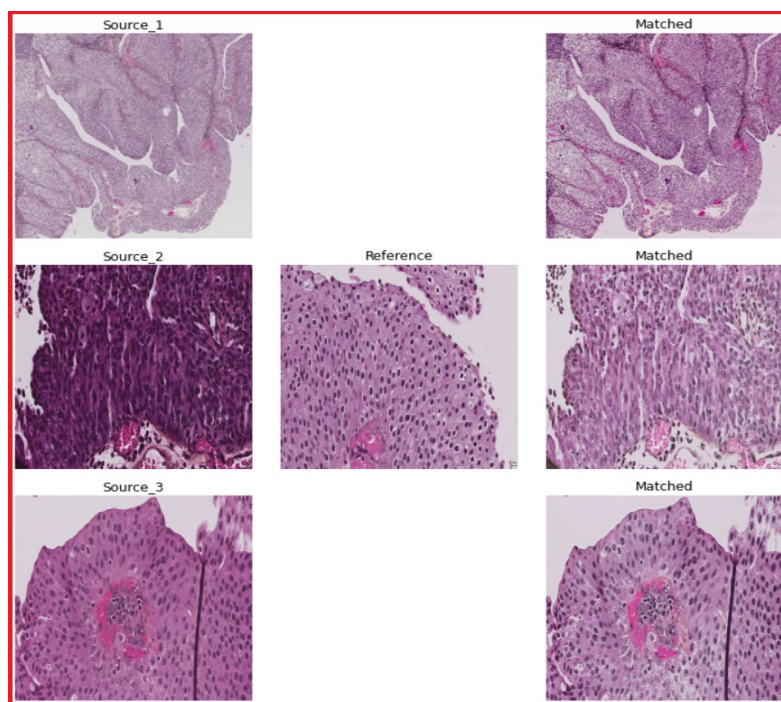
Figure 47: Result of histogram matching using Urothelium as a reference; Average processing time for a 256*256 patch is 0.03s.
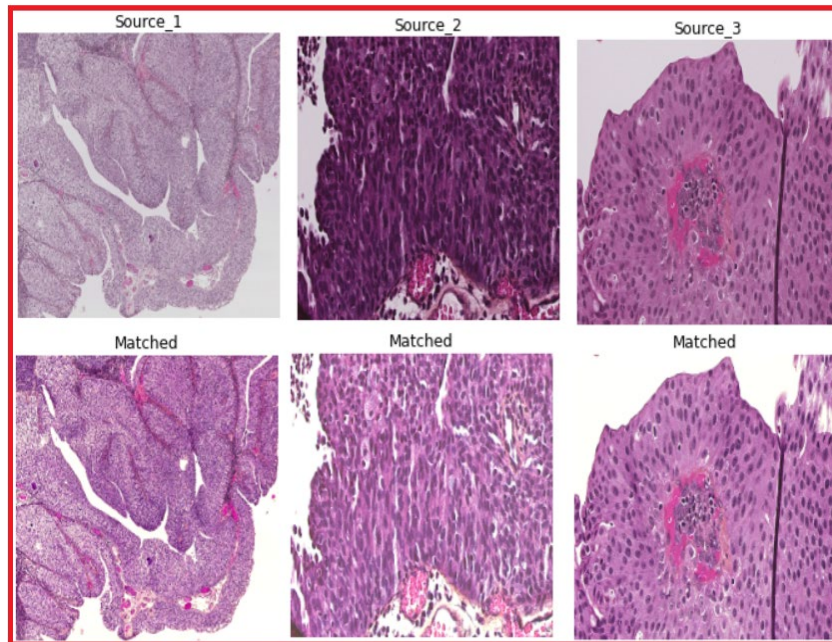
Figure 48: Implementation of lαβ color space proposed by Reinhard et al.; Average processing time for a 256*256 patch is 0.04s.

3.4.2.2 Color Deconvolution based Normalization

In contrast with the global normalization methods, color deconvolution considers the staining procedure used to obtain the histopathological images. As seen in previous sections, all color deconvolution techniques have in common: they separate stain color from structural information (stain concentration), offering a solid control on the information preprocessing and preserving the tissue structure, leading to the high fidelity of the original images. It is possible to normalize the images with the appropriate transformations as if they were stained with the same staining procedure and conditions. This approach can significantly reduce the problem of stain variation, but it can be computationally expensive.

The Color deconvolution-based normalization method also uses a template image. The concentrations C and the color-vector matrix M are estimated for both template and source image. Let C, M be the concentration and color-vectors for a source image and C_{temp}, M_{temp} corresponding to the template image. The concentration matrix C is often adjusted to have the same dynamic range as C_{temp}. The color matrix M is discarded and replaced with M_{temp} representing the colors of the objective image. The quality of the normalization depends on the quality of the BCD separation and the fidelity to the colors in the images. The proposed approaches often consider biological conditions and details that general approaches ignore. Color deconvolution-based stain normalization is explicitly designed for the normalization of histological images. Figure 49 illustrates the procedure, and an example using Macenko`s method is displayed in figure 50.
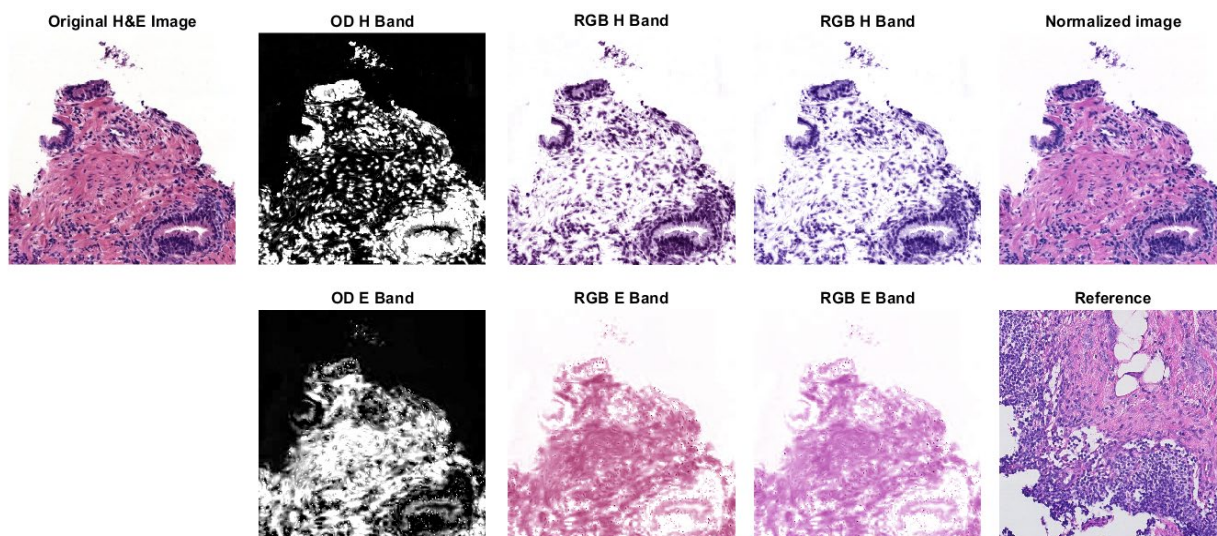
Figure 49: Example of color deconvolution-based normalization using the Super Gaussian method in [65]. The stains are separated and then individually modified to match those in the template image. The first column corresponds to the original image.
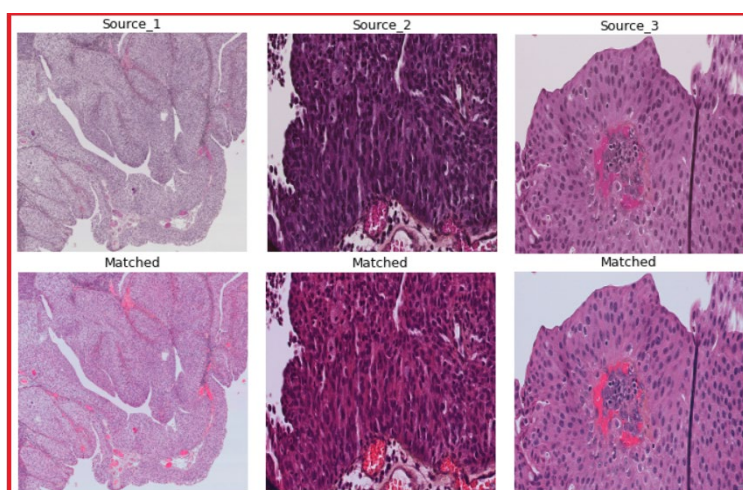
Figure 50: Color-deconvolution-based method to separate stain dyes and perform normalization, Macenko et al. [16]; Average processing time for a 256*256 patch is 0.21s.

3.4.3 Color Augmentation

Color augmentation can be considered as a type of data augmentation to simulate realistic variation in training data. These variations are aimed to mimic the appearance of future test samples. Color augmentation can be done in addition to normalization or avoiding color normalization as a preprocessing step, particularly during the training of AI models [17] [68]. In general, it can be stated that color normalization tries to avoid color variations in the data, whereas color augmentation accepts the color variations and tries to adopt a model. It can be counted as an approach to match training and test color distribution. Figure 51 shows different methods of patch augmentation.
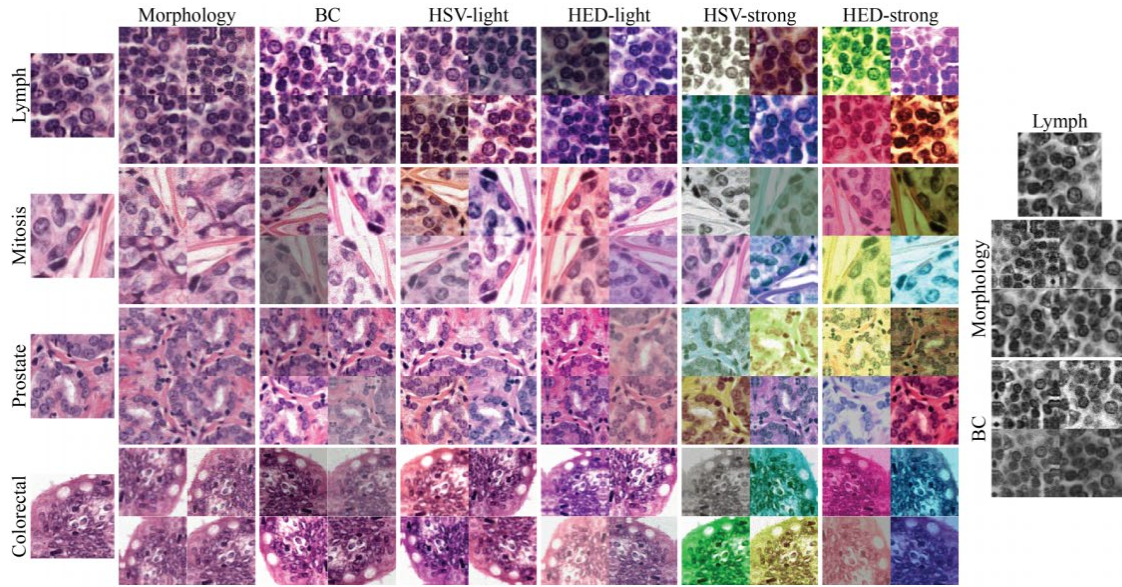
Figure 51: An overview of data augmentation techniques for four subtypes. The image depicts all morphological and color-based augmentation. The picture was taken from the original paper by Tellez et al. [17] and reprinted under arxiv.org's non-exclusive license to distribute.

In other words, this method is helpful to train the model when color variations are limited, and the model is expected to test different prospective data. Unlike other morphological augmentations in computational pathology where data is rotated, scaled, and flipped, the color augmentation techniques borrowed from computer vision can be applied on the fly and include several methods like introducing hue and contrast perturbations. Another stain augmentation based on color deconvolution was tested with ensembled networks [70]. Tellez et al. normalized OD matrix M$\in R^{3*3}$ (obtained by Beer Lambert law, as described in section 3.4.1) for RGB patches (P) and transformed them to HED color space S.

$$S = -log\,(P + \epsilon)\,M^{-1}$$

Where ϵ is positive bias to avoid numerical errors and each stain component i is individually multiplied with values α_i, β_i from a uniform distribution. We can simulate this stochastically by modifying component obtaining new RGB patch $P \in R^{N*3}$ from augmented (Hematoxylin-Eosin-Diaminobenzidine) HED color-space S'.

$$S' = \alpha_i S_i + \beta_i$$

$$P' = exp\,(-S'M) - \epsilon$$

Tellez et al. [17] have presented results after training ensembled CNN models using normalization and augmentation methods where color augmentation gives better results than color normalization by simulating stain variations. However, the results are directly proportional to more iterations to discover different possible color distributions. Moreover, some augmented samples result in unnatural colors where the nuclei are faded, which is not desired. This approach could be a possible motivation for the smaller dataset where diversification is needed to adjust variations. Xiao et al. [61] employed hue-based augmentation to increase the variety of data and refrain from slide-wise stain differences to improve model tolerance to new stains.

Figure 52 shows the output of color augmentation with ten iterations. It can be seen that every augmentation introduces different results due to multiplication by a random value. This approach is computationally costly and makes learning slower.
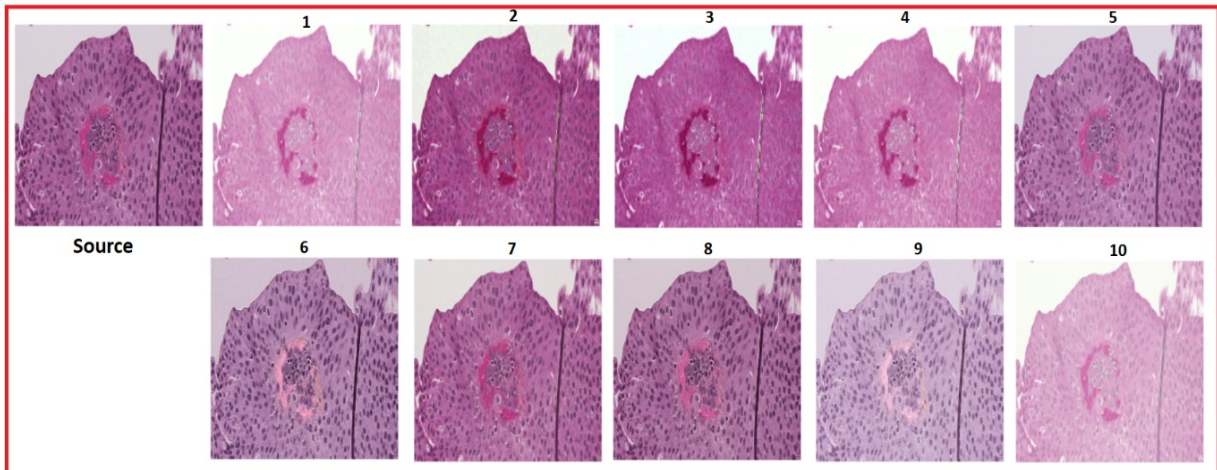
Figure 52: Color Augmentation approach with a source patch; Average processing time for a 256*256 patch is 2.4s.

To sum up, stain separation, normalization, transfer, and augmentation are popular image processing techniques to improve the generalization of a task [26] [70]. The most beneficial approach may vary depending on the task to perform and the system limitation [72]. Figure 53 provides a summary of visual outcomes from the discussion. It is important to be aware of the advantages and disadvantages of each approach in order to choose the most appropriate preprocessing pipeline. The most standard evaluation criteria for measuring the result of color variation techniques are described in the next section (3.4.4).
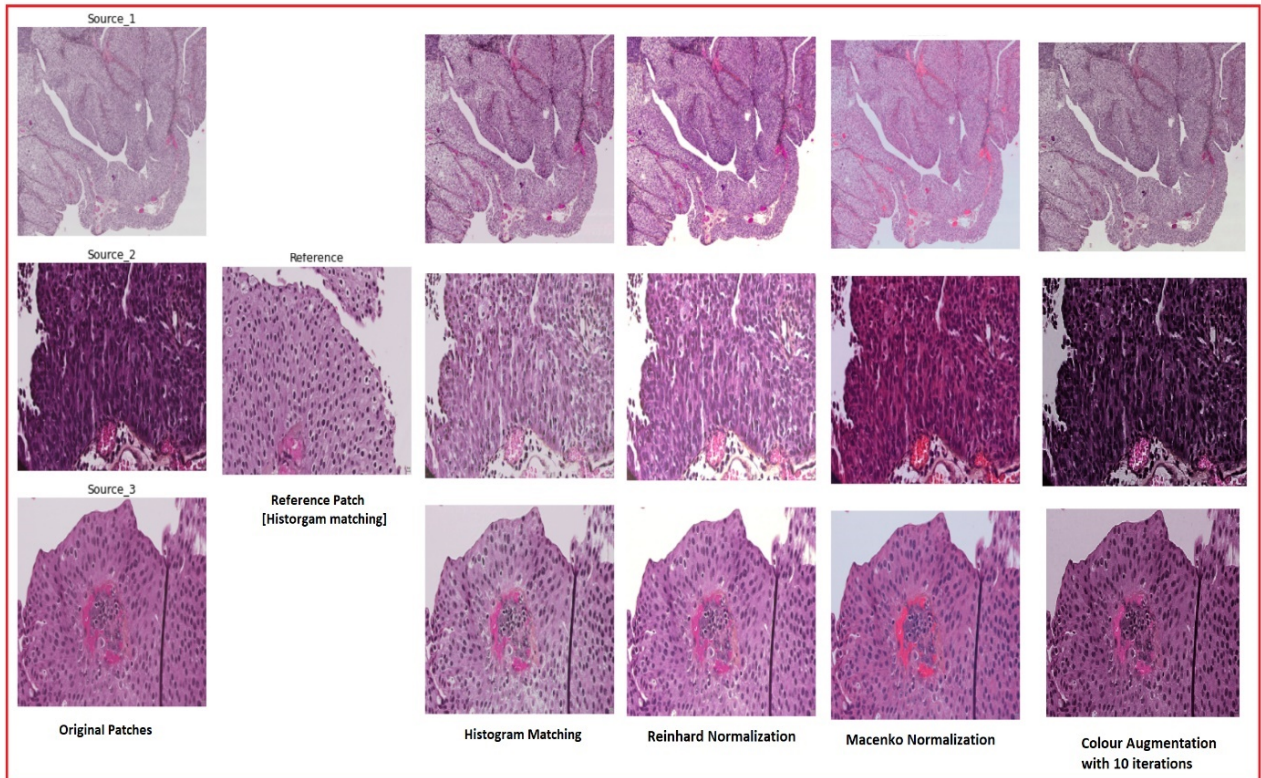
Figure 53: A summary of section 3.4 with outputs from all techniques. The picture shows a comparison of histogram equalization, Reinhard Method, and Macenko Method of normalization. The last column represents color augmentation output.

3.4.4 Normalization Quality Evaluation

The quality of the images is the primary concern in the preprocessing of histological samples. Apart from qualitative evaluation, usually done by a pathologist, the preprocessed images need to be quantitatively evaluated [73]. This section introduces the most expected quality metrics after applying a method to reduce color variation.

For BCD methods, the quality of their separation is usually measured where a single stain ground truth is required. It can be obtained through deconvolution when the color-vector matrix is determined by an expert pathologist [64]. Then, the quantitative evaluation can be carried out based on the similarity and correlation to the ground truth. The Structural Similarity Index (SSIM) [23] and the Peak Signal-to-Noise Ratio (PSNR) are often used for this purpose [67]. More recent works use the Quaternion Structure Similarity Index (QSSIM) to better measure color variation [74].

$$SSIM(x,y) = \left(\frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1}\right)\left(\frac{2\sigma_{xy} + c_2}{\sigma_x^2 + \sigma_y^2 + c_2}\right)$$

Where μ_x and μ_y mean values, σ_x and σ_y are standard deviations of source and normalized patch c_i are constants and σ_{xy} is a correlation coefficient between input x and output y. It focuses on scalar correlation to measure image quality. The numerical result of SSIM ranges from 0 to 1, where a higher value represents a better-normalized method.

Measuring the quality after color normalization is more complex as the ground truth is often not available. QSSIM and PSNR can be used on datasets like Mitos-Atypia [45] [43], where the same slides were digitized using two different scanners. Therefore, one of the scanners can be considered the ground truth. Using similarity metrics between the original and normalized image is not a good idea, as the best result will be obtained when the original image is not modified.

A common approach for measuring the quality of color normalization is the Normalized Median Intensity (NMI) [36][73]. NMI is employed to characterize the color constancy across all images in a dataset. A NMI value is calculated for each image, and then the standard deviation of the whole dataset (NMI SD) and coefficient of variation (NMI CV) (i.e., NMI SD divided by the mean) is calculated. Lower NMI SD and NMI CV values indicate minor color variation in the dataset and, therefore, a more consistent normalization. It is important to note that NMI does not measure the structure preservation during the normalization but only the color consistency of the images in the dataset.

$$NMI(I) = Med(u_i)P_{95}(u_i)$$

Where I is WSI, u_i denotes the mean of RGB channels for i-th pixel. Med refers to the median value, and P_{95} is the 95th percentile.

Finally, the most critical evaluation of a color preprocessing system is the CAD performance. All the preprocessing methods covered in this deliverable are intended to produce a better performance on WSI analysis. Therefore, the methods should be tested against final applications as classification or segmentation. In paper [66], the stain separation by different methods was tested against a set of shallow and deep classifiers (Random Forest, Xgboost, Gaussian Processes, and Deep Gaussian Processes). In [17], a complete quantification of

stain normalization and augmentation methods was performed using a benchmark of CNN classifiers.

3.5 Patch Selection based on ROI extraction

Grid patch extraction provides a redundant number of patches from the background and other tissue parts, which may not be significant for a specific classification task. The quality of the training data and labels is vital for the performance of the AI models during training and validation. As already mentioned, if the training set is too small or not diverse enough, the models are not robust or general enough to work for previously unseen data from other sources. A prime drawback is noise in training set in different ways. One type of noise is that the training set contains irrelevant samples concerning the task for which we intend to develop a model. It introduces the loss of discriminative details by introducing unnecessary histological features [76]. For example, if the training set contains patches of tissue types not relevant for the staging, the CNN model might suffer and struggle with confusing information [38].

Correspondingly, feeding a CNN with low-quality data will not end up with a good DL model. Often it will be desired to look at ROI and only use patches from these regions when training for specific tasks, as shown in figure 54. Such as for NMBC staging, borderline tissues of urothelium and stroma are valuable in identifying invasion in bladder cancer. A few related examples are explained in section 3.2.2, where pixel-level annotations help choose a specific region. This can be partly overlapping with the task of patch exclusion. The methods described in section 3.3 patch exclusion are significant for removing patches that we consider artifacts before selecting patches. Tissue classification can be an early step in the pipeline where the output of the tissue classifiers forms ROI to consider further. Many studies report 2-stage classification where patches are selected by the first classifier [41] [58] [76].
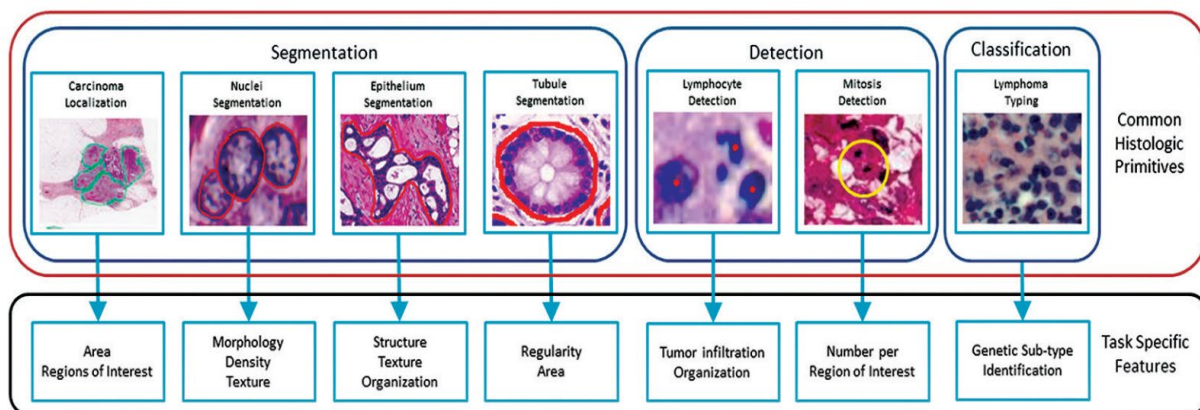

Figure 54: A box diagram showing generic histologic primitives and relevant features in patches for subsequent use in task-specific classifiers. These are some common higher-order tasks in DP research. It is reprinted from [22] under Creative Common Attribution-*Noncommercial-Share Alike 3.0 License*.

3.5.1 Tissue Classification

In some diseases, like NMBC, the main part of the diagnostic relevance lies in the urothelium parts of the tissue for grading and urothelium and stroma borders for staging. The WSI can have regions of blood, muscle, and cauterized areas in addition to the interesting parts such as urothelium and stroma [46]. Extracting engaging tissue from the other available patches is a type of ROI extraction [34] [35], and it can be regarded both as a system of its own or as a preprocessing part for models predicting diagnostics or prognostics. This part may somehow correlate with guided patch extraction in the sense that we want to add patches to our training set based on a task-specific classifier [60]. The use of grid patching raises the importance of ROI-based patch selection for training models such as mitosis detection. This step can be carried before color normalization to avoid waste of computational resources since patches relevant to classification tasks are only necessary.

In SML, segmented-guided patch extraction is performed where suspicious patches are marked with a positive label (automatic or manual), leading to improved performance relative to only using slide-level labels [78]. An analogous approach can be used in breast cancer combined with a smart-patching strategy where for nuclei-related tasks, a nuclei-guided segmentation can be used [77]. Nuclei-guided sampling approach terms patches as positive where a high density of nuclei is present, unlike adipose tissue that represents fats mainly. The appearance of the spatial distribution of nuclei in TNBC slides can be found using blob detection methods [29] [43].

3.5.2 Subsampling and Class Balancing

In histological images, high-resolution patches provide better learning characteristics to CNN. However, it results in hundreds of thousands of patches from a single WSI, and it might have a higher number of patches from one tissue type than another. Thus, selecting a subset of the possible patches (single or multiresolution) from each WSI can be necessary to control the number of patches in the training set and balance the classes. Another approach could be choosing a fixed number of patches from each class based on grade or scoring information [78] [79].

Patch subsampling can be planned by randomly choosing patches or predefined patterns as described in the patching methods section, possibly balancing the classes, including different resolutions, etc. Several color deconvolutions schemes have been proposed to identify non-informative regions [49] [66]. This step may help formulate bags of patches based on positive and negative classes, as shown in figure 55. Patches can further be used in supervised training with strong patch-based labels or with weak slide-based labels. Creating bags of patches for the Multiple Instance Learning (MIL) approach can be a helpful strategy when only weak labels are known [80] [81].
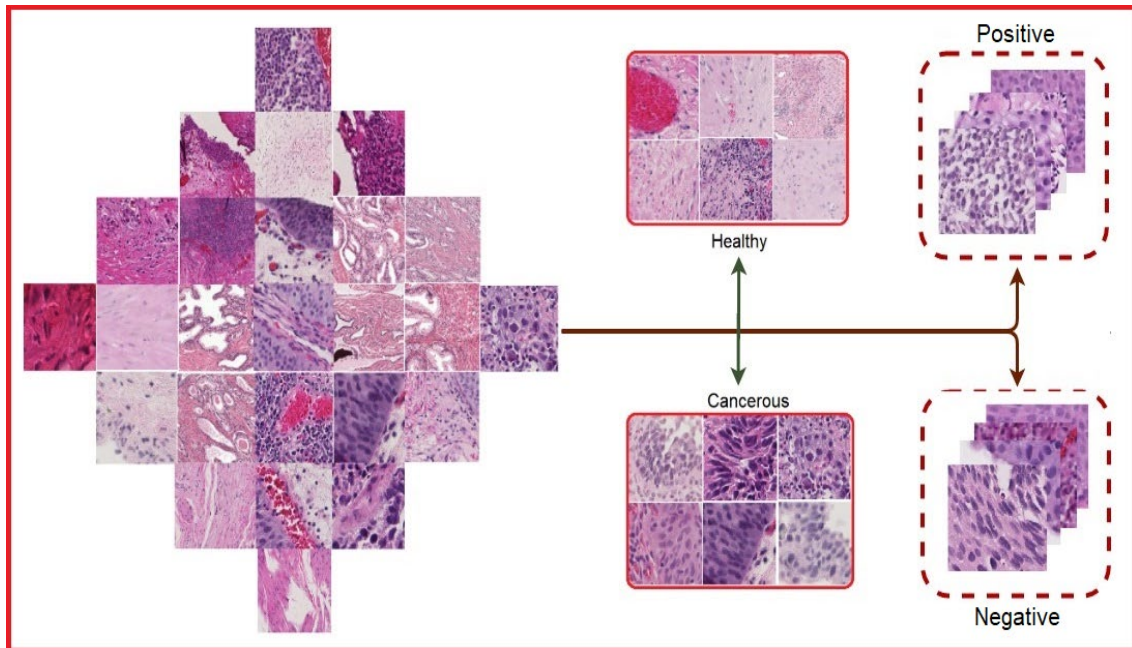
Figure 55: An example of a patch selection for various computational pathology tasks [80]. Different tasks require different groups. Labels can be based on manual annotation, automatic tissue segmentation, clustering of feature space, etc.

4 Recommendations and Opinions

A modular and sequential approach is deemed suitable to tackle different challenges in preprocessing gigapixel images. We have proposed a five-step pipeline to define a protocol for preprocessing WSI.

Image quality is the first important step before performing other computations. A quality check can empirically be performed from metadata parameters such as magnification and format and computing background/foreground ratios.

It is almost hard to process a complete WSI image with common hardware resources for a training model due to its high resolution and memory requirement. Patching WSI is a common practice where choosing a patch size is debatable subjective to histological features for training a classifier. Patching makes the process efficient, which is one of the targets in WP3.

Annotation-guided patching is an intelligent choice for extracting patches from ROI only. Not all patches are equally informative; patches with blur, burnt, folded tissue, blood, and air bubble can be identified with the help of classical image processing methods or methods based on CNN and excluded from further processing in terms of training or prediction. Deblurring steps can augment the computational complexity and may not result in very sharp and clear patches. If enough data is available, the deblurring step should be considered to exclude blurred patches. It addresses the WP3 requirement of extracting the most significant features.

Identification of color variation is an essential part of preprocessing and a significant action of task T3.1. Stain separation, normalization, and augmentation are common methods for improving the robustness of models against stain variations. Stain normalization based on deconvolution yields a better quantitative evaluation for structural similarity than global color normalization methods. On the other hand, global color normalization methods are fast and require less computational power. The selection of an appropriate method among these two can be evaluated based on computational requirements. Color augmentation can be done on the fly to improve the generalization of a model against different stain variants. This step is very reasonable for training datasets, whereas performing normalization for prediction or image retrieval is rational. There are qualitative and quantitative methods to evaluate the suitable choice. Section 3.4 provides more details and experimental outcomes on that.

Patch selection is an optional step for introducing class balance and training models based on the ROI. The patches can be bagged based on positive and negative classes for different learning schemes. The use of multiresolution patches can be beneficial to identify morphological features at a different level, similar to a clinical approach. It also complies with the requirements of T3.1 for subsequent analysis. Compilation of all these methods results in a ML pipeline which is the output of D3.1. Table 9 provides a summarized overview of these methods with advantages and limitations.

Challenge/ preprocessing step	Approach	Method	Advantages	Limitation
Quality Check	HistoQC Tool	• Evaluate metadata parameters. • Identify pen marking and coverslips. • Apply several filters to find blur ratio and air bubble ratio.	• Line plots make it easier to identify a WSI different from other WSI in the dataset. • Provides binary masks for selected modules.	• Requires a certain threshold on outputs to assess quality requirements. • Only suitable for WSI from the same scanner in the batch.
	Metadata Consultation	• Resolution • File type • Other metadata	• Easy to implement	• Only based on metadata, not content.
Patching Method **(What to patch)**	Grid Patching Method	• Divide complete WSI images into predefined patch size.	• It can be performed regardless of annotation.	• Memory Intensive.
	Guided Patching Method based on expert annotations.	• Patch the regions annotated with specific labels.	• Suitable for training a task-based classifier. • Identify non-informative patches based on annotation. • Reduces computation and memory requirements.	• Requires pixel-level/detailed annotation.
	Multiresolution Patches	• Extracting patches 25x,100x, and 400x from ROI.	• Patches at different resolutions may abstract better morphological features. • Using one patch size for the network and using different views.	• It may require post-processing techniques or multiple models for classification tasks.

73

Explicit vs. implicit Patching (when to patch)	Pre-patch		• Save patches as smaller images before further processing (learning/prediction).	• Patches saved after preprocessing do not need to be preprocessed again.	• Requires extra storage. • Reading from and writing to disk can be time-consuming. • Not flexible in terms of sliding window-patching, changing patch sizes, multiresolution formulations.
	Patch on-the-fly		• Save lists of coordinates to patches.	• Formulations can be flexible and easy to vary in terms of sliding windows, varying patch sizes, multiresolution formulations. • Attaching multiple tags/labels to a coordinate in the list makes it easy to pick out regions with different labels. • Does not require any additional storage. • Using appropriate libraries, only the part of the WSI defined by the coordinate lists and patch size is loaded into memory.	• It requires more time to perform the preprocessing in patching-on-the-fly. It is a bit hardware-dependent if the reading from disk or preprocessing is the most demanding part. • WSI has to be loaded and preprocessed during training every time. • Potential filtering or color normalization has to be done multiple times or stitched together to form large images already preprocessed.
Patch Exclusion	Foreground/ Background Segmentation of WSI to remove (patches from)	Discrete entropy of the histogram	• Calculates bins and measure average difference based on probability distribution.	• Possible to compute for gigapixel images without patching. • It helps to find WSI with a smaller tissue sample size if required for the omission.	• Does not perform well on WSI with dirty backgrounds or overall blurred slide.
		Otsu Thresholding	• Calculates single value to minimize intra-class variance.	• Good thresholding when the histogram is binomial.	• Does not provide a reasonable threshold well

	background areas			• Two-dimensional Otsu works better for noisy images.	with huge color variations and scattered tissue.
	Elimination of artifacts	Blurry patches	• Laplace filters	• Faster in finding attenuated frequencies in the image.	• Laplace may not detect very small blur/dot-like in more significant patches.
		Air Bubble	• Frangi filter for tube-like air bubble • Color transformation for bubbles on a corner and using a threshold for Value and Hue channel	• Classical image processing methods and require less computation.	• Fixing parameters may not give desired results for all color variations, especially with dirty slide backgrounds. • Hard to find a small bubble on the main tissue.
		Folded/Burnt/Blood patches	• Thresholding on HSV color-space transformation	• The hue channel presents better thresholding for blood. • Value channel is easier to a threshold for burnt and folded tissue.	• It works better when a patch has some typical tissue sample as well. • Hard to threshold with a full damaged patch.
		Classifier for Damaged/Blood patches	• Annotated slides are used to train a classifier to accommodate the shape and color of damaged and blood slides.	• Combines global contextual information by training on multiple magnification scales.	• Training a separate classifier requires a dataset with tiles from all classes present in WSI.
Color variation	Color normalization	Global Color Normalization (Reinhard Method)	• Convert the image from RGB to $l\alpha\beta$ and transfer color based on perception.	• Preserves structure and maintains contrast. • Estimates mean and variance from the target image.	• Transformation in $l\alpha\beta$ color space does not separate stains.

		Deconvolution based normalization (Macenko`s Method)	• Convert RGB to OD space and estimates stains by Single-Value decomposition (SVD).	• Relies on stain vectors estimation from the target image.	• Does not preserve all information of the source image. • Computationally heavy relative to other methods.
	Color deconvolution	Super Gaussian Priors for Blind Color Deconvolution	• Calculate concentration covariance after transforming RGB to OD space.	• Estimates stains with interference on neighboring pixels to produce a better color distribution.	• Stain estimation relies on the selection of Hematoxylin and Eosin reference pixels.
	Color Augmentation	HED Color Augmentation	• Convert image in HED color space and adds random numbers from range.	• Creates robust models to perform against the color variation. • Beneficial in smaller datasets with less variation	• Augmenting bigger values may result in undesired variations. • Not beneficial as normalization in prediction.
Patch Selection **(ROI and subset selection)**	Sub-Sampling and Class Balancing		• Sorting patches as positive or negative for the training of different tasks.	• Learn better models with balanced classes.	• Subsampling does not use all the available patches.
	Tissue classification for ROI extraction.		• Automatic selection of relevant patches for specific tasks based on CNN (such as staging).	• It helps in forming bags of, for example, cancerous and non-cancerous patches.	• It is fixed for a specific task.

Table 9: Recommendations and opinions about several approaches proposed in this document.

5 Conclusion

Research in digital pathology (DP) is observing an increase in the use of Machine Learning algorithms. Several research papers in medical imaging have proven that adding preprocessing steps helps in developing robust models. Combining traditional methods and preprocessing with deep learning frameworks can obtain a progressively high-performing CAD system. This document explains some popular methods for preprocessing WSI of three cancer diseases in CLARIFY project, namely NMBC, TNBC, and SML. The document also contains input from CLARIFY internal survey in every section, which can support understanding standard practices and preferred methods. The survey provides an abstract overview of preprocessing necessities where all early-stage researchers and supervisors have provided their feedback.

We have observed samples from datasets obtained from different institutions in the consortium. H&E images exhibit significant variations due to the scanner, age of the slide manual dying, and other artifacts. The deliverable also exhibits a set of existing artifacts in our datasets. In the light of its significance in DP, we propose various techniques to deal with color variability, selection of patches, and exclusion of non-informative patches based on their time complexity.

It is hard to establish a single preprocessing pipeline standard for all diseases and tasks in the CLARIFY project primarily due to the nature of different subsequent analyses for different cancer types and variations in datasets. Nevertheless, we have proposed a five-step preprocessing pipeline to cope with common challenges in handling high-resolution histological images.

1. At first, an Image quality check can be performed from metadata parameters and potentially from using tools like HistoQC.
2. A patching method has to be used to split the gigapixel image before performing further image analysis and classification.
3. Some patches are excluded based on foreground/background segmentation or artifacts detection.
4. Stain normalization is the most practiced step in preprocessing to adjust the variation in thickness of staining. In the light of its significance in DP, we propose various techniques to deal with color variability and their time complexity. We have also described possible evaluation methods for a normalized image. Alternatively, we propose color augmentation to develop a stain-resistant model by adding color variability for smaller datasets.
5. Lastly, based on the specific task at hand, patch selection is helpful to make an unbiased classifier or removing correlation from other parts of tissue types. The selection of patches can be useful to create a class balance or creating bags based on image particulars.

As indicated in Figure 14, steps 4 and 5 might be performed in the opposite order, especially if a patch-on-the-fly approach is chosen or if tissue segmentation is performed, including segmentation of damaged and blurry parts as well diagnostically relevant segments.

Different tasks, like "grading," "staging," "prognostic prediction," or "content-based image retrieval," might require different parts of the image, i.e., the patch selection is task-dependent. Nevertheless, we believe that the other steps of the preprocessing pipeline can be the same, independent of the task.

6   References

[1]   "Analog Intelligence." https://analogintelligence.com/artificial-intelligence-ai-startups-pathology-venture-meta-review-analysis/.

[2]   C. Gandi, L. Vaccarella, R. Bientinesi, M. Racioppi, F. Pierconti, and E. Sacco, "Bladder cancer in the time of machine learning: Intelligent tools for diagnosis and management," *Urol. J.*, vol. 88, no. 2, pp. 94–102, Jan. 2021, doi: 10.1177/0391560320987169.

[3]   G. Bueno, M. M. Fernández-Carrobles, O. Deniz, and M. García-Rojo, "New Trends of Emerging Technologies in Digital Pathology," *Pathobiology*, vol. 83, no. 2–3, pp. 61–69, 2016, doi: 10.1159/000443482.

[4]   N. Dimitriou, O. Arandjelović, and P. D. Caie, "Deep Learning for Whole Slide Image Analysis: An Overview," *Front. Med.*, vol. 6, no. November, pp. 1–7, 2019, doi: 10.3389/fmed.2019.00264.

[5]   S. Hosein, C. R. Reitblat, E. B. Cone, and Q. D. Trinh, "Clinical applications of artificial intelligence in urologic oncology," *Curr. Opin. Urol.*, vol. 30, no. 6, pp. 748–753, 2020, doi: 10.1097/MOU.0000000000000819.

[6]   S. A. Harmon *et al.*, "Multiresolution Application of Artificial Intelligence in Digital Pathology for Prediction of Positive Lymph Nodes From Primary Tumors in Bladder Cancer," 2020. doi: 10.1200/cci.19.00155.

[7]   F. Yi, J. Huang, L. Yang, Y. Xie, and G. Xiao, "Automatic extraction of cell nuclei from H&E-stained histopathological images," *J. Med. Imaging*, vol. 4, no. 2, p. 027502, 2017, doi: 10.1117/1.jmi.4.2.027502.

[8]   K. Fan, S. Wen, and Z. Deng, *Deep Learning for Detecting Breast Cancer Metastases on WSI*, vol. 145. Springer Singapore, 2019.

[9]   "CLARIFY." http://www.clarify-project.eu/.

[10]   M. Khened, A. Kori, H. Rajkumar, G. Krishnamurthi, and B. Srinivasan, "A generalized deep learning framework for whole-slide image segmentation and analysis," *Sci. Rep.*, vol. 11, no. 1, pp. 1–14, 2021, doi: 10.1038/s41598-021-90444-8.

[11]   M. He, D. Chen, J. Liao, P. V. Sander, and L. Yuan, "Deep exemplar-based colorization," *ACM Trans. Graph.*, vol. 37, no. 4, p. 16, 2018, doi: 10.1145/3197517.3201365.

[12]   A. Cruz-Roa *et al.*, "Accurate and reproducible invasive breast cancer detection in whole-slide images: A Deep Learning approach for quantifying tumor extent," *Sci. Rep.*, vol. 7, no. April, pp. 1–14, 2017, doi: 10.1038/srep46450.

[13]   V. Rastogi, N. Puri, S. Arora, G. Kaur, L. Yadav, and R. Sharma, "Artefacts: A diagnostic dilemma - A review," *J. Clin. Diagnostic Res.*, vol. 7, no. 10, pp. 2408–2413, 2013, doi: 10.7860/JCDR/2013/6170.3541.

[14]   "JCDR - artefacts, sectioning artefacts, staining artefacts, squeeze artefacts, crush artefacts, histopathology." https://www.jcdr.net/article_fulltext.asp?issn=0973-709x&year=2013&month=october&volume=7&issue=10&page=2408&id=3541

(accessed May 20, 2021).

[15]   E. Reinhard, M. Ashikhmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Comput. Graph. Appl.*, vol. 21, no. 5, pp. 34–41, 2001, doi: 10.1109/38.946629.

[16]   M. Macenko *et al.*, "A METHOD FOR NORMALIZING HISTOLOGY SLIDES FOR QUANTITATIVE ANALYSIS 3 Statistics and Operations Research , 4 Lineberger Comprehensive Cancer Center , 5 Renaissance Computing Institute , 6 Pathology and Laboratory Medicine , 7 Dermatology University of Nor," *IEEE Int. Symp. Biomed. Imaging*, pp. 1107–1110, 2009.

[17]   D. Tellez *et al.*, "Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology," *Med. Image Anal.*, vol. 58, 2019, doi: 10.1016/j.media.2019.101544.

[18]   G. Romero Lauro *et al.*, "Digital pathology consultations - A new era in digital imaging, challenges and practical applications," *J. Digit. Imaging*, vol. 26, no. 4, pp. 668–677, 2013, doi: 10.1007/s10278-013-9572-0.

[19]   A. L. Martel *et al.*, "An image analysis resource for cancer research: PIIP—Pathology image informatics platform for visualization, analysis, and management," *Cancer Res.*, vol. 77, no. 21, pp. e83–e86, Nov. 2017, doi: 10.1158/0008-5472.CAN-17-0323.

[20]   H. E. Källman, E. Halsius, M. Olsson, and M. Stenström, "DICOM metadata repository for technical information in digital medical images," *Acta Oncol. (Madr).*, vol. 48, no. 2, pp. 285–288, 2009, doi: 10.1080/02841860802258786.

[21]   J. Silva-Rodríguez, A. Colomer, M. A. Sales, R. Molina, and V. Naranjo, "Going deeper through the Gleason scoring scale: An automatic end-to-end system for histology prostate grading and cribriform pattern detection," *Comput. Methods Programs Biomed.*, vol. 195, p. 105637, Oct. 2020, doi: 10.1016/j.cmpb.2020.105637.

[22]   A. Janowczyk and A. Madabhushi, "Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases," *J. Pathol. Inform.*, vol. 7, no. 1, Jan. 2016, doi: 10.4103/2153-3539.186902.

[23]   Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, 2002, doi: 10.1109/97.995823.

[24]   A. Janowczyk, R. Zuo, H. Gilmore, M. Feldman, and A. Madabhushi, "HistoQC: An Open-Source Quality Control Tool for Digital Pathology Slides," *JCO Clin. Cancer Informatics*, no. 3, pp. 1–7, 2019, doi: 10.1200/cci.18.00157.

[25]   M. S. Hossain, F. Kimura, Y. Yagi, M. Yamaguchi, and T. Nakamura, "Practical image quality evaluation for whole slide imaging scanner," no. April 2018, p. 60, 2018, doi: 10.1117/12.2316764.

[26]   A. Anghel *et al.*, "A High-Performance System for Robust Stain Normalization of Whole-Slide Images in Histopathology," *Front. Med.*, vol. 6, Sep. 2019, doi: 10.3389/fmed.2019.00193.

[27]   "No Title（表示不可能）." http://www.andrewjanowczyk.com/wp-content/uploads/2019/04/HistoQC_w_supplemental.pdf (accessed May 18, 2021).

[28]   S. Hemati, S. Kalra, C. Meaney, M. Babaie, and H. R. Tizhoosh, "CNN and Deep Sets for End-to-End Whole Slide Image Representation Learning," pp. 1–11, 2021.

[29]   B. E. Bejnordi *et al.*, "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," *JAMA - J. Am. Med. Assoc.*, vol. 318, no. 22, pp. 2199–2210, Dec. 2017, doi: 10.1001/jama.2017.14585.

[30]   G. Aresta *et al.*, "BACH: Grand challenge on breast cancer histology images," *Med. Image Anal.*, vol. 56, pp. 122–139, 2019, doi: 10.1016/j.media.2019.05.010.

[31]   P. Bándi *et al.*, "From Detection of Individual Metastases to Classification of Lymph Node Status at the Patient Level: The CAMELYON17 Challenge," *IEEE Trans. Med. Imaging*, vol. 38, no. 2, pp. 550–560, 2019, doi: 10.1109/TMI.2018.2867350.

[32]   K. Roy, D. Banik, D. Bhattacharjee, and M. Nasipuri, "Patch-based system for Classification of Breast Histology images using deep learning," *Comput. Med. Imaging Graph.*, vol. 71, pp. 90–103, Jan. 2019, doi: 10.1016/j.compmedimag.2018.11.003.

[33]   T. A. Azevedo Tosta, P. R. de Faria, L. A. Neves, and M. Z. do Nascimento, "Computational normalization of H&E-stained histological images: Progress, challenges and future potential," *Artificial Intelligence in Medicine*, vol. 95. Elsevier B.V., pp. 118–132, Apr. 01, 2019, doi: 10.1016/j.artmed.2018.10.004.

[34]   R. Wetteland, K. Engan, T. Eftestøl, V. Kvikstad, and E. A. M. Janssen, "Multiclass tissue classification of whole-slide histological images using convolutional neural networks," *ICPRAM 2019 - Proc. 8th Int. Conf. Pattern Recognit. Appl. Methods*, pp. 320–327, 2019, doi: 10.5220/0007253603200327.

[35]   X. Zhu, J. Yao, F. Zhu, and J. Huang, "WSISA: Making survival prediction from whole slide histopathological images," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 6855–6863, 2017, doi: 10.1109/CVPR.2017.725.

[36]   H. Fatakdawala *et al.*, "Expectation-maximization-driven geodesic active contour with overlap resolution (EMaGACOR): Application to lymphocyte segmentation on breast cancer histopathology," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 7, pp. 1676–1689, 2010, doi: 10.1109/TBME.2010.2041232.

[37]   R. Wetteland, K. Engan, T. Eftestøl, V. Kvikstad, and E. A. M. Janssen, "A Multiscale Approach for Whole-Slide Image Segmentation of five Tissue Classes in Urothelial Carcinoma Slides," *Technol. Cancer Res. Treat.*, vol. 19, p. 153303382094678, 2020, doi: 10.1177/1533033820946787.

[38]   D. Komura and S. Ishikawa, "Machine Learning Methods for Histopathological Image Analysis," *Comput. Struct. Biotechnol. J.*, vol. 16, pp. 34–42, 2018, doi: 10.1016/j.csbj.2018.01.001.

[39]   P. Faridi, H. Danyali, M. S. Helfroush, and M. A. Jahromi, "Cancerous Nuclei Detection and Scoring in Breast Cancer Histopathological Images," 2016, [Online]. Available: http://arxiv.org/abs/1612.01237.

[40]   L. Hou *et al.*, "Dataset of segmented nuclei in hematoxylin and eosin stained histopathology images of ten cancer types," *Sci. Data*, vol. 7, no. 1, pp. 1–12, 2020, doi:

10.1038/s41597-020-0528-1.

[41] W. Bulten and G. Litjens, "Unsupervised Prostate Cancer Detection on H&E using Convolutional Adversarial Autoencoders," *IEEE Trans. Med. Imaging*, vol. 9, no. 1, pp. 1–11, 2018, [Online]. Available: http://arxiv.org/abs/1804.07098.

[42] M. Salvi, U. R. Acharya, F. Molinari, and K. M. Meiburger, "The impact of pre- and post-image processing techniques on deep learning frameworks: A comprehensive review for digital pathology image analysis," *Computers in Biology and Medicine*, vol. 128. Elsevier Ltd, p. 104129, Jan. 01, 2021, doi: 10.1016/j.compbiomed.2020.104129.

[43] H. Wang *et al.*, "Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features," *J. Med. Imaging*, vol. 1, no. 3, p. 034003, 2014, doi: 10.1117/1.jmi.1.3.034003.

[44] I. Keren Evangeline, J. Glory Precious, N. Pazhanivel, and S. P. Angeline Kirubha, "Automatic Detection and Counting of Lymphocytes from Immunohistochemistry Cancer Images Using Deep Learning," *J. Med. Biol. Eng.*, vol. 40, no. 5, pp. 735–747, Oct. 2020, doi: 10.1007/s40846-020-00545-4.

[45] C. Li, X. Wang, W. Liu, and L. J. Latecki, "DeepMitosis: Mitosis detection via deep detection, verification and segmentation networks," *Med. Image Anal.*, vol. 45, pp. 121–133, 2018, doi: 10.1016/j.media.2017.12.002.

[46] O. N. Dalheim, R. Wetteland, V. Kvikstad, E. A. M. Janssen, and K. Engan, "Semi-supervised tissue segmentation of histological images," *CEUR Workshop Proc.*, vol. 2688, pp. 1–15, 2020.

[47] S. A. Taqi, S. A. Sami, L. B. Sami, and S. A. Zaki, "A review of artifacts in histopathology," *Journal of Oral and Maxillofacial Pathology*, vol. 22, no. 2. Wolters Kluwer Medknow Publications, p. 279, May 01, 2018, doi: 10.4103/jomfp.JOMFP_125_15.

[48] O. Severino and A. Gonzaga, "A new approach for color image segmentation based on color mixture," *Mach. Vis. Appl.*, vol. 24, no. 3, pp. 607–618, 2013, doi: 10.1007/s00138-011-0395-z.

[49] A. C. Ruifrok and D. A. Johnston, "Quantification of histochemical staining by color deconvolution," *Anal. Quant. Cytol. Histol.*, vol. 23, no. 4, pp. 291–299, 2001.

[50] "Entropy Estimation," *Wiley StatsRef: Statistics Reference Online*, 2014. https://en.wikipedia.org/wiki/Entropy_estimation.

[51] H. Permuter, J. Francos, and I. H. Jermyn, "Gaussian mixture models of texture and colour for image database retrieval," *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, vol. 3, pp. 569–572, 2003, doi: 10.1109/icassp.2003.1199538.

[52] L. Hou, D. Samaras, T. M. Kurc, Y. Gao, J. E. Davis, and J. H. Saltz, "Patch-Based Convolutional Neural Network for Whole Slide Tissue Image Classification," 2016. doi: 10.1109/CVPR.2016.266.

[53] J. Xing, P. Yang, and L. Qingge, "Robust 2D Otsu's Algorithm for Uneven Illumination Image Segmentation," *Comput. Intell. Neurosci.*, vol. 2020, 2020, doi:

10.1155/2020/5047976.

[54]   H. Wu, J. H. Phan, A. K. Bhatia, C. A. Cundiff, B. M. Shehata, and M. D. Wang, "Detection of blur artifacts in histopathological whole-slide images of endomyocardial biopsies," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, vol. 2015-Novem, pp. 727–730, 2015, doi: 10.1109/EMBC.2015.7318465.

[55]   S. Pertuz, D. Puig, and M. A. Garcia, "Analysis of focus measure operators for shape-from-focus," *Pattern Recognit.*, vol. 46, no. 5, pp. 1415–1432, May 2013, doi: 10.1016/j.patcog.2012.11.011.

[56]   A. Levin, D. Lischinski, and Y. Weiss, "Colorization using optimization," *ACM SIGGRAPH 2004 Pap. SIGGRAPH 2004*, pp. 689–694, 2004, doi: 10.1145/1186562.1015780.

[57]   Y. Zheng, Z. Jiang, H. Zhang, F. Xie, J. Shi, and C. Xue, "Adaptive color deconvolution for histological WSI normalization," *Comput. Methods Programs Biomed.*, vol. 170, pp. 107–120, 2019, doi: 10.1016/j.cmpb.2019.01.008.

[58]   P. A. Bautista and Y. Yagi, "Detection of tissue folds in whole slide images," *Proc. 31st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. Eng. Futur. Biomed. EMBC 2009*, pp. 3669–3672, 2009, doi: 10.1109/IEMBS.2009.5334529.

[59]   Ş. Öztürk and B. Akdemir, "Effects of Histopathological Image Pre-processing on Convolutional Neural Networks," *Procedia Comput. Sci.*, vol. 132, no. Iccids, pp. 396–403, 2018, doi: 10.1016/j.procs.2018.05.166.

[60]   C. Bahlmann *et al.*, "Automated detection of diagnostically relevant regions in H&E stained digital pathology slides," *Med. Imaging 2012 Comput. Diagnosis*, vol. 8315, p. 831504, 2012, doi: 10.1117/12.912484.

[61]   Q. E. Xiao *et al.*, "Hematoxylin and Eosin (HE) Stained Liver Portal Area Segmentation Using Multi-Scale Receptive Field Convolutional Neural Network," *IEEE J. Emerg. Sel. Top. Circuits Syst.*, vol. 9, no. 4, pp. 623–634, 2019, doi: 10.1109/JETCAS.2019.2952063.

[62]   D. F. Swinehart, "The Beer-Lambert law," *J. Chem. Educ.*, vol. 39, no. 7, pp. 333–335, 1962, doi: 10.1021/ed039p333.

[63]   A. Vahadane *et al.*, "Structure-Preserving Color Normalization and Sparse Stain Separation for Histological Images," *IEEE Trans. Med. Imaging*, vol. 35, no. 8, pp. 1962–1971, 2016, doi: 10.1109/TMI.2016.2529665.

[64]   N. Alsubaie, N. Trahearn, S. E. A. Raza, D. Snead, and N. M. Rajpoot, "Stain deconvolution using statistical analysis of multi-resolution stain colour representation," *PLoS One*, vol. 12, no. 1, pp. 1–15, 2017, doi: 10.1371/journal.pone.0169875.

[65]   F. Pérez-Bueno, M. Vega, V. Naranjo, R. Molina, and A. K. Katsaggelos, "Fully automatic blind color deconvolution of histological images using super gaussians," *Eur. Signal Process. Conf.*, vol. 2021-Janua, no. 1, pp. 1254–1258, 2021, doi: 10.23919/Eusipco47968.2020.9287497.

[66]   F. Pérez-Bueno *et al.*, "A TV-based image processing framework for blind color

deconvolution and classification of histological images," *Digit. Signal Process. A Rev. J.*, vol. 101, p. 102727, 2020, doi: 10.1016/j.dsp.2020.102727.

[67]   F. Perez-Bueno, M. Vega, V. Naranjo, R. Molina, and A. K. Katsaggelos, "Super Gaussian Priors for Blind Color Deconvolution of Histological Images," *Proc. - Int. Conf. Image Process. ICIP*, vol. 2020-Octob, pp. 3010–3014, 2020, doi: 10.1109/ICIP40778.2020.9191200.

[68]   F. G. Zanjani, S. Zinger, B. E. Bejnordi, J. A. W. M. Van Der Laak, and P. H. N. De With, "Stain normalization of histopathology images using generative adversarial networks," *Proc. - Int. Symp. Biomed. Imaging*, vol. 2018-April, no. Isbi, pp. 573–577, 2018, doi: 10.1109/ISBI.2018.8363641.

[69]   S. Roy, A. kumar Jain, S. Lal, and J. Kini, "A study about color normalization methods for histopathology images," *Micron*, vol. 114, pp. 42–61, 2018, doi: 10.1016/j.micron.2018.07.005.

[70]   D. Tellez *et al.*, "Whole-Slide Mitosis Detection in H&E Breast Histology Using PHH3 as a Reference to Train Distilled Stain-Invariant Convolutional Networks," *IEEE Trans. Med. Imaging*, vol. 37, no. 9, pp. 2126–2136, 2018, doi: 10.1109/TMI.2018.2820199.

[71]   S. Roy, A. kumar Jain, S. Lal, and J. Kini, "A study about color normalization methods for histopathology images," *Micron*, vol. 114, no. March, pp. 42–61, 2018, doi: 10.1016/j.micron.2018.07.005.

[72]   S. Kothari *et al.*, "Automatic batch-invariant color segmentation of histological cancer images," *Proc. - Int. Symp. Biomed. Imaging*, pp. 657–660, 2011, doi: 10.1109/ISBI.2011.5872492.

[73]   D. Sebiskveradze *et al.*, "From preprocessing to fuzzy classification of IR images of paraffin embedded cancerous skin samples," *WHISPERS '09 - 1st Work. Hyperspectral Image Signal Process. Evol. Remote Sens.*, pp. 53–56, 2009, doi: 10.1109/WHISPERS.2009.5289025.

[74]   A. Kolaman and O. Yadid-Pecht, "Quaternion structural similarity: A new quality index for color images," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1526–1536, 2012, doi: 10.1109/TIP.2011.2181522.

[75]   A. Basavanhally and A. Madabhushi, "EM-based segmentation-driven color standardization of digitized histopathology," *Med. Imaging 2013 Digit. Pathol.*, vol. 8676, no. Figure 1, p. 86760G, 2013, doi: 10.1117/12.2007173.

[76]   B. M. Priego-Torres, D. Sanchez-Morillo, M. A. Fernandez-Granero, and M. Garcia-Rojo, "Automatic segmentation of whole-slide H&E stained breast histopathology images using a deep convolutional neural network architecture," *Expert Syst. Appl.*, vol. 151, Aug. 2020, doi: 10.1016/j.eswa.2020.113387.

[77]   Y. Zheng *et al.*, "Feature extraction from histopathological images based on nucleus-guided convolutional neural network for breast lesion classification," *Pattern Recognit.*, vol. 71, pp. 14–25, 2017, doi: 10.1016/j.patcog.2017.05.010.

[78]   A. Mahbod, G. Schaefer, I. Ellinger, R. Ecker, A. Pitiot, and C. Wang, "Fusing fine-tuned deep features for skin lesion classification," *Comput. Med. Imaging Graph.*, vol. 71, pp.

19–29, Jan. 2019, doi: 10.1016/j.compmedimag.2018.10.007.

[79]   A. Golatkar, D. Anand, and A. Sethi, "Classification of Breast Cancer Histology Using Deep Learning," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10882 LNCS, pp. 837–844, 2018, doi: 10.1007/978-3-319-93000-8_95.

[80]   L. Hou, D. Samaras, T. M. Kurc, Y. Gao, J. E. Davis, and J. H. Saltz, "Patch-Based Convolutional Neural Network for Whole Slide Tissue Image Classification," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 2424–2433, 2016, doi: 10.1109/CVPR.2016.266.

[81]   K. Roy, D. Banik, D. Bhattacharjee, and M. Nasipuri, "Patch-based system for Classification of Breast Histology images using deep learning," *Comput. Med. Imaging Graph.*, vol. 71, pp. 90–103, 2019, doi: 10.1016/j.compmedimag.2018.11.003.

[82]   G. Campanella *et al.*, "Clinical-grade computational pathology using weakly supervised deep learning on whole slide images," *Nat. Med.*, vol. 25, no. 8, pp. 1301–1309, 2019, doi: 10.1038/s41591-019-0508-1.

[83]   O. Jiménez del Toro *et al.*, "Convolutional neural networks for an automatic classification of prostate tissue slides with high-grade Gleason score," *Med. Imaging 2017 Digit. Pathol.*, vol. 10140, p. 101400O, 2017, doi: 10.1117/12.2255710.