Whole Slide Image Quality in Digital Pathology: Review and Perspectives

Romain Brixtel, Sebastien Bougleux, Olivier Lézoray, Yann Caillot, Benoit Lemoine, Mathieu Fontaine, Dalal Nebati, Arnaud Renouf

...

Digital Object Identifier ...

Whole slide image quality in digital pathology: review and perspectives

ROMAIN BRIXTEL[1], SÉBASTIEN BOUGLEUX[2], OLIVIER LÉZORAY[2] (Senior Member, IEEE), YANN CAILLOT[1], BENOÎT LEMOINE[1], MATHIEU FONTAINE[1], DALAL NEBATI[1] and ARNAUD RENOUF[1],

[1] Datexim, 24-26 rue Saint Martin, F-14000 Caen, FRANCE (e-mail: {romain.brixtel,yann.caillot, benoit.lemoine,mathieu.fontaine,dalal.nebati, arnaud.renouf}@datexim.ai)

[2] Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, F-14000 Caen, France (e-mails: {sebastien.bougleux,olivier.lezoray}@unicaen.fr)

Corresponding author: Romain Brixtel (e-mail: romain.brixtel@datexim.ai).

This work was supported in part by the Région Normandie and BPI France through a PTZI for Datexim's R&D team.

ABSTRACT With the advent of whole slide image (WSI) scanners, pathology is undergoing a digital revolution. Simultaneously, with the development of image analysis algorithms based on artificial intelligence tools, the application of computerized WSI analysis can now be expected. However, transferring such tools into clinical practice is very challenging as they must deal with many artifacts that can occur during sample preparation and digitization. Therefore, the quality of WSIs is of prime importance, and we propose a review of the state-of-the-art of computational approaches for quality control. In particular, we focus on WSI quality issues related to the presence of sample preparation artifacts, compression artifacts, color variations, and out-of-focus areas. An analysis of the monthly WSI clinical routine in a cytological laboratory confirms the importance of implementing quality control measures. Given this observation, we draw perspectives on how a computational quality process can be included in a computational pathology diagnosis pipeline.

INDEX TERMS Digital Pathology, Whole Slide Image, Computer-Aided Diagnosis, Computational Pathology, Quality Control, Artifacts.

I. INTRODUCTION

DIGITAL Pathology (DP) is the crossroads between the pathology world and the digitization revolution. Pathology, from pathos (πάϑος) and logia (λογία) respectively "suffering/experience" and "study of" is a medical science that involves the study and diagnosis of disease through the examination of surgically removed organs, tissues, and bodily fluids. Because each pathology subdomain (histology, cytology, etc.) needs adequate expertise to analyze dedicated samples, DP was first seen as an answer to fulfill the local lack of health specialists. As depicted at the First International Conference on Image Management and Communication in Patient Care [1]: "static and dynamic imaging in pathology represent ways to address the problem of maldistribution of specialty pathology services and to provide primary pathology diagnostic services in rural areas. These technologies are also valuable for consultation and educational programs. Because of the high information density in pathology specimens as compared with radiograms, image storage and image transport in pathology represent special challenges to video and communications engineers."

The first wave of changes affecting pathology began with the development of the Internet and the appearance of affordable scanners to digitize pathological samples. Whole slide images (WSIs) are numerical objects that are made to be displayable and shared through networks, see Preston *et al.* [2] for a vision from the 1980s. WSIs allow the screening of optical data at different magnifications and are described by a family of file formats that ease the automation of specific tasks (locating regions of interest, labeling a slide, etc.). The second wave in which DP was installed as a legitimate domain was the increase in available annotated data in the early 2000s. Digital pixel data enriched with expert labels allowed DP to benefit from machine learning (ML) tools and deep learning (DL) by approximately 2010 year. Nowadays definition of DP has followed these different evolutions, as depicted in "Introduction to Digital Pathology and Computer-Aided Pathology" [3]: "DP, which initially delineated the process of digitizing WSIs using advanced slide scanning technology, is now a generic term that includes Artificial Intelligence (AI)–based approaches for detection, segmentation, diagnosis, and analysis of digitalized images".

The main goal of DP, through different periods, is to investigate and study a digital object with the same diagnostic properties as the original biological sample. DP was first involved in easing the delivery of diagnostics by connecting specialists. Then it mutated to computer-aided diagnosis (CAD) to add valuable information to experts. Regardless of the case, adding the digitization step to the standard pathology pipeline involves certain requirements for the subsequent stages. Is the quality requirement to diagnose a slide the same for an AI as for a human expert? Does digitization embeds sufficient information to obtain an equivalent diagnostic for both a sample and its digital representation? What are the requirements to obtain acceptable digitization of a slide, and are these requirements the same for a human expert and an AI?

In a pragmatic way of thinking, a laboratory entering the DP realm should expect an additional Quality Control (QC) step to ensure that each slide and its digital clone are well prepared. Janowczyk *et al.* [4], highlighted the QC need for laboratories diving in a DP world: "Manual review of glass and digital slides is laborious, qualitative, and subject to intra- and inter-reader variability. Therefore, there is a critical need for a reproducible automated approach to precisely localize artifacts to identify slides that need to be reproduced or regions that should be avoided during computational analysis". Thus, when considering the issue of quality in DP, two main related topics emerge:

- QC in laboratories: computer-aided verification of failures in the digital pipeline to check whether each WSI meets the diagnostic requirements according to its domain (histology, cytology, etc.) and to the laboratory (specific machines preparing slides and scanner),
- Quality of a representative corpus from a CAD perspective: This ranges from the versatility of learned models for machine learning engines to the ability to process and detect out-of-scope or "never seen before" manifestations (rare visual anomalies induced by any prior steps of the laboratory pipeline).

Figure 1 depicts the trends in DP scientific publications that involve WSIs processing. The same rise in publications can be observed for either generic DP or WSI related papers or QC in DP, highlighting the interest in this emerging field. It should be noted that pathology subdomains do not receive the same attention, with many more studies focusing on histology rather than cytology. In addition, there is some confusion between pathology and histology: many articles of DP use the term pathology instead of histology even if all the considered examples are from this subdomain. In addition, even in specific categories, the level of interest is different, focus detection and color adaptation are the two main tasks in which QC is mainly involved.

This study focuses on QC issues in DP, driven by a laboratory point of view: how to guarantee the quality of WSIs or associated labeling information, how to detect a failure in a pipeline when analyzing a WSI, which decision to take when a quality dysfunction is detected, etc. are some of the questions we address here.

In the next sections, we run through different stages of DP, from sampling to diagnosis. We elaborate on some technical aspects addressed in the literature, from focus, blur, or artifact detection to color deviation. We end this survey with a review of quality issues of a 1-month stream of WSIs in a DP laboratory to confront the previously exhibited ideas with reality. Last sections provide perspectives and a conclusion.

II. DIGITAL PATHOLOGY PIPELINES

To understand the challenges in DP, we begin with a description of the processes involved in pathology laboratories, with a digital accent, to understand the impact of QC at different stages.

A. LABORATORY POINT OF VIEW

To obtain a specific accreditation that allows them to operate, laboratories must follow guidelines certified by a dedicated consortium (European cooperation for Accreditation[1], International Laboratory Accreditation Cooperation[2], International Accreditation Forum[3], etc.). Accreditation organisms, based on norms (e.g., NF EN ISO 15189) with specifications relative to the countries where they are applied, guarantee the technical and diagnostic reliability of the pathology examinations occurring in a laboratory. These guidelines include pre-analytic, analytic, and post-analytic risk management and require laboratories to analyze each step of their processes according to: what has already happened (retrospective analysis), what occurs currently (ongoing quality measures, user returns), what could happen (risk analysis) and what has happened in other laboratories (bibliographical and clinical studies). To show just one example, here are some requirements for medical Pathology Services from the Australian Department of Health [5][4]: SB8.4 Acceptable test performance must be confirmed by the ongoing use of internal quality control material, SB8.8 Uncertainty measurement must be estimated for each test procedure where relevant and possible, SB8.9 The Medical Pathology Service must have evidence that its uncertainties measurements meet clinical requirements.

From sampling to diagnosis, the biological material goes through many different and entangled stages; the firsts are made to prepare the next stages until an expert diagnoses the sample. These stages differ according to the pathological sub-domain involved. However, the pipelines are broadly the same in both histology and cytology, with samples collected, registered in the laboratory management system, prepared, disposed on a slide, stained, preserved, digitized, and analyzed. However, significant differences can be noticed when

[1] https://european-accreditation.org/
[2] https://ilac.org/
[3] https://iaf.nu
[4] https://www1.health.gov.au/internet/main/publishing.nsf/Content/health-npaac-publication.htm

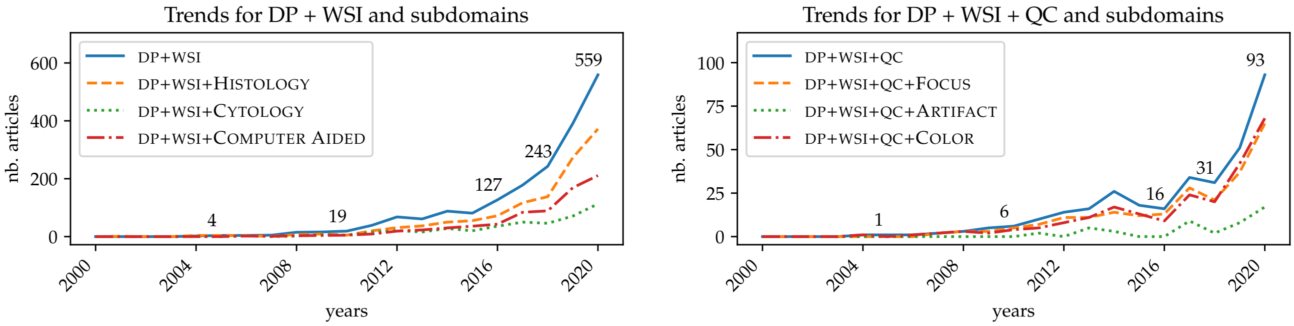

Figure 1: Google scholar trends for publications involving Digital Pathology (DP), Whole Slide Images (WSI), Quality Control (QC) and specific subdomains.

preparing the sample, which can lead to various quality and performance issues (discussed later in Section III).

Laboratories that have completed their digital transition can screen slides on monitors and be helped by various artificial intelligence assistance tools. Therefore, precautions should be taken by the DP laboratory. Indeed, the digitization, storage, and computation steps have minimal requirements to operate as intended and these should be controlled. Digitization is located at the end of the pipeline, and its outputs (pixel data in the WSI and other meta-data information) are directly used for the diagnosis of an AI or a pathologist. We describe the main steps in various DP pipelines to highlight the potential quality issues that impact diagnosis.

B. FROM SAMPLE TO SLIDE

Biological samples to be examined by pathologists must follow a specific sample preparation pipeline because of the requirements of brightfield microscopy. In particular, the requirements for a sample to be examined under the microscope are as follows: i) the sample is well preserved, ii) the sample is transparent so that light can pass through it, iii) the sample is thin so that a single layer of cells is present and iv) some components of the sample can be distinguished by different colors. Because of these requirements, the preparation pipelines differ in histology and cytology. In addition, these differences lead to specific quality issues according to the specificity of the step in which they occur. Figure 2 depicts the typical issues in each preparation step for both histology and cytology. Some errors due to the failure of the information system management are included; they have been addressed in the early ages of DP. Aspects that alter the final diagnosis are the main topics discussed here.

In histology, the following steps are performed [6], as depicted in the left column of Figure 2:

- Tissue collection: A specimen is collected by a surgeon who removes a piece of tissue from the body, and a sample is cut from it.
- Tissue processing: To preserve the cells of a sample, it is fixed with a chemical product, placed into a cassette, processed (dehydration, clearing), and then embedded in paraffin wax.
- Tissue sectioning: The paraffin block is cut into thin sections (3-10 microns) to allow visualization through a microscope and placed on glass slides.
- Tissue staining: slides are stained to reveal the structural details of the tissue sample, and covered with a glass cover slip.

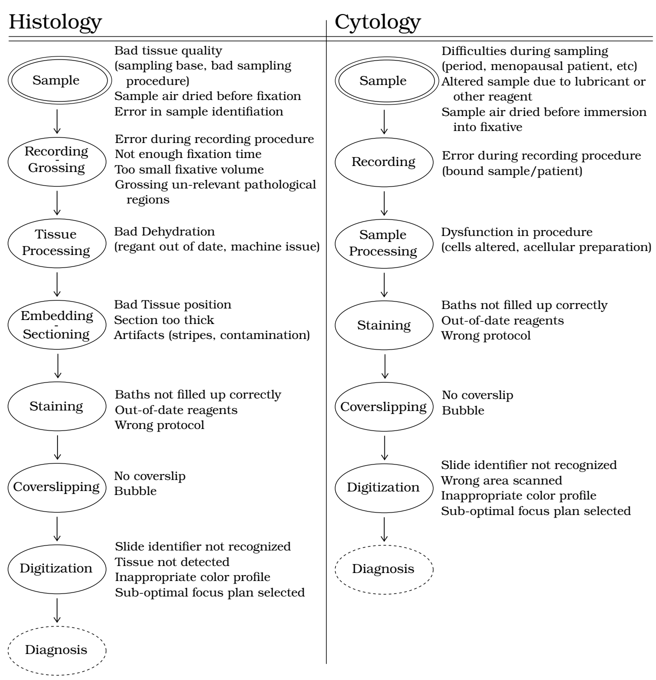

Figure 2: Histology and Cytology pipeline steps and their related issues

In cytology, the following steps are performed [7], see Figure 2, right column:

- Fluid collection: A biofluid is extracted from a patient and placed into a tube.
- Fluid preparation: Cells are processed and displayed on a glass slide. It can include centrifugation or filtration

steps to concentrate the cells in the tube.

- Fluid spreading and staining: cells are distributed within a disk on a glass slide, and the slides are stained and covered with a glass cover slip.

Therefore, preparing a histological or cytological sample for microscopic examination is a technically complicated step that is rather sensitive. In addition, it requires specialized equipment and expertise. Even if we consider a specific slide preparation pipeline (e.g., using either centrifugation or filtering with cytological samples), differences between laboratories can appear. Any problem that occurs during the preparation of the slides can have an impact on the quality of the prepared samples, which is of utmost importance for the final diagnosis taken by the pathologist. In the next section, we focus on the heart of the DP pipeline: the digitization step and its related constraints.

C. DIGITIZATION OF SLIDES

Once the slides are prepared, they can be transformed into digital slides using WSI scanners. There are many different types of WSI scanners. Table 1 provides a (non-exhaustive) list of the most common ones. WSI scanners have two

Company	WSI scanner
3DHistech	Pannoramic SCAN II, 250 Flash
Hologic	Genius
Hamamatsu	NanoZoomer RS, HT, S360, S210, S60 and XR
Huron	TISSUescope 4000, 4000XT, HS
Leica	ScanScope AT, AT2, CS, FL, SCN400, GT450
Olympus	vS120-SL
Omnyx	vL4, vL120
OptraScan	OS 15, OS-Ultra (80, 160, 320, 480), OS-SiA
Perkinelmer	Lamina
Philips	Ultra-Fast Scanner, iSyntax
Sakura Finetek	visionTek
Unic	Precice 500, Precice 600x
Ventana	iScan Coreo, iScan HT
Zeiss	Axio Scan.Z1

Table 1: List of whole slide image scanners.

major components. The hardware components include a microscope with lens objectives, a light source (bright field or fluorescent), robotics to load and move glass slides, digital cameras for (line or tile) image capture and a built-in computer. The software components include digital slides I/O, management, and visualization. Farahani *et al.* [8] and Patel *et al.* [9] reviewed the different properties of WSI scanners including slide capacity, image magnification and resolution, file format, and scan speed. Some scanners can digitize a slide at several focal planes, which can be of interest for cytological slides that are thicker than histological slides. With many different WSI scanners available using different technologies, the digitization of slides can produce very different results. Consequently, the parameters of each WSI scanner component can impact the quality of the WSI and therefore its further interpretation. In addition, as pointed out by Ogura *et al.* [10] the same slide repeatedly scanned with the same WSI scanner will not produce completely identical WSIs. As such, the FDA has designated WSI systems as class III (highest risk) medical devices. When using a WSI, validation must be performed first to ensure that the diagnostic performance based on digitized slides is at least equivalent to that of glass slides and light microscopy [11]. Pantanowitz *et al.* provided guidelines on this topic [12]. They have analyzed the impact of differences between digitized slides. Then, WSI scanners can be validated through stability measurement with regard to the quality of the information captured in the WSI. Shi *et al.* [13] concluded that one should use $20\times$ magnification scans for diagnostic workouts and $40\times$ for challenging cases. In [14], after a review of different studies led on eight different WSI scanners, the mean diagnostic concordance of WSI and light microscopy (LM), weighted by the number of cases per study, was established at 92.4%. A similar finding was reported by Rajaganesan *et al.* in [15], with four different scanners. The diagnostic accuracy of LM was 95.44%, and that of WSI was 93.32%. They also reported rates from other studies in the literature and the diagnostic concordance rate was always $> 96\%$. Interestingly, this study also estimated that the mean digital image artifacts (out of focus and stitching) appearance rate was 6.8% (we will discuss this later in the paper). Given these studies, digitization produces digital slides that do not seem to affect diagnostic performance as long as their quality is sufficient. However, as the visual quality of the digital slides (in terms of color and introduced digital artifacts) can vary significantly among WSI scanners, this must be carefully monitored. In particular, only two scanners from Philips™ and Leica™ companies received FDA approval for the review and interpretation of digital surgical pathology slides prepared from biopsied tissue. This shows that obtaining WSIs of good visual quality is still a challenge.

D. COMPUTATIONAL PATHOLOGY

With the advent of WSI scanners, the development of image analysis algorithms based on artificial intelligence tools, and an increase in the computational power of computers, the application of computerized image analysis to WSIs can now be expected. In 2011, Fuchs *et al.* introduced the term "computational pathology" (CP) [16] in a journal special issue dedicated to WSI processing [17]. This was the first appearance of this term and it encompasses all approaches that make use of AI methods on WSIs to analyze patient samples [3]. However, what gain can be provided to pathologists with an automated system using computational pathology? First, the samples analyzed by pathologists are most often benign and can be easily distinguished from cancerous ones. This is a potentially huge waste of time, and any system that could help pathologists in localizing cancerous areas in slides would be beneficial [18]. Therefore, computational pathology can be useful for computer-aided diagnosis. Second, it can be used to predict disease outcome and survival. In particular, grading approaches have been established for prostate and breast cancers (Gleason and Elston/Ellis) which are correlated with patient outcome and long-term survival. Computational pathology can be very interesting in providing such a grade estimation based on a quantitative analysis of the slide biological structures. From all these potential outcomes, we can roughly say that computational pathology [19] aims at using computational methods to analyze patient specimens (and especially WSIs) for the study of disease [3], [20]. Currently, these computational methods are often referred to as machine learning or artificial intelligence methods [21] and particularly to deep learning methods based on neural networks. In [22], Janowczyk *et al*. showed in tutorial cases how deep learning can be used for computational pathology tasks: segmentation (nuclei, epithelium and tubule), detection (lymphocytes and mitosis), and classification (lymphoma sub-type). This paper is important to the community as it has shown how pathologists can benefit from such computational approaches. In [23], a similar approach was adopted to demonstrate the incorporation of AI and machine learning tools into clinical oncology. Many approaches have been proposed so far for the processing and analysis of WSIs. Recent approaches use Vision Transformers and Multiple Instance Learning (e.g., [24], [25], to quote a few). A review of such approaches is beyond the scope of this paper. We refer the reader to recent comprehensive reviews for more insights on AI and deep learning in computational pathology [26]–[29], with specific reviews for histology [30]–[34] and cytology [35]–[37]. If some ethical issues have appeared in the use of computational pathology in clinical routine [38], most pathologists are in favor of their use [39].

At this time, most of the proposed computational pathology tools are restricted to research use only (RUO) and can only be used to provide a complementary analysis to that performed by a pathologist. In [40], [41], it was shown that the combination of computational pathology and human pathologists has the potential to improve accuracy and efficiency in gastric cancer diagnosis. However, with the use of computational pathology, remarkable progress has been made beyond RUO. For the Gleason grading of prostate cancer [42]–[44] and the Elston/Ellis grading of breast cancer [45]–[47], commercial solutions recently reached the market[5,6]. Similarly, for cancer screening in cervical cytology, some commercial solutions have been proposed[7,8]. One can see that computational pathology enables the addressing of many tasks such as tissue detection, segmentation, and classification with a very broad focus from computer-aided decision (that aims to assist pathologists in routine diagnosis) to precision medicine (that aims to study the clinical outcome of a patient). However, this is at the cost of collecting large sets of manually annotated WSIs to train deep learning models efficiently.

[5]https://ibex-ai.com/solutions/the-galen-platform/
[6]urlhttps://www.paige.ai/hospitals-and-labs
[7]https://www.hologic.com/hologic-products/cytology/geniustm-digital-diagnostics-system
[8]https://datexim.ai/cytoprocessor/

E. QUALITY CONTROL IN DIGITAL PATHOLOGY

With the development of DP and CP, it becomes now commonplace to acquire digital slides and to analyze them with AI tools. However, new problems have appeared that are important to overcome for routine clinical practice. Indeed, slide preparation can introduce many artifacts that can impact the readability of the slide. This is also the case with slide digitization (e.g., with blurred areas). Both these kinds of artifacts can be annoying for either a visual analysis by a pathologist or an automated analysis by an AI tool. As a consequence, some approaches have been developed to minimize the appearance of these artifacts or to cope with them by automated detection or correction. As previously mentioned, clinical laboratories have to follow guidelines for quality control and both DP and CP have to conform to these to be usable in practice. As a result, quality control is an emerging topic in DP and CP [26], [35], [48]–[50]. In the next sections, we review the approaches proposed to assess the quality of WSIs. As both the slide preparation and slide digitization can have an impact on the final WSI quality, we analyze both separately. Finally, we present approaches for quality control at the slide level, and we provide recommendations for improving the quality of slides considered as unanalyzable by computational pathology pipelines.

III. QUALITY OF SLIDE PREPARATION

In a digital pathology flow for sample analysis and diagnosis, the first factor that can strongly impact the quality of a WSI (the digital item) is related to the quality of the slide preparation (the physical item). As slide preparation is very technical and sensitive, mistakes can be made and artifacts can appear. In histology and cytology, an artifact is a structure which should not be present in living samples. In some situations, the presence of an artifact can compromise an accurate diagnosis with an examination under a microscope [51]. Consequently, artifacts due to slide preparation can also cause potential mistakes in quantitative analysis involving the processing of WSIs. As the preparation of slides can be very heterogeneous between and within institutions, artifacts are inevitable. What type of artifacts can be encountered? There is no perfect answer to that question as an artifact is in essence an item that was not foreseen to appear. In a very detailed technical note, Leica™ presented a list of artifacts that can be annoying for the interpretation of (digital or not) slides. As several artifacts are very rare, we present the most common ones grouped into two categories.

- **Domain-independent artifacts** (in histology or cytology) can be generated from the sample preparation: (*i*) The sample can have been contaminated with an unexpected biological element (e.g., blood, bacteria, mucus), or a foreign (non-biological) object (e.g., surgical contaminant). (*ii*) The glass slide can be dusty or dirty, affecting its background. (*iii*) Air bubbles can appear during the glass cover slip. (*iv*) Staining is a critical step, it can cause severe visual color artifacts if it is not well monitored (e.g, by checking the validity of the coloration chemicals). Artifacts can also be generated after sample preparation. For example, pathologists often use markers on glass slides to identify specific areas for analysis. These markers can mask biological components.

- **Domain-dependent artifacts** are due to the specific characteristics of the specimens and their preparation. In histology, thin tissues can fold on themselves [52], Yagi *et al.* [53] showed that there is a correlation between the quality of slide preparation (the thinner the tissue, the better) and the quality of WSIs. In cytology, if the centrifugation is not well performed, many clusters will appear, preventing their interpretation [54]. In addition, in cytology, slides are usually thicker [55], which can lead to focus problems (discussed in the section on digitization).

Figure 3 illustrates such preparation artifacts that affect the quality of the slide. As we just have seen it, artifacts that occur during slide preparation can have a significant impact on the accuracy of slide analysis [58], whether this analysis is performed by a human pathologist or a computer-aided decision system. This has received attention from the computational pathology community only recently [48], [59]. Recent works proposed to perform quality control (automatic or not) of slide preparation to identify the type and/or severity of artifacts from WSIs. The output of such systems can then be used as additional input for computer-aided decision systems to establish a more accurate analysis of WSIs. We review some of these artifacts' detection methods in the sequel. At this level, we focus only on artifacts related to slide preparation. Digitization artifacts will be further considered (even if some of the cited works also consider digitization artifacts such as out-of-focus areas). They are summarized in Table 2, specifying the pathology domain, artifact types, and detection methods.

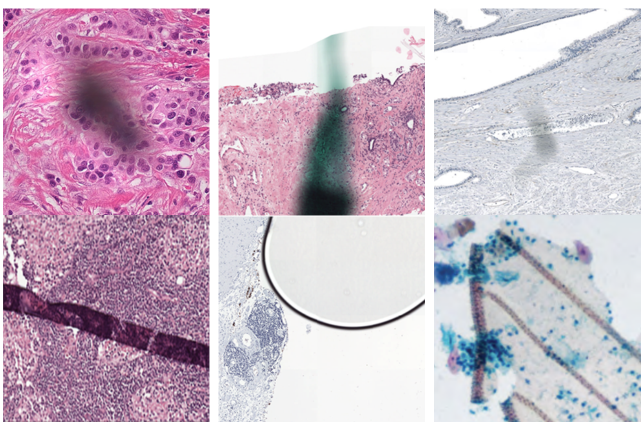

Figure 3: Different types of artifacts: foreign object, marker, dust, tissue fold, air bubble, biological contaminant. Samples are from the dataset https://grand-challenge.org/algorithms/quality-assessment-of-whole-slide-images-through-a/ [56] and [57].

In [60], Avanaki *et al.* proposed automatic quality estimators by adapting image quality assessment (IQA) methods that were originally developed for natural images. In particular, they considered IL-NIQE, a no-reference IQA, to detect artifacts and showed that the scores provided by this estimator enable discrimination of artifacts following the ratings given by a pathologist. In [61], Kothari *et al.* proposed a method for identifying tissue-fold artifacts in histological WSIs. Assuming that tissue folds are different from the other slides' areas w. r. t. color saturation and intensity, they considered the difference between both. The difference is then thresholded at different values and the distribution of the remaining connected components is studied to define two thresholds that enable the extraction of tissue-folded regions. A similar approach was proposed in [62]. In [63] Palokangas *et al.* proposed the extraction of folds using k-means clustering in a saturation-intensity feature space. In [4], Janowczyk *et al.* proposed HistoQC[9], an open-source quality control tool for digital pathology slides. This tool uses a combination of handcrafted features extracted from a digital slide (related to color, brightness, contrast, and edges) that can be fed to machine learning techniques for artifact extraction. The software can then be used to identify artifacts and artifact-free areas. In [64], Chen *et al.* used HistoQC for quality control of renal biopsy WSIs. They aimed at identifying the slides that were unsuitable for computational analysis because of the presence of artifacts. They used different modules of HistoQC to identify artifacts such as tissue folds, pen markers, air bubbles, and ink stain variations. They showed that HistoQC could identify batch effects in the slides' cohorts of three pathological laboratories and concluded that a quantitative process is necessary for robust and reproducible quality control of digital slides. Kumar *et al.* [65], proposed a method for the identification of artifacts in cytological cervical smears WSIs. Cells that are not epithelial are considered as artifacts: e.g., blood cells. The cells were extracted using classical image processing methods. Features (describing intensity, shape, and texture) are extracted from the segmented cells and the latter are classified as artifact or non-artifact cells with an SVM. They obtained an accuracy of 86.64%. In [66], [67], Shakhawat *et al.* proposed the detection of artifacts (air bubbles and tissue folds) in histological WSIs at low resolution, as the considered artifacts can be seen at low magnification and have visual properties that are very different from the surrounding tissue. They extracted luminance and saturation features, and Haralick gray-level co-occurrence matrix texture features from 100×100 patches. Only relevant features were retained using sequential feature selection and fed to an SVM classifier. They obtained an accuracy of 98.98%. Smit *et al.* [56] proposed a multi-class deep learning model[10] for the semantic segmentation of artifacts caused by tissue folds,

[9]https://github.com/choosehappy/HistoQC
[10]https://github.com/DIAGNijmegen/pathology-artifact-detection

	Pathology domain		Type of Artifact						Type of method		
Reference	Histology	Cytology	Air bubble	Marker	Ink	Tissue fold	Foreign object	Biological contaminant	Image processing	Machine learning	Deep learning
[60]	✓		✓			✓			✓		
[61]–[63]	✓					✓			✓		
[4]	✓			✓	✓	✓			✓	✓	
[64]	✓		✓	✓	✓	✓	✓		✓	✓	
[65]		✓						✓		✓	
[66], [67]	✓		✓			✓				✓	
[56]	✓		✓	✓	✓	✓					✓
[68], [69]	✓		✓	✓	✓	✓	✓				✓
[70]	✓					✓					✓
[71]	✓			✓							✓
[72]	✓				✓						✓
[73]	✓		✓	✓	✓	✓	✓				✓

Table 2: Computational methods for the detection of artifacts in WSIs.

ink, air bubbles, dust, and markers for histological WSIs of multiple tissues and staining types. The network architecture was an encoder-decoder network with EfficientNet-B2 as the encoder and DeepLabV3+ as the decoder. The semantic labeling is performed on 1024×1024 patches. They obtained an accuracy of 89.45%. In [68], Foucart *et al*. proposed a deep residual network for artifact detection in H&E and IHC stained WSIs. Their approach relies on rough annotations and works at a low resolution of the WSI pyramid to enable a fast analysis. Semantic labeling was performed using 128×128 patches. They also considered a specific data augmentation technique as artifact areas are much less present than non-artifact areas. They obtained an accuracy of 89.77%. They recently extended their results in [69] and showed that a deep learning approach to artifact segmentation[11] can produce interesting results as long as learning strategies are adapted to dataset characteristics. In particular, artifacts in digital pathology slides are ill-defined objects, which makes them particularly challenging to annotate precisely, and their work addresses this aspect of imprecise annotations[12]. Babaie *et al*. [70], proposed the use of a pre-trained DenseNet201 CNN as a feature extractor to characterize tissue folds, and fed an SVM classifier with these features. The DenseNet201 CNN was fine-tuned on a dataset of folded and fold-free 255×255 patches along with data augmentation. They obtained an accuracy of 96.7%. Ali *et al*. [71] proposed a fully automatic CNN for the classification and restoration of WSIs containing pen ink markers[13]. First, a CNN detects tiles corrupted by a pen marker. Second, in these corrupted tiles, a Yolo CNN detects the bounding boxes of pen marker areas. Third, corrupted pixels are restored by a domain-adaptive cycle-consistent-adversarial generative model. Their approach can produce visually coherent marker-free WSIs while enhancing their quality (as assessed using PSNR, SSIM, and VIF IQA measures). In [72], Zhang *et al*. proposed a method for assessing the staining quality of WSIs for Gram staining. They considered a MobileNet CNN to estimate the staining quality of the slides' tiles. They obtained an accuracy of 86.8%. The results were shown to pathologists in the form of a stain-quality heat map. In [73], [74], Haghighat *et al*. proposed a multi-task ResNet18 deep neural network to determine the presence of artifacts on 228×228 patches at low resolution[14]. From the patch-level artifact estimation statistics, they also provided a slide-level "usability" index that estimates wether the slide is appropriate for establishing a clinical diagnosis and an indication of the impact of artifacts on WSI quality. This can help pathologists determine whether the slide needs to be re-scanned or re-stained. They obtained an accuracy of 98.7%.

From these representative studies on artifact detection in WSIs, we observed that this problem has rarely been addressed in the literature. Most recent studies rely on deep learning and obtain very good detection results. However, the integration of the artifact detection result within a computer-aided diagnosis system still needs to be explored.

IV. QUALITY OF SLIDE DIGITIZATION

A. QUALITY OF WSI FORMAT

Many commercial scanners have appeared in the market and each has introduced its proprietary file format. Consequently, there is no established file format for storing and exchanging WSIs produced by WSI scanners (see Table 3). However, these proprietary file formats can share similar properties. Indeed, given the flexibility of the TIFF file format, many have adopted it to store WSI data, often in the form of a tiled multi-resolution pyramid (Figure 4). Nevertheless, this abundance of proprietary file formats can be a strong barrier to their use even if some formats are very similar. Indeed, as we will point it out later in this review, deep learning methods for computational pathology require large and diverse datasets from different centers potentially using different scanners. Therefore, the quality of the WSI file format is an important factor for its use in the digital

[11] https://github.com/adfoucart/deephisto
[12] https://zenodo.org/record/3773097
[13] https://github.com/sharib-vision/histopathology-inkRemoval
[14] https://github.com/MaryamHaghighat/PathProfiler

Company	WSI file format	Compression
3DHistech	.mrxs	Jpeg, JpegXR
Hamamatsu	.vms,.ndpi	Jpeg, Jpeg2000
Hologic		Jpeg
Huron	.tif	Jpeg, Jpeg2000
Leica	.svs, .scn, .cws	Jpeg, Jpeg2000
Olympus	.vsi	Jpeg, Jpeg2000
Omnyx	.jp2, .rts	Jpeg, Jpeg2000
OptraScan	.otiff	Jpeg, Jpeg2000
Perkinelmer	.qptiff	LZW
Philips	.tiff, iSyntax	Jpeg, Jpeg2000, LZW
Sakura Finetek	.svslide	Jpeg
Unic	.tmap	Jpeg
Ventana	.bif,.tif	Jpeg
Zeiss	.czi	Jpeg, Jpeg2000, LZW

Table 3: Different file formats of WSI scanners.

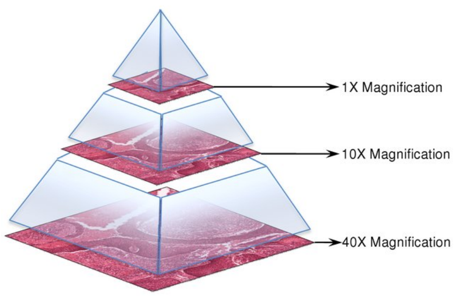

Figure 4: A WSI pyramidal file [80]

computational pathology analysis pipeline. For instance, the macro-image (that represents a snapshot of the entire glass slide) provides a low-magnification overview of all of the tissue pieces. It can be used to guide the scanner's tissue detection system or for focus-point selection. As reported in [75], [76], an incorrect macro-image can generate technical problems such as automatic tissue detector failure, or poor scan coverage. Even if a standard specification for WSI data has been published by the Digital Imaging and Communications in Medicine (DICOM) Working Group [77], its adoption is rather limited. Consequently, reading and writing WSIs can be technically difficult, even if some open-source libraries have made significant progress [78], [79]. The quality of the WSI file format, in terms of interoperability, must be considered when choosing a WSI scanner. In addition, the quality of a digitized slide can be altered by many factors which we review in the sequel.

B. COMPRESSION

A parameter that can impact image quality is the use of compression. Indeed, all WSI scanners use lossy compression to ensure reasonable file sizes. Typically, image compression is measured using the quality factor (QF) or compression rate. An uncompressed image has a QF of 100. If the scanner uses a QF that is too low, compression artifacts will appear, which are mostly visible as block effects. Figure 5 presents examples of such compression artifacts on histological and cytological images. As shown, a QF that is too low can severely affect the image content making it more difficult to diagnose.

Some studies investigated the optimization of compression standards for WSI scanners. Sharma *et al.* [81], considered WSIs obtained from a scanner at different compression rates for 12 different stain types. Then, they determined the most suitable QF for pathologists to be able to perform a diagnosis on compressed images. They concluded that a QF of 50 was suitable for all stains. In [82], Bug *et al.* investigated Scalable High Efficiency Video Coding (SHVC) as a replacement for the JPEG and JPEG2000 standards currently found in most WSI formats. They showed that SHVC can provide a gain in compression performance but introduces blurring artifacts. In [83], Helin *et al.* defined an optimized parameterization for JPEG 2000 image compression specifically used with histopathological WSIs. Their parameterization is based on allowing a very high degree of compression on the background part of the WSI while using a conventional amount of compression on the tissue-containing part of the image.

Although there is no consensus regarding acceptability of image compression levels, JPEG is thought to allow a compression rate between 10:1 and 20:1, and between 30:1 and 50:1 for JPEG 2000, without the loss of diagnostic information [77], [84]. Therefore, lossy compression is not an issue because few visible compression artifacts will appear making the interpretation difficult. However, many studies have used deep learning to analyze WSIs. Compression can be a problem with such methods. In [85], Zanjani *et al.* studied the impact of JPEG 2000 compression on a CNN for detecting tumor metastases in H&E-stained tissue sections. Their experiments showed that the CNN model is robust against a compression rate of up to 24:1 when trained on uncompressed images. In addition, they showed that when the CNN was trained on compressed images, the performance is not much impacted even at high compression rates. Chen *et al.* [86] also investigated the effect of compression on the performance of deep learning approaches for segmentation and detection tasks in WSIs. Their findings are similar: the images can be compressed by 85% while still maintaining the performance of the algorithms at 95% of what is achievable without any compression. In particular, they observed that the minimum acceptable QF for diagnosis by a pathologist corresponded to a decrease in the performance of the deep learning algorithms. These results are in line with the findings of Dodge *et al.* [87] for natural images. They studied the robustness of four state-of-the-art CNNs against five distortions. The CNNs were highly resilient to compression and it was only at very low QF that their performance began to decrease (less than 10 for JPEG and less than 30 for JPEG2000).

From these works, we can conclude that the usual lossy compression employed by whole slide scanners does not have a strong impact on the image quality and its use in machine learning tasks, even at high compression rates. Therefore, fixing the compression rates at established values ensures the visual quality of WSIs [84].

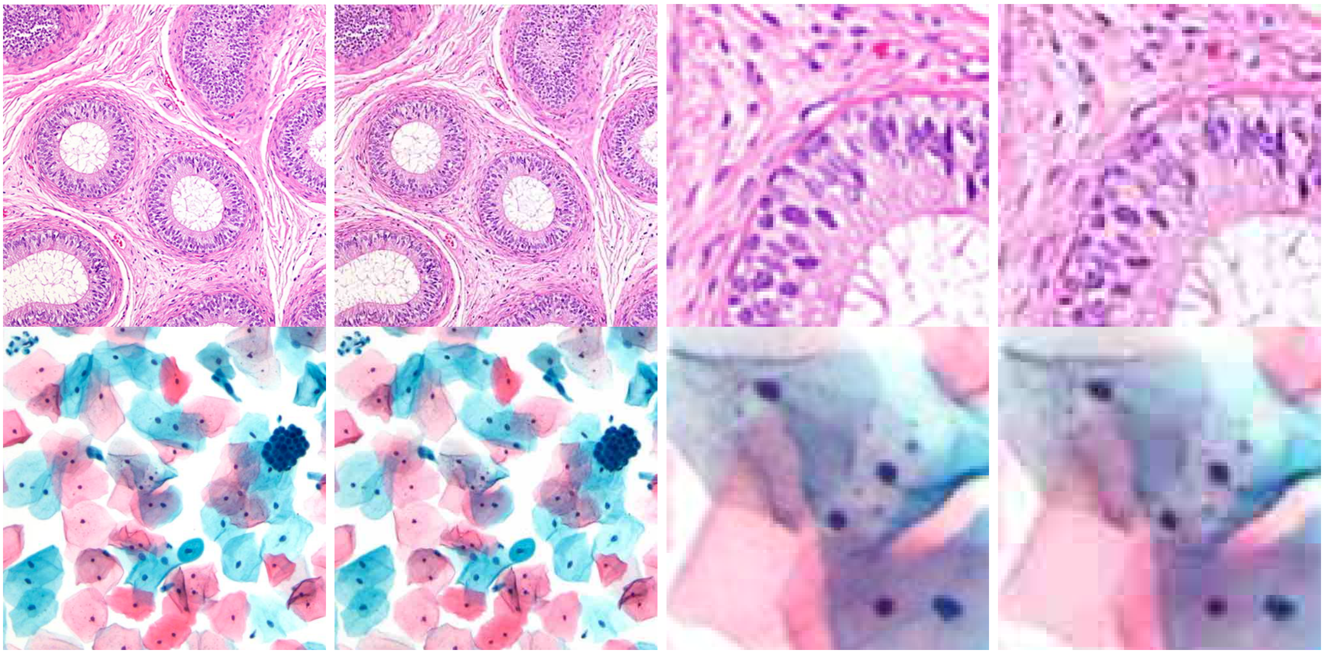

Figure 5: Images with compression artifacts. Each line presents an original uncompressed image, its compressed version in JPEG with loss (compression rate of 5:1), and zoomed and cropped parts from the uncompressed and compressed images. Block effects can be seen in compressed images.

C. COLOR

Assessing and enhancing the color quality of WSIs is probably what has been mostly addressed in the pre-processing of WSIs. Several steps can have a significant influence on color quality in the preparation of digital slides [88]. First, the slides are stained with chemical dyes to highlight the cellular structures and enable their interpretation. If standardized staining protocols [89], [90] can help to reduce variations in staining results, many factors can affect the stain color in practice: the use of different staining equipment, dye brands, and staining protocols. Second, the slides are digitized using WSI scanners that can provide very different results depending on their electronic components and internal color calibration (if any). Consequently, it is inevitable for WSIs to have variations in their color appearance among different institutions, because they use different staining protocols and scanners (see Figure 6). The computational pathology community has embraced this problem and has tried to address it through different means: color calibration or color normalization. We provide a review of the most representative works in this section.

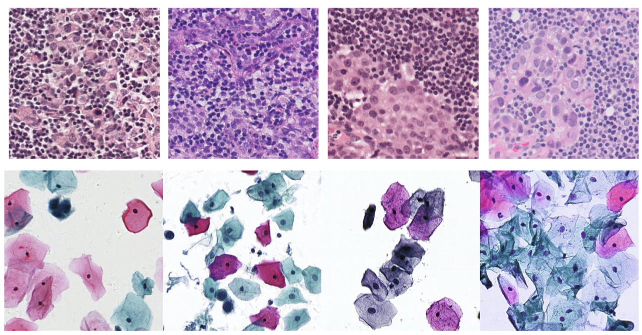

Figure 6: Images with staining from different institutes and scanners (top: H&E histological images, bottom: Papanicolaou cytological images).

1) Color calibration

Color calibration is an established routine in the print and photography industries (to quote a few) that has been adopted by most digital systems with the use of ICC (International Color Consortium) color profiles. The calibration process consists in comparing the known colors of a set of color patches with their digitization from a digital device. The difference between the two can then be used to define a color correction embedded in the ICC color profile. The US Food and Drug Administration has released recent guidance [91] stating the need to develop a method to control color reproduction throughout the digitization process in whole-slide imaging for primary diagnostic use. They stated that color control is essential in digital pathology and recommended the use of a target slide with spectral characteristics similar to those of stained biological components. The ICC Medical Imaging Working Group has started pooling resources to develop a calibration system for digital microscopes[15]. Unfortunately, this is still not standardized among WSI scanners' vendors, and even if

[15] https://tinyurl.com/ysxz6vjv

calibration slides have appeared recently[16,17], their use is not widespread. A general review of this color calibration problem in pathology can be found in [88].

To meet these requirements of color calibration, in [92] Yagi *et al.* were the first to propose a target slide containing nine filters with color patches selected for H&E-stained slides. This has been further explored by Bautista *et al.* in [93] that proposed a color correction procedure based on the comparison between the spectral colors of the patches and their scanned colors. Using this procedure, they have shown that the color difference between two slides scanned from different scanners is significantly reduced after correction. In [94], Chen *et al.* eliminated the need for a color calibration slide by measuring the spectral transmittance of a reference biological tissue sample. However, this is difficult to use in practice. In [95], Shrestha *et al.* proposed an alternative to using a specific color calibration slide. They considered a standard IT8-target transmissive film on a slide and proposed a matrix-based calibration method compliant with the ICC standards, without the use of external color measurement devices such as colorimeters. They showed that with this method there is no visible difference between calibrated slides scanned by different scanners in terms of CIE-Delta2000 JND (Just Noticeable Difference). In [96], a color calibration slide was used to calibrate a WSI scanner for cytology. They showed that this reduces the color variation to less than 2·JND and maintains color fidelity. Therefore, color calibration minimizes system-to-system variability, producing repeatable color output regardless of system age or optical component variation.

At this time, even if color calibration should be a requirement of any whole slide scanner, this is still not the case. A recent work by Ogura *et al.* [10] has shown that this can be a strong problem as several digitization of the same slide with the same scanner and the same parameters provide slightly non-identical images. Although the fact that color calibration from reference targets (such as MacBeth or IT8) are established methods in displays, print, and photography, the lack of such an established reference target slide is still a barrier to color calibration in digital pathology.

2) Color normalization

Color normalization is the transformation process from one image to another which affects the colors. There are various algorithms for color normalization, most of which were presented in [97], [98]. Color normalization methods can be categorized into three categories from the most ancient to the most recent: global normalization, stain separation, and generative model-based approaches. All the methods are summarized in Table 4 including the pathology domain, the type of staining under consideration, and the category of the method for color normalization.

[16]https://tinyurl.com/55xnd2as
[17]https://ffei.ai

Global color normalization methods apply either histogram matching or color transfer. In histogram matching, a normalized histogram is computed, its probability distribution function (PDF) is estimated and the PDF of the image is matched to a reference image [99]. Because histogram matching on entire images ignores local differences in image content, colors associated with one stain may be matched to irrelevant colors. In color transfer, a statistical analysis is used to impose the characteristics of a reference image on other images. In [100], Reinhard *et al.* matched the mean and standard deviations of reference and query images in Lab color space. Because images stained by multiple chemical dyes may have different color distributions, colors associated with different biological components may blend after color transfer. To address this problem, before the color transfer, an image can be divided into different regions. Magee *et al.* [101], applied a Gaussian mixture probabilistic model for automatic segmentation at the pixel level in three classes: Hematoxylin (H), Eosin (E), and background (B) which was followed by color transfer on the corresponding classes from the reference to the query image. Nevertheless, color transfer methods cannot ensure that the structural features of biological components are preserved.

Stain separation methods aim at estimating the main stain vectors of an image that can be used for stain intensity correction and stain replacement. Generally, these methods compute stain vectors using Singular Value Decomposition (SVD) or Non-negative Matrix Factorization (NMF), as the stain concentration cannot be negative. Both involve factorization of the optical density (OD) space matrix. In [102], Ruifrok *et al.* proposed that each pixel color can be represented as a linear combination of different stains using a stain matrix. They proposed a method called color deconvolution, that decomposes the optical densities of stain mixtures into stain-specific channel information. The method is supervised as the OD values of the pure stains must be provided in the form of a stain matrix that describes how color is affected by stain concentration. In [103], Macenko *et al.* proposed an algorithm using plane fitting with SVD to determine the stain matrix from the stain vectors of an image. Prior knowledge of the stains can also be used in the plane fitting process as proposed in [104]. In [105], [106][19], CNNs were used to estimate stain vectors faster. Gupta *et al.* [107][20] performed the color normalization in three steps: illumination correction, stain color vector correction from an alignment and a rotation of their SVD estimation, and stain quantity correction. Khan *et al.* [108] learned an image-specific stain color matrix from a color-based classifier using a stain color descriptor. A non-linear mapping of the channel statistics obtained after color deconvolution enables the reconstruction of a normalized image. Kather *et al.* [109][21] proposed an optimized variant

[19]https://github.com/Zhengyushan/ssc
[20]https://github.com/shivgahlout/GCTI-SN
[21]https://zenodo.org/record/35083

	Pathology domain		Type of staining				Category of method		
Reference	Histology	Cytology	H&E	IHC DAB	Jenner-Giemsa	Pap	Global	Stain separation	Generative
[99]	✓		✓				✓		
[100]	✓	✓	✓	✓	✓	✓	✓		
[101]	✓		✓	✓			✓		
[102]	✓		✓	✓				✓	
[103]	✓		✓					✓	
[104]	✓		✓					✓	
[105]	✓		✓					✓	
[106]	✓		✓					✓	
[107]	✓	✓	✓		✓			✓	
[108]	✓		✓	✓				✓	
[109]	✓			✓				✓	
[110]	✓		✓					✓	
[111]	✓		✓	✓				✓	
[112]	✓		✓					✓	
[113]	✓		✓					✓	
[114]	✓		✓						✓
[115]	✓		✓						✓
[116]	✓		✓						✓
[117]		✓				✓			✓
[118]	✓		✓						✓
[119]	✓		✓						✓
[120]		✓				✓			✓
[121]	✓		✓						✓
[122]	✓		✓						✓
[123]	✓		✓						✓
[124]	✓		✓						✓

Table 4: Computational methods for color normalization of WSIs. The StainLib software [125] [18] contains the popular methods of [100], [103], [111].

of [102] for immunostained images.

To avoid the supervision required to define the stain matrix in previous methods, unsupervised methods have been developed. In [110] Li *et al.* proposed the use of illuminant and spectral estimation from a reference image. For each query image, the image illuminant is estimated and an NMF-based stain spectral estimation is performed with an initialization using saturation-weighted statistics to enable a better convergence. In [111] Vahadane *et al.* performed stain separation to estimate the stain matrix by an NMF that incorporates a sparseness constraint followed by a structure-preserving color normalization. Lei *et al.* [112] proposed an improvement to this method by using CNNs to estimate the stain matrix instead of a sparse NMF. These methods do not require prior information and preserve the structure of the original image. However, they do not preserve all the color information of the source images. Bejnordi *et al.* [113] used color and spatial information to classify the image pixels into different stain components. The chromatic and density distributions for each of the stain components in the hue-saturation-density color model were then aligned to match the corresponding distributions of a reference.

Generative model-based methods have recently attracted considerable attention. They use adversarial learning of deep neural networks to perform a style transfer to a WSI. In addition to being more adaptive than the global color normalization and stain separation approaches, they suppress the need of selecting reference images to define the stain parameters. Indeed, generative adversarial network (GAN) [126] based methods consider the overall dataset of the target style as the template and approach the problem of color normalization as an image-to-image translation. GAN-based color normalization approaches can be divided into supervised and unsupervised methods. Supervised methods require paired images of different staining protocols and use L1 and adversarial losses to optimize the generative networks. For instance, in [114][22], Salehi *et al.* proposed the use of conditional GANs where the generator is trained to generate restained images conditioned by input gray-scale images. However, paired images of different styles require multiple staining which is difficult to perform in practice. In contrast, unsupervised methods do not require paired images and are much more appealing. Cho *et al.* [115][23] proposed a stain-style transfer GAN composed of two transformations: gray-level normalization to have a laboratory-independent image, followed by a colorization that fits the stain-style of a chosen laboratory. The latter used a specific loss that combines reconstruction loss, conditional GAN loss, and

[22] https://github.com/pegahsalehi/Stain-to-Stain-Translation
[23] https://github.com/hanwen0529/DSCSI-GAN

feature-preserving loss to ensure the preservation of the extracted features (essential for subsequent analysis of the normalized image). In [116][24], Shaban *et al*. proposed a method called StainGAN that uses Cycle-Consistent Adversarial Networks (CycleGAN) for one-to-one domain stain transfers. Cycle consistency allows the images to be mapped to different color models but preserves the structures of biological components. This method was also used by Runz *et al*. [127]. In [122], Lee *et al*. proposed an extension of [115], [116] that uses UNet as the generator for better structure preservation, and a Markovian discriminator with local receptive fields. To improve the quality of the generated image, they also introduce a color classifier that provides feedback to the generator on the normalized color content. Comparisons were led for classification tasks on H&E breast cancer WSIs and the classification performance was better than the compared methods, i.e., [100], [103], [111], [115], [116]. In [117][25], Kang *et al*. proposed a faster alternative to StainGan using a 1×1 convolution. In [118][26], Liang *et al*. proposed a method [27] that uses a different reconstruction loss (based on the use of a structural similarity index matrix and directional statistics-based color similarity index) to better preserve the texture, structure, and color of the biological components. In [119][28] Shrivastava *et al*. proposed a self-attentive adversarial stain normalization approach for normalizing of multiple stain appearances to a common domain. This enables many-to-one domain stain transfer and the feature preserving loss of [116] is replaced by a structural cycle consistency loss. Chen *et al*. [120] proposed to normalize an input image by style removal and reconstruction, as in [115]. Style removal generates a grayscale image using a color-encoding mask. For style reconstruction, the loss contains an intra-domain adversarial loss, an L1 penalty, and an inter-domain adversarial loss. Ren *et al*. [121] considered a Siamese network as the generator to regularize the normalization. In [123], Nazki *et al*. have proposed an unsupervised adversarial network to normalize histological WSIs while preserving the structural features of the tissue. To that aim, a single generator is trained to normalize images. The preservation of the fine salient anatomical structures is performed using an auxiliary feature extraction network and a perceptual loss to minimize the perceptual distance between the normalized and the original images. Experiments were led for color normalization between different WSI scanners. Recently, some authors have proposed performing color normalization with self-supervised techniques. In [124], Zhao *et al*. first perform stain separation to estimate H and E dyes from H&E stained WSIs. Then, a UNet-based network learns how to re-stain the grayscale images in a self-supervised manner using a combination of adversarial, color, and staining losses. Their approach outperforms the methods [100], [103], [111], [116] in terms of image quality assessment measures (SSIM, PSNR, etc.), and of performance for segmentation and classification of H&E stained breast cancer WSIs.

[24] https://github.com/xtarx/StainGAN
[25] https://github.com/khtao/StainNet
[26] https://github.com/hanwen0529/DSCSI-GAN
[27] https://github.com/hanwen0529/DSCSI-GAN
[28] https://github.com/4m4n5/saasn-stain-normalization

3) Influence of color

We reviewed approaches for color calibration and normalization, but how important are these two steps in a computer-aided diagnosis? Some recent studies have attempted to answer that question by studying the influence of color normalization on decision systems. In [128], Leo *et al*. considered H&E stained histological prostate WSIs. They have shown that textural and structural features can become very unstable when scanned on different scanners. In [129], Jia *et al*. have also studied the influence of color on H&E stained WSIs by comparing the color spectrums with an MDS embedding, which enables to easily identify low staining quality and abnormal staining conditions. However, no solution is proposed on how to use this information for quality control of the staining. Liu *et al*. [130] evaluated the degree of color similarity between two images based on the volume of their color gamut. They compared the normalized images obtained using the methods of [100], [103], [111] with a reference obtained by a hyperspectral imaging microscopy system. They observed that the methods in [103], [111] reduced significantly the color gamut and that the one in [100] better preserved the color gamut but was unable to fully preserve the color information. In [131], Ziaei *et al*. proposed an evaluation of the color normalization methods' ability to normalize images obtained from one scanner to match the color rendering from another scanner. The comparison was based on the CIE Delta E color difference. They compared the methods in [100], [103], [111], [116]. Their experimental results showed that color normalization was effectively able to reduce color variation and that StainGAN [116] performed significantly better than the other global and stain separation approaches. However, these two works studied only the color variation of the color normalization methods and not their influence on the subsequent machine learning process. In [132], Aubreville *et al*. showed that the influence of the color domain shift introduced by different scanners strongly affects the performance of CNN-based mitosis detection on histological H&E slides. In [133], Nisar *et al*. have also shown that it is possible to detect and estimate the staining shift in digital histopathology. To that aim, they considered a domain shift metric measuring the differences between two domains' distributions using features extracted from pre-trained neural networks. As attended, this shift can impact the generalization performance. This advocates the use of color normalization. Pontalba *et al*. [134] evaluated the necessity and impact of color normalization for CNN nuclei segmentation methods on histological H&E slides. They considered the color normalization methods of [100], [103], [108], [116]. As in [131], they observed that color variations are less important with StainGAN [116]. For nuclei segmentation, the segmentation performance varied significantly depending on the applied color normalization method. Only the approach in [108] preserved the segmentation performance with respect to those with un-normalized images. In [135], Swiderska-Chadaj *et al.* noticed that for prostate cancer classification, StainGAN normalization can improve CNN robustness as compared to classical color normalization. In [136], Bianconi *et al.* evaluated the effect of color normalization methods [100], [103], [108] on automated classification methods (based on either classical machine learning or deep learning techniques) of histological H&E slides of different types of cancers. Their results showed that in most cases color pre-processing did not improve the classification accuracy and could even result in a noticeable reduction in accuracy. Similar findings were reported by Gadermayr *et al.* [137]. They studied the combination of several color normalization methods [100], [103] with feature extraction methods (Fisher vectors, LBP, color histograms, and a VGG CNN) for the classification of patches in the glomerulus and non-glomerulus tissue with slides stained by alpha-smooth muscle actin or periodic acid Schiff staining. They observed that the use of color normalization always causes a loss in accuracy. Tellez *et al.* [138] studied the effect of color normalization [103], [113], [116] on the performance of several different CNN classification tasks on histological H&E slides. They also observed that color normalization was not necessary to achieve better performance. Hameed *et al.* [139] have considered six intermediate layers of the pre-trained Xception model to extract features for the classification of histological H&E breast cancer images. None of the normalization methods they considered [100], [102], [103], [111] was able to outperform the results of the original un-normalized dataset. However, with GAN-based approaches, the classification results of the unnormalized and normalized images can be very similar. In [140], Ciompi *et al.* investigate the influence of color normalization methods [103], [113] on tissue classification of colorectal cancer tissue samples in H&E-stained images. In contrast to the previously mentioned studies, they reported a significant gain in performance with the color normalization of [113]. Finally, in [141], the influence of several color normalization algorithms [100], [103], [106], [108], [111], [113] has been studied for the classification of three different histological cancer H&E WSIs (ovarian, breast, and pleural) using a ResNet18 network. Their finding is that color normalization does not improve performance when WSIs are from the same center. However, when normalized datasets from several centers are used for learning, an increase of classification performance can be obtained.

In conclusion, color normalization is important for obtaining images of similar colors for analysis by pathologists. However, it is still not demonstrated that color normalization can enhance the performance results of a computer-aided system, even if the recent GAN-based approaches for color normalization appear very promising [142], [143] for learning stain invariant features to improve the generalization of CNNs.

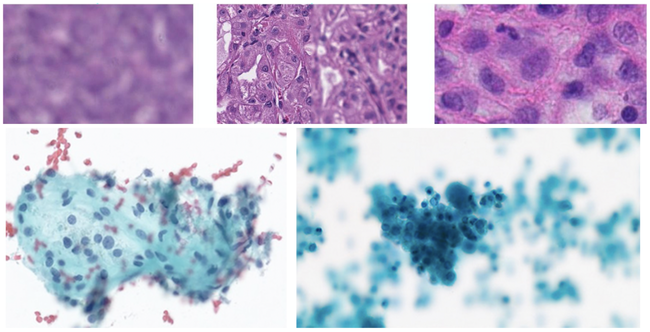

Figure 7: Different types of blur (global, regional, local) that can be encountered in histology (first row) or cytology (second row).

D. OUT-OF-FOCUS

Despite a controlled environment and an autofocus process, scanners can produce blurred WSIs, either globally, regionally, or locally (see Figure 7). This is mostly caused by a poor focusing on the objects of interest [8], known as the out-of-focus (OOF) effect. During acquisition, several problems may appear and contribute to this effect, such as thermal variations, internal or external vibrations, errors in the focus determination at a focus point, or in the generation of the WSI focus map (interpolation from the focus points). Some of these errors are caused by preparation issues, such as tissue folds, bubbles, dirt, or the distribution of objects along the z-axis in liquid-based preparations. Specific tissue types, such as fat, may also affect slide sections.

The objective of automated focus quality assessment (FQA) is to: i) determine whether a slide must be rescanned, either locally, regionally, or globally, and ii) provide an FQA map for visual inspection and possible weighting of further processing steps (a kind of confidence map). To perform this, FQA methods locally estimate a focus class (or score) at the patch or tile level, with high magnification levels (usually $20\times$ or $40\times$). The local scores are directly used to define the pixel values of an FQA map, usually in the form of a heatmap (as shown in Figure 8). Eventually, regional or global focus scores can be deduced from the local ones. Some works have also proposed to directly modify the acquisition process to perform learned auto-focusing in WSI scanners, either for histological [159] or cytological [160] samples. All the FQA methods are summarized in Table 5 including the pathology domain, the type of staining, the

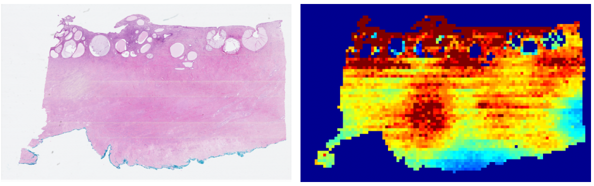

Figure 8: Focus Heat Map (from [148]) with a color range from blue to red corresponding to low to high focus quality.

	Pathology domain		Type of staining			Data used for training and Testing					Category of method				Transfer Test	
Reference	Histology	Cytology	H&E	IHC	Other	Magnif. N×	≠ scanners	Data MAN	STK	SYN	#Focus Classes	Type IP	HFC	LFC	Other Stain	Other Scan.
[144], [145]	✓		✓			$1.25\times - 20\times$		✓			2	✓				✓
[146]	✓		✓	✓		$20\times$, $40\times$				✓	2	✓				✓
[147], [148]	✓		✓		✓	$40\times$			✓		2		✓			
[149]	✓		✓			$20\times$				✓	2		✓			
[150]	✓			✓		$20\times$, $40\times$		✓	✓		2		✓			
[151]		✓		✓		$20\times$		✓			2		✓			
[152]	✓		✓	✓		$20\times$		✓			2		✓		✓	
[153]	✓		✓		✓		✓			✓	6		✓			
[154]	✓		✓			$20\times$				✓	2			✓		✓
[155]	✓		✓	✓		$40\times$			✓		2			✓		✓
[156]	✓		✓	✓		$40\times$	✓			✓	30			✓		✓
[157]	✓		✓		✓	$40\times$			✓		12			✓	✓	✓
[158]	✓		✓		✓	$40\times$			✓		12			✓		

Table 5: Computational methods for the detection of out-of-focus areas in WSIs.

data used, the number of focus classes, the category of the method for OOF estimation (IP: Image Processing, HFC: Handcrafted Feature Classification, LFC: Learned Feature Classification), and possible transfer test with other stains or scanners.

The FQA methods used either during or after digitization are essentially similar. They are inspired by auto-focusing techniques and both determine a local quality score. However, images after digitization usually contain some degradation due to post-processing, e.g.,, JPEG compression. This section focuses on FQA methods at the patch or the tile level once the WSI has been acquired.

1) Image Processing methods

The firstly proposed OOF estimation works used classical image processing tools. In [161], Walkowski *et al.* considered Haralick contrasts and entropies. In [144], [145], Ameisen *et al.* considered maximum local variations on the brightness. In [162], Zerbe *et al.* considered the Tenenbaum gradient. Once these locally computed features are obtained, thresholds are used for binary blur scoring. In [146], Jimenez *et al.* conducted an individual analysis of such features. They suggested predicting a patch as OOF if two of the four features predict it as OOF with a manually adjusted threshold for each feature. Obviously it is very difficult to establish manual thresholds that will perform accurately in all focus configurations. As a consequence, works have rapidly shifted towards learned predictions to better adapt to image focus variations. However, this requires annotated datasets.

2) OOF datasets

To be able to learn to detect OOF areas, FQA research works have considered annotated datasets. The latter can be obtained by different techniques (resulting in different dataset types). The most commonly employed techniques to label the focus quality of patches in WSIs are:

- Fully manual annotation (MAN): patches are extracted from the WSI and labeled by human experts.
- Automatic z-stacks (STK): patches extracted from real z-stacks with an offset value from the in-focus plane, or an absolute z-level from this location.
- Semi-synthetic z-stacks (SYN) simulate the previous one by applying several degrees of synthetic blur to the in-focus patches.

The labels associated to the patches are usually binary (in-focus or out-of-focus) but in some works [153], [156], [157] several levels of focus have been considered. FQA methods designed on subjectively labeled data (MAN) should provide a response closer to manual FQA according to the assessor cohort, which is usually able to distinguish a maximum of six OOF grades. Because the creation of MAN datasets is tedious, thus limiting their size, they are mainly used to evaluate the performances of FQA methods optimized with objectively labeled data (STK and SYN). The latter data, related to auto-focusing and z-stacking, are easier to create and overcome the limitations of the former (size and dependence on assessors). Note that STK and SYN are not completely objective (manually calibrated and checked), and SYN is an augmentation technique. While Gaussian blur is commonly used, Bokeh blur ($2D$ Heaviside step function) is known to be closer to the perception of real OOF occurring in photography and microscopy. As observed in [156], the human perception of blur follows an exponential relationship with the OOF level rather than a linear one. In particular, Gaussian blur tends to underestimate strong OOF which is less the case for the Bokeh blur model.

Even if labeled datasets are mandatory to design a learned FQA method, still few datasets are available. FocusPath[29] [147], and its extended version[30], contain 864 (resp. 8640) 1024×1024 pathological images with 16 absolute z-levels as scores (STK), and acquired with Huron TissueScope

[29] https://sites.google.com/view/focuspathuoft/
[30] https://zenodo.org/record/3926181

LE1.2 at $40\times$. TGCA@Focus[31] [157], selected from the Cancer Genome Atlas [163], contains 14 371 manually annotated patches (MAN, 2 classes) from 52 organ types. Both of them are restricted to histology.

3) Handcrafted feature classification

In [150], Gao *et al*. proposed the first work on learned WSI FQA from handcrafted features. They used 44 standard measures of image quality as features (neighborhood contrasts, derivative-based, local intensity statistics, wavelet-based), which have already been identified as good candidates for characterizing the OOF effect in the context of microscopy and auto-focusing [164], [165]. Then, an AdaBoot-based binary classifier is used to estimate if a patch is in or out of focus. They considered MAN and STK datasets [150], both at $20\times$ and $40\times$, and trained their binary AdaBoost classifier based on 44 features for the four datasets independently. For the MAN dataset, the classification accuracy was higher at $20\times$ with a large improvement (approximately 5%). For the STK datasets, it was higher at $40\times$ with a slight improvement ($<1\%$). Globally, STK achieved better performances than MAN (>92.7 and just above 91% respectively.). The individual behavior of these handcrafted features was further analyzed by Moles Lopez *et al*. [152]. They observed that the discriminatory ability of these features varies from one biological structure to another for the same stain (most importantly for IHC), and from one stain to another. To cope with this, they considered several binary classifiers based on height discriminatory features and decision trees (DT), trained either on a stain-specialized dataset (H&E or IHC) or on a mixed version, with objective scores (MAN). All three learned models achieved an accuracy of at least 96%. To test the transfer ability of these three learned models, they tested them on an independent MAN dataset, and the performances dropped by up to 89%, particularly for classifiers that rely on IHC. As classification accuracy was reduced for the mixed version in any case, mainly owing to the presence of IHC patches, they retained the stain-specialized versions with a reduced number of four features (Haralick features, mean gradient magnitude, Tenenbaum gradient, and noise used in [149]) and decision tree depths (approximately seven). In the context of liquid-based cervical cytology, Lahrmanb *et al*. proposed a binary classifier for patches containing cells automatically selected using Otsu segmentation and HSV analysis [151]. The classification is based on a SVM and five handcrafted features expressing the mean quantity of edges and local variations, and the differences between sharpened, smoothed, and blurred versions of the patches. High accuracy and sensitivity ($>98\%$) were obtained for the MAN dataset. Inspired by previous studies, Campanella *et al*. [153] considered a Random Forest (RF) model with 13 features to predict a blur class (among six). Based on a feature selection on a Gaussian-SYN dataset, a RF regression model composed of 19 trees and 10 features was retained.

More recently, Hosseini *et al*. designed several sharpness measures (HSV-MaxPol [147] and FQPath [148]) based on a symmetric FIR kernel that mimics the ability of the human visual system to boost high-frequency domain magnitudes in a balanced manner. It is defined as a superposition of multiple even-order derivative kernels, either fixed (HSV-MaxPol) or to fit the inverse PSF of the scanner optic (using the Born & Wolf model) up to a threshold frequency (FQPath). By using the FocusPath dataset (STK) to tune the parameters, good correlations with the FocusPath scores were obtained by HSV-MaxPol and FQPath compared to several state-of-the-art measures, close to the behavior of Maximum Local Variations (MLV) [166], for a reduced computation time. Similar results were obtained for the binary classification [157] (ROC$\simeq$0.94, PR$\simeq$0.97). However, when using a different dataset (MAN) to test the transfer ability (other scanners and stain-tissue types), the performances were considerably reduced and were different for HSV-MaxPol, FQPath, and MLV ($0.56<$PR<0.67). This confirms the observations made in previous studies on the transfer ability of handcrafted features.

4) Learned feature classification

During the past five years, several data-driven methods based on CNNs have been proposed for FQA of digital microscopy images [167] and WSI to improve transfer ability. They can be divided into three main categories: i) standard architectures eventually adapted with minor adjustments, ii) truncated at a lower level, and iii) architectures developed specifically for WSI FQA. Standard architectures, which are usually pre-trained with natural images are all re-trained for WSI FQA. Campanella *et al*. provided the first example with ResNet-18 [153], but restricted to grey-level patches for comparison to RF. In Contrast, the CNN models considered in all the other studies take RGB color patches as input data.

Senaras *et al*. proposed DeepFocus[32] [155], a binary classifier based on five convolutions and three max-pooling layers to extract the features, and two fully-connected layers followed by a Softmax and manual threshold to determine the OOF class. Trained by using an STK dataset and categorical cross entropy as a loss function, DeepFocus showed high classification accuracy (93.2%) for an independent dataset (from the same scanner and stains, H&E and IHC), and confirmed that H&E-stained tiles are much easier to classify than IHC ones ($<90\%$ for 2 slides among 3). Kohlberger *et al*. proposed ConvFocus [156], which is a similar network that distinguishes 30 OOF levels. The network is a truncated Inception V3 architecture composed of six layers to extract the features (three conv. + max pooling + average pooling + dropout), and a fully connected layer followed by a SoftMax. ConvFocus was trained using cross entropy with two Bokeh-SYN datasets, coming from

[31] https://zenodo.org/record/3910757

[32] https://github.com/cialab/deepfocus

two different scanners, and augmented by various transformations (orientation, brightness, contrast hue, saturation, translational jitter, Poisson noise, and JPEG compression). It was tested using two MAN datasets (similar scanners, same stains, six grades). The predictions were highly correlated with the expert scores, mainly for one scanner (SRCC approximately 0.8 and 0.93 resp.).

Wang *et al.* proposed FocusLiteNN[33] [157], a shallow CNN that provides a sharpness score from a single convolution layer (with k filters for each color channel) and non-linear pooling as an activation function (linear combination of the extremum feature values). FocusLiteNN was compared to the knowledge-based methods discussed in [148], including FQPath, HSV-MaxPol [147], [148], and MLV [166], as well as five standard CNN architectures similar to those considered in previous works (DenseNet-13, EONSS, and different configurations of ResNet). All CNNs were trained on FocusPath using their initial loss function (PLCC for FocusLiteNN). Performances on FocusPath and a TGCA@Focus datasets were always higher for CNN-based models (ResNet-10 performed best), with interesting transfer ability ($0.97 < \text{PR} < 0.99$ for FocusPath and $0.9 < \text{PR} < 0.87$ for the TGCA@Focus dataset). Differences between shallow and deeper CNN architectures were only slight, while shallow CNN improved computation time by a large margin. Albuquerque *et al.* [158] performed a comparison using FocusPath to train seven other standard CNN architectures (MobileNet_V2, AlexNet, GoogleNet, ResNet18, VGG16, truncated ShuffleNet, and SqueezeNet) for a classification task (12 OOF levels). Contrary to Wang *et al.* [157], data augmentation (shifting, zooming, flipping, and rotation) was applied. As most previously tested models were trained according to a non-ordinal loss function (typically cross-entropy), Albuquerque *et al.* [158] also considered five different ordinal loss functions (ordinal encoding, binomial unimodal, regularized cross-entropies, ordinal entropy). The performances for FocusPath, w. r. t. classification accuracy, MAE, and Kendall's τ measure showed that cross-entropy was the second-best option. MobileNet_V2 with ordinal encoding was the best performer and improved the correlation (SRCC, PLCC) with the FocusPath scores previously obtained by Wang *et al.* [157] for all CNNs with a reasonable computation time compared with FocusLiteNN.

5) Discussion

To conclude this tour of WSI FQA, best performances can be obtained by CNN-based methods trained on semi-synthetic or stack datasets. However, most methods are dedicated to specific stains and scanners. The results on transfer ability are encouraging for subjectively assessed data containing other stain-tissue types and acquired by different scanners, even for very shallow networks [157]. While shallow networks seem to provide a more appealing ratio between computation time and classification accuracy than handcrafted feature-based methods, their behavior for specific stain-tissue types (e.g., IHC) and cytology images has been rarely studied. The datasets used for learning or evaluation are generally free of artifacts owing to slide preparation or manipulation. Therefore, even if the quality of the images produced by autofocus and FQA during digitization is becoming higher and higher, FQA after digitization is still necessary. Another important issue concerns the decision at the slide level from OOF estimation at the patch level for automatic rescan decisions, which has only been addressed in a few studies, for example [144], [151], as described in the following section.

V. QUALITY DRIVEN BY DIAGNOSIS

With the advent of computational pathology, computer-aided (CAD) systems are now able to assist pathologists in establishing their daily diagnosis, for example for tumor detection or cancer grading. However, if these systems can reach the performance of pathologists for some diagnostic tasks, very few can be considered as being directly clinically applicable in clinical practice [168], [169]. Indeed, when such algorithms, which were developed and validated in pure research settings, are applied in routine diagnostics, they face many variations that can have occurred at each step of the digital slide preparation. As we have seen in the previous sections, both the quality of slide preparation and digitization can introduce severe artifacts. If the latter can be annoying for a visual examination by a pathologist, they will represent a potential failure case for computer-aided systems, as they will have never learned to recognize these before. For example, Wright *et al.* [170] showed that the quality issues of digital slides can have a strong impact on the performance of classification algorithms. Some recommendations have recently been released to facilitate the implementation of computational pathology workflows in pathology laboratories [171]. Standardizing the sample preparation in laboratories is of course a solution to such issues. The CAP NSH WSI Quality Improvement Program[34] is an initiative toward this purpose. Labs can have the quality of their histological H&E WSIs estimated and feedback is provided to help the lab in preventing the appearance of preparation or digitization artifacts. However, standardization will not eliminate the appearance of artifacts, and computational pathology methods must be able to address them. As we have seen in the previous section, methods have been developed to handle these problems of slides' preparation artifacts, staining variations, and out-of-focus areas. These can be considered as the first steps toward quality control of WSIs. They should be integrated within any CAD pipeline in digital pathology to establish if a slide is of sufficient quality to be analyzed, and if not, what countermeasures have to be taken. However, how can we establish such a quality of a WSI towards its analysis? In

[33] https://github.com/icbcbicc/FocusLiteNN

[34] https://www.nsh.org/learn/histoqip

this paper, we consider that a slide is analyzable, and can lead to a trusted CAD diagnostic, if:

- The sample is sufficiently representative according to the analysis requirements (sufficient tissue in histology, enough cells in cytology). For instance, in cytology, the Bethesda System for Reporting Cervical Cytology (TBS) imposes a minimum of 5000 analyzed cells to obtain a sample of sufficient quality. In histology, the Elston-Ellis breast cancer grading requires the number of mitotic figures per 10 consecutive high-power fields.
- No element of the sample is obfuscated, potentially leading to doubt in the diagnostic. In cytology, the TBS imposes that less than 50% of the cells can be obscured.
- If part of the sample is obfuscated but an abnormality is detected in a clear area, then the slide can lead to a diagnosis if no other clues are needed.

Indeed, even if a slide contains preparation artifacts such as air bubbles, or digitization artifacts such as out-of-focus areas, if the slide contains enough information of good visual quality to establish a diagnosis, the slide can be considered as "analyzable". The methods we reviewed for the detection of artifacts, the estimation of out-of-focus areas, and staining problems provide the results of their analysis in two main forms: i) a heatmap that assigns a quality score to patches of the WSI. Figure 8 presents a heatmap for the estimation of out-of-focus areas. ii) a semantic segmentation map that assigns a label to each pixel (or patch) of the WSI as predefined semantic labels. Figure 9 presents such a semantic segmentation map for the labeling of artifacts and non-artifact areas.

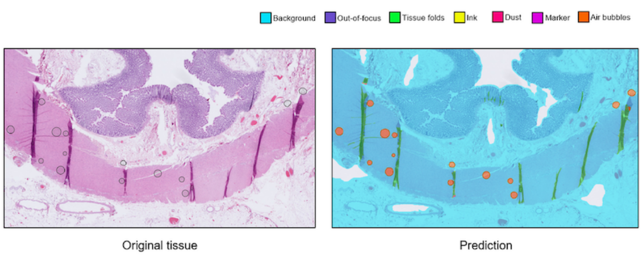

Figure 9: Artifact semantic segmentation map [56] (from https://www.computationalpathologygroup.eu/projects/artifact_detection/)

Both outputs can be used as inputs to computational pathology methods that can integrate this information into their analysis (e.g., by not analyzing identified artifacts). However very few methods have been designed to assess the analyzability level of a whole slide: is it suitable for CAD analysis or should it be re-prepared, re-stained, or re-scanned? These methods are reviewed in the sequel (Table 6 provides a summary).

In [151] Lahrmanb *et al.* proposed an approach for scoring the focal quality of cytological slides. The slide was divided into 16 low-resolution regions from which cells are extracted with Otsu thresholding. A total of 200 cells were randomly chosen and cropped at high resolution for each region. Each cell is described by sharpness image processing features that are fed to an SVM, which classifies each cell as in or out of focus. The percentage of in-focus cells in each region is divided by the number of regions. If this value is lower than a user-defined threshold, the slide has to be re-scanned. Ameisen *et al.* [144], [145] proposed a method to assess the focus quality of a WSI. Blank tiles were excluded based on their color saturation values. The remaining tiles were characterized using image processing methods in terms of sharpness, contrast, brightness, and color. A combination of in-house thresholds is applied to these features to determine whether a tile is of sufficient quality. Depending on the magnification in the tile pyramid, the image is considered sharp if 90% of the tiles are sharp at 2x magnification (70% at 10x magnification). Their results were consistent with those of pathologists for 100 WSIs with various blurred areas. Their work has recently been extended as a software [35] that was presented in [172]. A similar approach was proposed in [173], [174], but without providing a slide-level quality decision. In [153], Campanella *et al.* first performed background detection to avoid blank tiles. A set of 10 sharpness measures were extracted from the patches and fed to a random forest classifier that assigns a blur score to each patch. This method provides a blur heat map as output but also provides a blur slide-level score as the percentage of the blurred surface of the WSI. Zhang *et al.* [72] proposed using a CNN to assess the quality of WSIs stained with Gram staining. They considered a MobileNet CNN to estimate the staining quality of the slide tiles. From the quality of the tiles, they generate assessments (good/average/low) of the slide quality from empirically selected thresholds in terms of staining, density, and artifacts presence. In [66], [67], Shakhawat *et al.* proposed the detection of artifacts (air bubbles and tissue folds) at low resolution in histological WSIs using an SVM trained from handcrafted features extracted on non-blank tiles. The tiles were then classified into three groups (good/average/poor) based on empirical thresholds. The final slide-level quality was estimated as

$$\frac{\#Good}{\#Tiles} \times 0.25 + \frac{\#Average}{\#Tiles} \times 0.50 + \frac{\#Poor}{\#Tiles} \times 0.75$$

where # means "number of". The slide can be proposed to be re-scanned or not depending on the slide quality. In [56], Smit *et al.* proposed a multi-class deep learning model for the semantic segmentation of artifacts. The segmentation output is used by a quality control module in the form of a decision tree and proposes four possible actions: no action (the slide is analyzable), re-prepare (re-cut or clean up), and re-scan. In [73], Haghighat *et al.* proposed the pathProfiler tool for quality control in a large retrospective cohort of prostate WSIs. After extracting the tissue regions to avoid blank tiles, a multi-task deep neural network

[35] https://magiq.xyz/en/pathology/

	Pathology domain		Type of criterion			Type of decision			
Reference	Histology	Cytology	Artifacts	Staining	Focus	Analyzable	Re-prepare	Re-stain	Re-scan
[151]		✓			✓	✓			✓
[144], [145]	✓				✓	✓			✓
[153]	✓				✓	✓			✓
[72]	✓		✓	✓		✓	✓	✓	
[66], [67]	✓		✓			✓			✓
[56]	✓		✓	✓	✓	✓	✓		✓
[73], [74]	✓		✓	✓	✓	✓	✓	✓	✓

Table 6: Slide-level quality control methods.

performs a quality estimation at 5x magnification. With this model, a tile is described by six inferred quality measures assessing usability, no artifact presence, staining artifacts, focus artifacts, tissue folding, and presence of other artifacts (dirt, ink, air bubble, etc.). The tile-level statistics are then aggregated and fed to a fully connected neural network to predict quality at the slide-level. Three slide-level scores are provided to predict the WSI usability, focus, and staining qualities. They also used handcrafted features extracted from HistoQC [4] (this tool does not provide a slide-level analysis) and found that learned deep features perform much better. This work is interesting as it is the only one that: i) considers the usability of WSIs to assess if the slide is appropriate for clinical diagnosis and ii) simultaneously estimates artifact presence, staining problems, and out-of-focus areas. This study is probably the closest to an ideal quality control tool.

Now, if we compare the number of approaches in Table 6 with those presented for artifact detection, staining quality, and focus quality, we can see that very few approaches proposed a global slide-level quality analysis. Indeed, most state-of-the-art approaches we have seen in the previous sections do provide only a quality heat map or a semantic segmentation map, and usually for only one quality criterion.

To conclude, the field of quality control of WSIs is just in its infancy. The main issues are known and are related to artifact detection, staining quality, and focus quality. If some methods have been proposed separately for each quality issue (as exposed in the previous sections), they all have limitations in terms of accuracy and reproducibility on other datasets from other digital pathology centers. In addition, their output is at the moment limited to heat and segmentation maps that can be overlaid on the WSIs but not to evaluate the global slide quality. The two recent approaches of [56], [73] provide good directions towards what the community should converge to: a quality control based on artificial intelligence able to recognize all the quality problems and to decide what to do with the slide from rejecting it, advising re-stain or re-scan, or considering the slide as analyzable by a CAD system.

VI. LABELED DATA QUALITY AND QUANTITY: HOW TO LEARN THE UNEXPECTED

In the previous sections, we reviewed the most recent state-of-the-art approaches for WSI quality control in digital pathology and observed how they can be integrated within a slide-level quality analysis for a computational pathology process. Regardless of the QC to be performed (artifact and out-of-focus detection or stain normalization), the most recent and efficient state-of-the-art methods do rely on deep learning approaches. Many recent reviews of the field of deep learning in computational pathology have been published, and we refer interested readers to these [22], [26], [27], [30]–[32], [35], [36], [36], [47], [175]. If deep learning approaches enable astonishing results that have put forward the field of computational pathology, these techniques require a large number of labeled examples to guarantee good generalization. Generalization is a well-known issue in machine learning. Generalization refers to the ability of a learned algorithm to adapt properly to new and previously unseen data (hopefully drawn from the same distribution as that used to create the model). It has been studied from a theoretical point of view with the help of the VC Dimension [176]. The VC dimension is a measure of the capacity (or complexity) of a set of functions, which can be learned using a classification algorithm. Baum and Haussler [177] proposed that if a generalization level of 90% is desired, the number of training samples should be about 10 times the VC dimension [176]. It has been recently proved [178] that the VC dimension of a multi-layer perceptron of W weights and L layers is $(W \cdot L \cdot log(W))$. For instance, a shallow MLP with an input vector of size 100 to be classified into 5 classes with $L = 2$ layers containing each 50 neurons has $W = \sum_{i=0}^{L}(1 + H_i)H_{i+1} = 7855$ weights (with H_i denoting the number of neurons of a layer) and it will theoretically require more than one million examples to achieve a good generalization. Deep Neural Networks (DNN) such as Convolutional Neural Networks (CNN) are very different and their convergence properties are still not well understood. Recently, in [179] Long and Sedghi presented several theorems bounding the generalization error of a class of CNNs with a high probability by $O\left(\sqrt{\frac{W(\beta+L+log(\lambda))+log(1/\delta)}{n}}\right)$, where W is the number of weights, L is the number of layers (both convolutional and fully-connected layers), β is the distance from initialization in the operator norm, λ is the margin, n is the number of sample data, and the bound holds with a probability of at least $1 - \delta$. Thus, with very deep neural networks, it can be expected that attaining good generalization will require a very huge number of examples if the number of weights and the depth of the CNN are large. As this is usually the case (with the popular ResNet 101, W is more than 44 million with $L = 101$ layers), a key issue of applying DNNs in computational pathology is therefore highly related to the availability of a huge and well-labeled learning dataset.

The quantity and the quality of a learning dataset are also something that has to be taken into account when designing a computational pathology approach for CAD. Indeed, the quality of the dataset used for learning can potentially have a strong impact on the quality of the induced deep learning algorithm. We already addressed this partly when we reviewed the influence of color and focus of WSIs: once a DL algorithm has been trained, it is built only to work on data that are similar to those used to train it. Therefore, if one wants an algorithm to be as versatile as possible, it should have learned to recognize patterns in many different situations. Therefore the evaluation of machine learning algorithms for digital pathology can be very delicate [180]. In [181], Wahab *et al.* have studied, for the classification of different breast cells in H&E WSIs, the quality of annotations in terms of completeness, exhaustiveness, diversity, and agreement. They concluded that standardization of annotation protocols is necessary and proposed a new one. Two direct consequences of annotation quality are that: i) datasets have to be very carefully labeled to have high-quality training corpus (to cope with the lack of quality of labels), and ii) the DL algorithm can use specific learning strategies to make them more robust to variations (to cope with the lack in quantity of data). These two aspects are considered in the sequel.

A. LACK IN QUALITY OF LABELS

In [182], Marée *et al.* summarized the potential data collection issues in digital pathology and proposed guidelines for constructing realistic ground-truth datasets while controlling their quality:

- Examples should come from different staining and scanner equipments.
- Examples should be equally represented in each class.
- Biological variations of the examples should be covered in each category of examples.
- Artifacts should be considered as additional categories to avoid too many false positives.
- Annotations should be performed by several experts.

Unfortunately, very few datasets encompass all these requirements (see in [32] for a list of available cancer histology datasets). Recently, Hosseini *et al.* [183] proposed a new digital pathology dataset called the "Atlas of Digital Pathology". In particular, they demonstrated the quality of their image labels through pathologist validation and by training three state-of-the-art neural networks for tissue type classification. This stresses that obtaining annotations of large cohorts of representative WSIs is a very sensitive step. This is mainly because of the expertise required to generate quality labels and the limited availability of qualified experts. Therefore, pathologists require efficient annotation tools (preferably open-source) that easily enable them to label patterns in WSIs . Hopefully, such software are available (see [184] for a complete review):

- Icy[36] [185].
- QuPath[37] [186].
- Cytomine[38] [187].
- SlideRunner[39] [188].
- Quick Annotator[40] [189]

Regardless of the chosen annotation tool, the quality of the labeled dataset must be carefully monitored, because labeling is usually performed by a single pathologist. This can have a strong influence on the learned deep learning generalization quality and introduce some bias as recent studies have stressed [190]–[192], but this is a well-known issue of any deep learning based computer vision approach [193].

B. LACK IN THE QUANTITY OF DATA

When the objective of a method is to detect rare items in WSIs (such as artifacts), the collection of annotated data is all the more problematic. Indeed, because rare items are by definition not often encountered, the constitution of a representative dataset is almost impossible and this is particularly true for artifact detection in quality control. Therefore, other strategies have been envisioned to address this lack of data. In computer vision, when a dataset is strongly imbalanced or its size is too limited to train a deep learning algorithm, typical strategies exist to alleviate the scarcity of annotated data [194]: data augmentation, transfer learning, domain adaptation, and weakly-supervised learning. Data augmentation [195] artificially generates synthetic data from the initial dataset to enlarge it and improve the performance of the model. Augmentation can be performed using many different transformations, such as geometric transformations or color shifts. Transfer Learning [196] works by training a network on a large dataset such as ImageNet and then using those weights as the initial weights in a new classification task. Transfer learning works only if the data to be processed are similar to have a valid transfer (e.g., natural images). Domain adaptation is a type of transfer learning method. A DL algorithm learns from a source domain with a large labeled dataset and aims at achieving comparable performance for the same task on a target domain with few labeled data [197]. Weakly supervised learning [198] consists in training the models with labels less expensive to collect than image-level annotations (e.g., grades at the slide-level). Such labels are often easier to obtain with limited efforts, and also in

[36] http://icy.bioimageanalysis.org
[37] https://qupath.github.io
[38] https://cytomine.be
[39] https://github.com/maubreville/SlideRunner
[40] https://github.com/choosehappy/QuickAnnotator

large quantities. All these strategies have been explored in computational pathology, an excellent review of which can be found in [30]. We quote only some representative recent works for each strategy in the sequel to provide a rough overview.

1) Data Augmentation

Data augmentation is a frequently used solution to introduce a certain degree of invariance to a CNN and to tackle class imbalance by artificially increasing the learning dataset. In [138], Tellez *et al*. conducted a study on different kinds of data augmentation (rotations, mirroring, scaling, elastic deformation, Gaussian blur and noise, brightness and contrast, Hue-Saturation-Value and Hematoxylin-Eosin-Dab color shifts) and showed that this can lead to performance gains. In particular, stain color augmentation is crucial for achieving the best performance. In [199], Teramoto *et al*. demonstrate a consequent gain in performance with data augmentation for the classification of benign and malignant cells in cytology with a VGG CNN. In [200], Annuscheit *et al*. investigated data augmentation techniques (color-based, geometric-based, filter-based transformations, and erasing) on different datasets using multiple network architectures (VGG, Inception, DenseNet). They observed that geometric-based techniques increase the model performance but color-based augmentations have no significant effect. This result is not in concordance with those of [138], but their approach has strong sensitivity towards data augmentation hyperparameters. To address this problem, in [201] Faryna *et al*. proposed to use automated and computationally efficient data augmentation, as classical data augmentation requires extensive hyper-parameter tuning and can lead to sub-optimal generalization performance. Based on the RandAugment framework, they considered several domain-specific modifications relevant to histopathological images (based on these of [138]). They showed that this automated data augmentation could outperform the approach of Tellez *et al*. [138] where data augmentation was manually tuned.

2) Transfer Learning

Two strategies were considered in transfer learning. The first strategy consists in using off-the-shelf features extracted from a source pre-trained network that are fed to a specific classifier for the target task. The second strategy consists in using a source network with pre-trained weights and fine-tuning the weights on the target domain. In [202], Sharma *et al*. demonstrated the ability of the pre-trained Xception model to perform breast cancer histopathological image classification in contrast to handcrafted approaches. Li *et al*. [203] showed that off-the-shelf features learned from natural images can be reused in computational pathology, but the amount of information that could be transferred heavily depended on the complexity of pathology images. Some papers have studied this transfer ability from general models such as ImageNet in pathology. In [204], Sharma *et al*. compared the performance of features extracted from networks trained on ImageNet and histopathology data. They demonstrated that specific encoders (ResNet) trained on multiple histopathology datasets result in superior features than their ImageNet trained counterparts and should be used for weight initialization in histopathology tasks. Aitazaz *et al*. have recently shown [205] that the vision transformer works better on histopathology images than CNN models pre-trained on ImageNet. In [206], Mormont *et al*. investigated various deep learning transfer learning strategies on histological and cytological datasets. They compared transfer strategies for seven different network architectures. They observed that fine-tuning always outperformed off-the-shelf features from the last layer of the network, regardless of the network. Similar conclusions were drawn by Jang *et al*. in [207] who studied transfer learning between different types of cancers. In [208], Mormont *et al*. investigated multi-task learning as a way of pre-training networks. They gathered 22 digital pathology datasets into a single dataset and used multi-task training. The features extracted from their model are superior to those of ImageNet. This shows that domain-specific pre-training can be an interesting alternative to fine-tuning. Another way is to try to consider the existence of generalizable knowledge between different problems and to use weight distillation [209] for cross-knowledge transfer.

3) Domain adaptation

When transfer learning is considered, the source and target domains are assumed to follow similar distributions. Domain adaptation deals with cases where a model trained on a source distribution is used in the context of a different (but related) target distribution. This problem has been recently addressed by employing adversarial learning: a discriminator is trained to distinguish source and target data using features extracted from a deep neural network as inputs, while the deep neural network is tuned to confuse the discriminator. This helps to map the source and target data close. In [210], BenTaieb *et al*. proposed the use of GANs to learn dataset-specific staining properties to transfer stains across datasets. Fine-tuning the obtained stain transfer network with images from a new domain can enable normalizing training images with respect to the new domain distribution. Ren *et al*. [121], [211] proposed using unsupervised domain adaptation to transfer discriminative knowledge obtained from the source domain to the target domain. Adaptation is achieved through adversarial training to find an invariant feature space in the source domain along with a Siamese network architecture on the target domain to enforce regularity. Shi *et al*. [212] also use stain style transfer to translate the style of a small image dataset into a large dataset style with cycleGAN. Because an inception CNN has been trained on a large dataset, it can be applied to datasets from other centers. In [213], a different approach has been proposed. The authors train a "universal" model to recognize diverse histological tissue types from a source domain dataset of healthy slides from various organs. They can then adapt the model to transfer diagnostically relevant labels for tissue and disease classification into target domains without any re-training or fine-tuning, by using prior histological knowledge.

4) Weakly supervised learning

Building a large dataset of cell-level labels on WSIs is tedious and time-consuming. In contrast, slide-level labels (e.g., grading or diagnosis) are easier to obtain as they are readily available from diagnostic reports. This is the motivation of weakly supervised learning where the slide-level diagnoses constitute weak labels. Recent reviews on that topic can be found in [214], [215]. It is then assumed that if a WSI has a cancer diagnosis, at least one of its tiles must contain cancer cells. An application of this principle, referred to as multiple instance learning (MIL), was performed by Campanella *et al*. [168]. They used a dataset of 44,732 WSIs using only slide-level diagnoses as labels with impressive results on test sets of prostate cancer, basal cell carcinoma, and breast cancer metastases. In [216], Kanavati *et al*. trained a CNN based on the EfficientNet-B3 architecture, using transfer learning and weakly-supervised learning, to predict carcinoma in WSIs. They also compared fully-supervised learning and weakly-supervised learning and demonstrate that: i) fully-supervised learning performs best when cell-level labels are available, and ii) when only slide-level diagnoses are available, weakly supervised learning can be performed but requires a much larger dataset of WSIs. In [217], Teramoto *et al*. developed a weakly supervised method for the classification of benign and malignant lung cells in cytological images using attention-based deep multiple instance learning (AD MIL). Images were divided into patches images and stored in bags. Each bag was then labeled as benign or malignant, and classification was conducted using AD MIL. Their weakly supervised learning with AD MIL was able to reach the accuracy obtained with supervised learning and in addition, enables the visualization of the regions that contributed to the decision by the attention mechanism. In [218], Lu *et al*. proposed a Clustering-constrained Attention Multiple instance learning (CLAM) that also only requires slide-level multi-class labels. CLAM uses attention-based learning to automatically identify sub-regions of high diagnostic value to accurately classify the whole slide, while also utilizing instance-level clustering over the representative regions identified to constrain and refine the feature space. In contrast to classical MIL, CLAM uses an attention-based pooling function to aggregate the patch-level features into slide-level representations for classification. This makes their approach much more efficient than classical MIL and diminishes the need for very large sets of slide-level labels. This last work showed that weakly supervised learning can be beneficial in computational pathology. This has been confirmed by the recent study [219] of Ghaffari Laleh *et al*. that compared weakly-supervised deep learning pipelines for whole slide classification in computational pathology. In particular, they showed that a classical weakly-supervised approach using Vision Transformers (that are very new in computational pathology) can outperform MIL and CLAM. As the field continuously evolves, there is also much attention now on self-supervised learning. Contrary to weakly-supervised learning that uses slide-level annotation, self-supervised learning aims at pre-training a self-supervised model on an unlabeled set to obtain task-agnostic feature representations. The model can then be fine-tuned on a limited amount of labeled data to obtain task-specific features [220]–[222]. If these models are appealing, they still need many computational resources to be trained.

VII. IMPORTANCE OF QUALITY CONTROL IN A ROUTINE LABORATORY PROCESS

To support our conclusion that quality control is essential in digital pathology in a routine laboratory process, we consider the monthly WSI output of a laboratory for cytological slides. Indeed, with the large number of slides that have to be digitized daily in a laboratory, quality issues are inevitable. However, it might be beneficial to detect these quality problems automatically. To support this assumption, we analyze the obtained monthly set of cytology WSIs both qualitatively and quantitatively. First, we provide details on the corpus of WSIs. They were obtained from a single laboratory equipped with a P250 3DHISTECH scanner generating `.mrxs` files. The whole corpus was digitized consecutively in a 30-day window between October 10, 2020, and November 29, 2020 (i.e., during 21 days) where slides were scanned. It represents 2093 WSIs, each scanned at different magnifications from 5x to 40x. At the highest resolution, the number of microns per pixel (*mpp*) for each slide is approximately 0.24, in other words, 1 micron of a glass slide need 4-5 pixels to be stored digitally. It represents approximately 2.85 terabytes of data. On this basis, we will analyze different aspects related to QC, each involving perturbation in the diagnosis. The aim is to be as close as possible to reality when analyzing the need for a laboratory when moving to the digital world.

A. SELECTED CRITERION

A digital slide is considered "unanalyzable" or "unreadable" (denoted as READ -) if there is doubt in the diagnosis according to all the pixel data available. For example, if some visible nuclei are noticeable but obfuscated by any factor (preparation issue, sample quality, out-of-focus area, etc.), doubt is allowed because these nuclei could potentially change the diagnosis for the whole slide. Several criteria have been defined for monitoring the different manifestations of these obfuscations. The labels used to annotate the corpus are described in Table 7 and are discredited with levels denoted by "-", "+", or "++" ordered by ascending order. These labels were defined according to the background of the annotator in cytology diagnosis with or without AI assistance. They also correspond to the main issues that the literature has already addressed according to our survey. Two main categories of elements are considered:

- Issues that an automatic method can handle from a quality diagnostic perspective (BLUR, BUBB)
- Elements that affect an automatic analysis that should be considered when collecting training corpus (ATRO, PAUC, BLOO, POLY, MUCU, BACT)

The intuition behind this selection is that several concomitant elements that slightly affect readability can accentuate the perturbation in the establishment of a reliable diagnosis.

label	description
ATRO	characteristic of the sample altering the rendering of nuclei and cytoplasm ("-" no atrophy, "+" atrophy sample)
PAUC	presence of visible nuclei ("-" low presence, "+" acceptable presence)
BUBB	a bubble appears in the digitization ("-" no bubble, "+" one or more bubbles)
BLUR	a zone of the slide is blurred ("-" no blur, "+" small zone, "++" large zone)
BLOO	presence of blood ("-" no trace of blood, "+" some traces, "++" a lot of blood)
POLY	presence of poly-nuclear cells ("-" few, "+" noticeable presence, "++" huge presence)
MUCU	presence of mucus ("-" few, "+" noticeable presence, "++" huge presence)
BACT	presence of bacteria ("-" few, "+" noticeable presence, "++" huge presence)

Table 7: Corpus label description.

All annotations were made by a single expert of the domain (a Cyto Technologist (IAC) involved in the diagnostic of both glass/digital slides and in collecting corpus cell examples to train machine learning based nuclei classification) with more than 10 years of experience.

B. DIACHRONIC AND QUANTITATIVE ANALYSIS

The 2093 slides were annotated according to the labels and their scale, at two or three different levels. WSIs are split into two main categories, "readable" (READ+) and "unreadable" (READ-) with their levels ("-" "+" or "++"). Each WSI was annotated with each category label and a level of readability. A synthesis of all annotations made is given in Table 8. Approximately 15% of WSI have been labeled as "hard to diagnose" (READ -). One can see that, even if considered as readable, the slides always contained preparation (air bubbles, blood, mucus and bacteria) and digitization (blurred areas) artifacts. Therefore, artifacts are not always a problem in terms of the readability of the slide. However, as expected, some labels determined the readability of the slide more than others. WSIs with a sparse distribution of cells (PAUC +/++) often lead to the READ - classification. The same effect appears in the presence of bubbles (BUBB +), blur (BLUR +/++) or poly (POLY +/++). For the other labels, the correlation between a given label and readability is less noticeable. But one can notice that for approximately 90% of the "hard to diagnose" slides, there was a joint presence of several preparation artifacts such as blood, mucus and bacteria. Similar studies have been led for histological slides [223], [224] and reported a rate of around 20% of quality issues (mainly due to out-of-focus and stitching).

LABEL	READ + (1787)			READ - (306)		
LEVEL	-	+	++	-	+	++
ATRO	1537	250		248	58	
PAUC	1787	0		254	52	
BUBB	1787	0		122	184	
BLUR	1787	0	0	73	229	4
BLOO	1597	179	11	253	41	12
POLY	1104	460	223	93	165	48
MUCU	1668	118	1	276	24	6
BACT	1579	200	8	275	31	0

Table 8: Corpus label distribution.

Figure 10 presents the stream of slides according to their labels. Some labels and days were gathered for readability purposes when the amount of associated data in each was too small. Regardless of the label, approximately the same

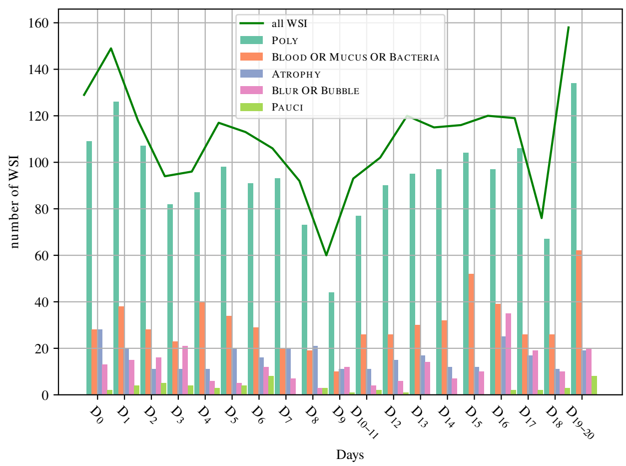

Figure 10: Characteristics of slides according to time.

proportion of WSI arrives at each time step. In addition, from the perspective of capturing examples to build a machine learning training dataset, some labels are scarcer than others. In this laboratory, it would take approximately a month to collect 100 paucicellular WSIs and a few days for atrophic WSIs.

To complete this journey through the labels, we analyze the data to check the interdependence between the labels. In other words, we want to ensure that the chosen labels are good descriptors with few redundancies in the overall information they provide. The uncertainty coefficient (entropy coefficient or Theil's U) is a non-symmetric measure based on the conditional entropy between two phenomena A and B. Knowing phenomenon A, it checks if we can predict phenomena B. We use this measure to check the predictability of labels according to others.

$$U(A|B) = \frac{H(A) - H(A|B)}{H(A)}$$

with H the conditional entropy as

$$H(A|B) = -\sum_{a,b} P_{A,B}(a,b)log(P_{A|B}(a|b))$$

$P_{A,B}$ the joint distribution between A and B

$P_{A|B}$ the conditional distribution of A knowing B

The uncertainty coefficient is non-symmetrical. From a WSI, if knowing the degree of B implies that we can estimate its A feature, we cannot infer systematically that knowing its degree of A can estimate how it is affected by feature B.

Figure 11: Correlation between labels in the corpus.

Figure 11 shows that the READ class is mainly due to the presence of the labels BLUR and BUBB. Similarly, the presence of BLUR and BUBB affects the readability status READ of the slide. Conversely, the PAUC label affects the READ class of the slide, but knowing the READ characteristic of the slide rarely implies a PAUC label. None of the other labels used were correlated according to the uncertainty coefficient. This acknowledges that the chosen labels form an orthogonal semantic space. Therefore, the annotations build a good foundation for describing digital slides from a diagnostic perspective.

As the last window opened on those data, we analyze the impact of concomitant labels on the WSI. If a single noise disturbs the interpretation of a WSI, the addition of a new perturbation can increase the difficulty of the diagnosis. Figure 12 shows the proportion of WSI READ+ / READ- according to the number of different labels characterizing the WSI positively. For instance, we counted 3 for a WSI labeled as BLOO+, ATRO+, and MUCU+. The number of labels and the readability characteristic of the WSI are correlated, the more a WSI is affected by disturbances, the less trustworthy is its diagnosis. This highlights that analyzing WSI through different angles is a necessary strategy:

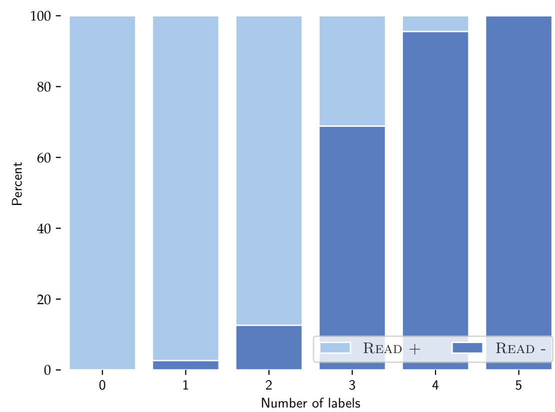

Figure 12: Evolution of the readability according to the number of labels attributed to WSIs.

taken individually, each label doesn't necessarily warn of a potential issue on the slide. A WSI can be categorized with regard to its analyzability only when analyzed holistically with orthogonal features. As well, by defining each case that can be encountered in a "divide and conquer" approach, the detected cause can lead to adequate countermeasures if anticipated. In the case of multiple causes compromising the readability of a WSI, the best set of actions to take is the one with the lowest cost for the laboratory, both in terms of time and material resources.

VIII. PERSPECTIVES

Out of this state of affairs, we can see that there are still many improvements to be expected for QC of WSI for routine clinical diagnosis. In particular, if some recent methods [56], [73] have opened the way toward global quality assessment of WSIs, there is still much to do. First, the methods have to be able to detect simultaneously different quality problems and to globally assess the slide quality in terms of analyzability for a final diagnosis (by a human or a computer-aided diagnosis system). Second, quality control should be able to detect out-of-the-scope items to avoid their analysis by an AI. In particular, too few works have been led on artifact detection in WSIs. Third, the methods have to be robust to the different quality situations that can occur in real practice. In particular, most of the methods for quality control are developed on specific small-size datasets. If they can work reasonably well on the learned datasets, their generalization abilities to data coming from other centers (using different slide preparation and digitization protocols) are not enough explored. Only the recent work of [225] has explored how deep neural networks perform on corrupted images (with compression, focus, color, and artifact issues) and shown that this can severely decrease the prediction accuracy. This makes none of the existing quality assessment tools ready for clinical use. However, it can be expected that novel techniques using larger datasets will converge to models close to being deployed for clinical practice.

We propose to draw an outline of what a QC tool for whole slide images should do and how it could be included within a computational pathology pipeline. The pipeline is as generic as possible as it encapsulates any histology and cytology routines. A similar pipeline was proposed in [226]. This is illustrated in Figure 13. In this

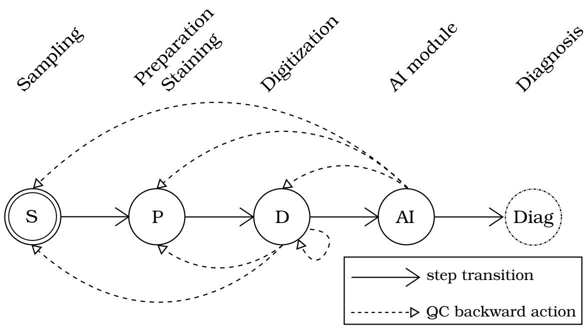

Figure 13: Holistic approach in QC – prototypical pipeline

prototypical pipeline, quality control checks are performed on two nodes: after digitization and conjointly with the AI module used in this DP pipeline. After the preparation of the biological sample, a scan is performed by a WSI scanner. The latter must be calibrated using a calibration slide to ensure color fidelity and reproducibility. The slide can be rejected if it needs to be re-scanned (digitization problem), re-stained (staining preparation problems), re-prepared (does not contain sufficient material), or accepted and analyzed for CAD. Just after digitization, several checks configured with the laboratory constraints and material (protocol for staining, the chemical used, machine for several preparation steps) are made:

D to D - re-digitize: Check if the minimum requirements are met in terms of magnification, microns per pixel, and size of the scanned area. Automatic glitch, blur, foreign body detection, and other feature detection with a relative pipeline agnostic characteristic can lead to scanner reconfiguration or glass slide cleaning before regenerating a new WSI.

D to P - re-prepare: Check gamut deviation according to laboratory standards (depending on the protocol, chemicals, and machine used). If the stain preparation was incorrect, the colors can be too far from what is expected. Such a gamut test can be performed using the estimated stain vectors [105].

D to S - re-sample: Some issues need another extraction from the initial sample e.g., when a large fold is detected in a histological WSI.

After these checks, the WSI can be used by a pathologist to perform its examination. With the addition of an AI module as a certified assistant for a pathologist, new constraints and new indicators can be performed, each eventually leading to backward action to any prior step of the pipeline.

AI to D - re-digitize: the scanner needs to meet the requirements for creating WSI in a compatible format and with enough information to allow the use of the AI module as intended.

AI to P - re-prepare: Sometimes, color deviation from the standard is not captured by quality control made in the prior steps. According to a fine analysis of the train data and data augmentation, the AI module should be self-aware of its range of acceptability. Countermeasures can be applied with color/style transfer from the domain of the processed WSI to the domain described by the training dataset. In the same way, an AI can be self-aware of its weaknesses by incorporating a module for the detection of known perturbation (artifact, blur, glitch, or pipeline-specific disturbance such as a large presence of blood). Any error detection module listed above must indicate the backward action to take to bypass the issue encountered.

AI to S - re-sample: In the same way, automatically detected issues can lead to the preparation of a brand-new sample (lack of biological material or un-recoverable preparation error).

Backward actions do not have to be automatic. When issues are located precisely on a WSI, joining data analysis with a map of trusted/corrupted areas allows the pathologist to obtain the final word. WSI containing blurred areas but visible suspicious regions should not be filtered out blindly. Finally, if a set of causes leads to difficulties in establishing a diagnosis (as depicted in our corpus analysis, Section VII) the best set of actions to take is the one minimizing the number of steps that have to be redone. Any other phenomenon that does not fit this ontology should be considered by enhancing the AI module with this new type of object or changing the laboratory workflow with a preemptive procedure.

IX. CONCLUSION

In this paper, we considered the quality of WSIs in digital pathology. Many factors in a digital pathology pipeline can have a strong influence on the quality, from slide preparation to slide digitization. In particular, we reviewed issues related to quality concerning the presence of sample preparation artifacts, compression artifacts, color variations, and out-of-focus areas. We have proposed a review of all the computational methods that have been proposed in the state-of-the-art for their detection.

In the quality process driven by diagnosis, we have established the notion of the analyzability of a WSI. The latter can be obtained from the previous computational methods that analyze the quality of slides. They can be blended to assess whether a slide is of sufficient quality to be used by a computer-aided diagnosis system, or should be re-prepared (too many physical artifacts), re-stained (too much color deviation), or re-scanned (too many out-of-focus areas).

As the most recent competing methods rely on deep learning, we also addressed issues related to the quality of the data used for learning deep models, and ways to cope with this problem.

Finally, to illustrate the importance of quality control in the daily practice of a real laboratory, we have labeled and analyzed the quality issues of cytological WSIs digitized during one month. This confirms that the presence of preparation artifacts (e.g., air bubbles) or digitization artifacts (e.g., out-of-focus areas) occurred for at least 15% of the slides. The greater the presence of artifacts, the less the slide is analyzable by a human expert and therefore might also be difficult to analyze using an AI-based system.

Based on this observation, we have drawn perspectives on how a computational quality process can be included in a computational diagnosis pipeline.

References

[1] Ronald S Weinstein, Kenneth J Bloom, and L Susan Rozek, "Static and dynamic imaging in pathology," in 1989 The First International Conference on Image Management and Communication in Patient Care: Implementation and Impact. 1989, pp. 77–78, IEEE Computer Society.

[2] Kendall Preston Jr, "High-resolution image analysis.," Journal of Histochemistry & Cytochemistry, vol. 34, no. 1, pp. 67–74, 1986, Publisher: SAGE Publications Sage CA: Los Angeles, CA.

[3] Nam Soojeong, Chong Yosep, Jung Chan Kwon, Kwak Tae-Yeong, Lee Ji Youl, Park Jihwan, Rho Mi Jung, and Go Heounjeong, "Introduction to digital pathology and computer-aided pathology," J Pathol Transl Med, vol. 54, no. 2, pp. 125–134, 2020.

[4] Andrew Janowczyk, Ren Zuo, Hannah Gilmore, Michael Feldman, and Anant Madabhushi, "HistoQC: An Open-Source Quality Control Tool for Digital Pathology Slides," JCO Clinical Cancer Informatics, , no. 3, pp. 1–7, 2019.

[5] National Pathology Accreditation Advisory Council, Requirements for Medical Pathology Services, vol. 1, Australian Department of Health, Australia, second edition, 2018.

[6] Geoffrey Rolls, "An introduction to the preparation of specimens for microscopy in the histopathology laboratory.," Tech. Rep., Leica Microsystems, 2021.

[7] Albert J Mach, Oladunni B Adeyiga, and Dino Di Carlo, "Microfluidic sample preparation for diagnostic cytopathology," Lab on a chip, vol. 13, no. 6, pp. 1011–1026, Mar. 2013.

[8] N. Farahani, A. Parwani, and L. Pantanowitz, "Whole slide imaging in pathology: advantages, limitations, and emerging perspectives.," Pathology and Laboratory Medicine International., vol. 7, pp. 23–33, 2015.

[9] Ankush. Patel, Ulysses. Balis, Jerome. Cheng, Zaibo. Li, Giovanni. Lujan, David. McClintock, Liron. Pantanowitz, and Anil. Parwani, "Contemporary whole slide imaging devices and their applications within the modern pathology department: A selected hardware review," Journal of Pathology Informatics, vol. 12, no. 1, pp. 50, 2021.

[10] Maki Ogura, Tomoharu Kiyuna, and Hiroshi Yoshida, "Impact of blurs on machine-learning aided digital pathology image analysis," Artificial Intelligence in Cancer, vol. 1, no. 1, pp. 31–38, 2020, Publisher: Baishideng Publishing Group Inc.

[11] Olsi Kusta, Charlotte Vestrup Rift, Torsten Risør, Eric Santoni-Rugiu, and John Brandt Brodersen, "Lost in digitization – a systematic review about the diagnostic test accuracy of digital pathology solutions," Journal of Pathology Informatics, vol. 13, pp. 100136, 2022.

[12] Liron Pantanowitz, John H. Sinard, Walter H. Henricks, Lisa A. Fatheree, Alexis B. Carter, Lydia Contis, Bruce A. Beckwith, Andrew J. Evans, Avtar Lal, and Anil V. Parwani, "Validating Whole Slide Imaging for Diagnostic Purposes in Pathology: Guideline from the College of American Pathologists Pathology and Laboratory Quality Center," Archives of Pathology & Laboratory Medicine, vol. 137, no. 12, pp. 1710–1722, Dec. 2013.

[13] Wangpan Shi, Petros Georgiou, Aqsa Akram, Matthew C Proute, Tatsiana Serhiyenia, Mina E Kerolos, Roshini Pradeep, Nageshwar R Kothur, and Safeera Khan, "Diagnostic Pitfalls of Digital Microscopy Versus Light Microscopy in Gastrointestinal Pathology: A Systematic Review.," Cureus, vol. 13, no. 8, pp. e17116, Aug. 2021.

[14] Edward Goacher, Rebecca Randell, Bethany Williams, and Darren Treanor, "The Diagnostic Concordance of Whole Slide Imaging and Light Microscopy: A Systematic Review," Archives of Pathology & Laboratory Medicine, vol. 141, no. 1, pp. 151–161, July 2016.

[15] Sathyanarayanan. Rajaganesan, Rajiv. Kumar, Vidya. Rao, Trupti. Pai, Neha. Mittal, Ayushi. Sahay, Santosh. Menon, and Sangeeta. Desai, "Comparative assessment of digital pathology systems for primary diagnosis," Journal of Pathology Informatics, vol. 12, no. 1, pp. 25, 2021.

[16] Thomas J. Fuchs and Joachim M. Buhmann, "Computational pathology: Challenges and promises for tissue analysis," Computerized Medical Imaging and Graphics, vol. 35, no. 7, pp. 515–530, 2011.

[17] Olivier Lézoray, Metin Gurcan, Ali Can, and Jean-Christophe Olivo-Marin, "Special issue on whole slide microscopic image processing," Computerized Medical Imaging and Graphics, vol. 35, no. 7, pp. 493–495, 2011.

[18] Filippo Fraggetta, "Clinical-grade Computational Pathology: Alea Iacta Est," Journal of pathology informatics, vol. 10, pp. 38–38, Dec. 2019.

[19] David N Louis, Michael Feldman, Alexis B Carter, Anand S Dighe, John D Pfeifer, Lynn Bry, Jonas S Almeida, Joel Saltz, Jonathan Braun, John E Tomaszewski, John R Gilbertson, John H Sinard, Georg K Gerber, Stephen J Galli, Jeffrey A Golden, and Michael J Becich, "Computational pathology: A path ahead," Archives of pathology & laboratory medicine, vol. 140, no. 1, pp. 41–50, Jan. 2016.

[20] Esther Abels, Liron Pantanowitz, Famke Aeffner, Mark D Zarella, Jeroen van der Laak, Marilyn M Bui, Venkata Np Vemuri, Anil V Parwani, Jeff Gibbs, Emmanuel Agosto-Arroyo, Andrew H Beck, and Cleopatra Kozlowski, "Computational pathology definitions, best practices, and recommendations for regulatory guidance: a white paper from the Digital Pathology Association," The Journal of pathology, vol. 249, no. 3, pp. 286–294, Nov. 2019.

[21] Miao Cui and David Y. Zhang, "Artificial intelligence and computational pathology," Laboratory Investigation, vol. 101, no. 4, pp. 412–422, 2021.

[22] Andrew. Janowczyk and Anant. Madabhushi, "Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases," Journal of Pathology Informatics, vol. 7, no. 1, pp. 29, 2016.

[23] Kaustav Bera, Kurt A. Schalper, David L. Rimm, Vamsidhar Velcheti, and Anant Madabhushi, "Artificial intelligence in digital pathology - new tools for diagnosis and precision oncology," Nature Reviews Clinical Oncology, vol. 16, no. 11, pp. 703–715, 2019.

[24] Richard J. Chen, Chengkuan Chen, Yicong Li, Tiffany Y. Chen, Andrew D. Trister, Rahul G. Krishnan, and Faisal Mahmood, "Scaling vision transformers to gigapixel images via hierarchical self-supervised learning," in IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022. 2022, pp. 16123–16134, IEEE.

[25] H. Zhang, Y. Meng, Y. Zhao, Y. Qiao, X. Yang, S. E. Coupland, and Y. Zheng, "Dtfd-mil: Double-tier feature distillation multiple instance learning for histopathology whole slide image classification," in 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022, pp. 18780–18790, IEEE Computer Society.

[26] Chen Li, Xintong Li, Md Rahaman, Xiaoyan Li, Hongzan Sun, Hong Zhang, Yong Zhang, Xiaoqi Li, Jian Wu, Yudong Yao, and others, "A comprehensive review of computer-aided whole-slide image analysis: from datasets to feature extraction, segmentation, classification, and detection approaches," arXiv preprint arXiv:2102.10553, 2021.

[27] Aïcha BenTaieb and Ghassan Hamarneh, "Deep Learning Models for Digital Pathology," CoRR, vol. abs/1910.12329, 2019, _eprinttype: arXiv.

[28] Massimo Salvi, U Rajendra Acharya, Filippo Molinari, and Kristen 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, p. 104129, 2020, Publisher: Elsevier.

[29] Vipul Baxi, Robin Edwards, Michael Montalto, and Saurabh Saha, "Digital pathology and artificial intelligence in translational medicine and clinical practice," Modern Pathology, vol. 35, no. 1, pp. 23–32, 2022.

[30] Chetan L. Srinidhi, Ozan Ciga, and Anne L. Martel, "Deep neural network models for computational histopathology: A survey," Medical Image Analysis, vol. 67, pp. 101813, 2021.

[31] Jeroen van der Laak, Geert Litjens, and Francesco Ciompi, "Deep learning in histopathology: the path to the clinic," Nature Medicine, vol. 27, no. 5, pp. 775–784, 2021.

[32] Liangrui Pan, Zhichao Feng, and Shaoliang Peng, "A review of machine learning approaches, challenges and prospects for computational tumor pathology," CoRR, vol. abs/2206.01728, 2022.

[33] João Pedro Mazuco Rodriguez, Rubens Rodriguez, Vitor Werneck Krauss Silva, Felipe Campos Kitamura, Gustavo Cesar Antônio Corradi, Ana Carolina Bertoletti de Marchi, and Rafael Rieder, "Artificial intelligence as a tool for diagnosis in digital pathology whole slide images: A systematic review," Journal of Pathology Informatics, vol. 13, pp. 100138, 2022.

[34] J. Zhu, M. Liu, and X. Li, "Progress on deep learning in digital pathology of breast cancer: a narrative review," Gland Surg., vol. 11, no. 4, pp. 751–766, 2022.

[35] André Victória Matias, João Gustavo Atkinson Amorim, Luiz Antonio Buschetto Macarini, Allan Cerentini, Alexandre Sherlley Casimiro Onofre, Fabiana Botelho De Miranda Onofre, Felipe Perozzo Daltoé, Marcelo Ricardo Stemmer, and Aldo von Wangenheim, "What is the state of the art of computer vision-assisted cytology? a systematic literature review," Computerized Medical Imaging and Graphics, p. 101934, 2021, Publisher: Elsevier.

[36] Md Mamunur Rahaman, Chen Li, Xiangchen Wu, Yudong Yao, Zhijie Hu, Tao Jiang, Xiaoyan Li, and Shouliang Qi, "A Survey for Cervical Cytopathology Image Analysis Using Deep Learning," IEEE Access, vol. 8, pp. 61687–61710, 2020.

[37] Keluo Yao, Evita Sadimin, Sue Chang, Daniel Schmolze, and Zaibo Li, "Current applications and challenges of digital pathology in cytopathology," Human Pathology Reports, vol. 28, pp. 300634, 2022.

[38] F. McKay F, B.J. Williams, G.Prestwich, D. Bansal, N. Hallowell, and D. Treanor, "The ethical challenges of artificial intelligence-driven digital pathology," J Pathol Clin Res, vol. 8, no. 3, pp. 209–216, 2022.

[39] K. Rakovic, R. Colling, L. Browning, M. Dolton, M.R. Horton, A. Protheroe, A.D. Lamb, R.J. Bryant, R. Scheffer, J. Crofts, E. Stanislaus, and C. Verrill, "The use of digital pathology and artificial intelligence in histopathological diagnostic assessment of prostate cancer: A survey of prostate cancer uk supporters," Diagnostics (Basel), vol. 12, no. 5, pp. 1225, 2022.

[40] Wei Ba, Shuhao Wang, Meixia Shang, Ziyan Zhang, Huan Wu, Chunkai Yu, Ranran Xing, Wenjuan Wang, Lang Wang, Cancheng Liu, Huaiyin Shi, and Zhigang Song, "Assessment of deep learning assistance for the pathological diagnosis of gastric cancer," Modern Pathology, vol. 35, no. 9, pp. 1262–1268, 2022.

[41] Muhammad Joan Ailia, Nishant Thakur, Jamshid Abdul-Ghafar, Chan Kwon Jung, Kwangil Yim, and Yosep Chong, "Current trend of artificial intelligence patents in digital pathology: A systematic evaluation of the patent landscape," Cancers, vol. 14, no. 10, 2022.

[42] Wouter Bulten, Maschenka Balkenhol, Jean-Joël Awoumou Belinga, Américo Brilhante, Aslı Çakır, Lars Egevad, Martin Eklund, Xavier Farré, Katerina Geronatsiou, Vincent Molinié, Guilherme Pereira, Paromita Roy, Günter Saile, Paulo Salles, Ewout Schaafsma, Joëlle Tschui, Anne-Marie Vos, Brett Delahunt, Hemamali Samaratunga, David J. Grignon, Andrew J. Evans, Daniel M. Berney, Chin-Chen Pan, Glen Kristiansen, James G. Kench, Jon Oxley, Katia R. M. Leite, Jesse K. McKenney, Peter A. Humphrey, Samson W. Fine, Toyonori Tsuzuki, Murali Varma, Ming Zhou, Eva Comperat, David G. Bostwick, Kenneth A. Iczkowski, Cristina Magi-Galluzzi, John R. Srigley, Hiroyuki Takahashi, Theo van der Kwast, Hester van Boven, Robert Vink, Jeroen van der Laak, Christina Hulsbergen-van der Kaa, Geert Litjens, and ISUP Pathology Imagebase Expert Panel, "Artificial intelligence assistance significantly improves Gleason grading of prostate biopsies by pathologists," Modern Pathology, vol. 34, no. 3, pp. 660–671, 2021.

[43] Felicia Marginean, Ida Arvidsson, Athanasios Simoulis, Niels Christian Overgaard, Kalle Åström, Anders Heyden, Anders Bjartell, and Agnieszka Krzyzanowska, "An Artificial Intelligence–based Support Tool for Automation and Standardisation of Gleason Grading in Prostate Biopsies," European Urology Focus, vol. 7, no. 5, pp. 995–1001, 2021.

[44] Ali Hasan Md Linkon, Md Mahir Labib, Tarik Hasan, Mozammal Hossain, and Marium-E. Jannat, "Deep learning in prostate cancer diagnosis and Gleason grading in histopathology images: An extensive study," Informatics in Medicine Unlocked, vol. 24, pp. 100582, 2021.

[45] Liron Pantanowitz, Douglas Hartman, Yan Qi, Eun Yoon Cho, Beomseok Suh, Kyunghyun Paeng, Rajiv Dhir, Pamela Michelow, Scott Hazelhurst, Sang Yong Song, and Soo Youn Cho, "Accuracy and efficiency of an artificial intelligence tool when counting breast mitoses," Diagnostic Pathology, vol. 15, no. 1, pp. 80, 2020.

[46] Soorya Shaji, M. Sreeraj, Jestin Joy, and Alphonsa Kuriakose, "On Computer-Aided Diagnosis of Breast Cancers Using Nuclear Atypia Scoring," in Advances in Communication Systems and Networks, J. Jayakumari, George K. Karagiannidis, Maode Ma, and Syed Akhter Hossain, Eds., Singapore, 2020, pp. 665–676, Springer Singapore.

[47] Andrea Duggento, Allegra Conti, Alessandro Mauriello, Maria Guerrisi, and Nicola Toschi, "Deep computational pathology in breast cancer," Seminars in Cancer Biology, vol. 72, pp. 226–237, 2021.

[48] Farzad Ghaznavi, Andrew Evans, Anant Madabhushi, and Michael Feldman, "Digital imaging in pathology: whole-slide imaging and beyond," Annual Review of Pathology: Mechanisms of Disease, vol. 8, no. 1, pp. 331–359, 2013.

[49] Stephan W Jahn, Markus Plass, and Farid Moinfar, "Digital pathology: Advantages, limitations and emerging perspectives," Journal of Clinical Medicine, vol. 9, no. 11, pp. 3697, 2020, Publisher: Multidisciplinary Digital Publishing Institute.

[50] Shyamali Mitra, Nibaran Das, Soumyajyoti Dey, Sukanta Chakraborty, Mita Nasipuri, and Mrinal Kanti Naskar, "Cytology Image Analysis Techniques Toward Automation: Systematically Revisited," ACM Comput. Surv., vol. 54, no. 3, Apr. 2021.

[51] Varun Rastogi, Naveen Puri, Swati Arora, Geetpriya Kaur, Lalita Yadav, and Rachna Sharma, "Artefacts: a diagnostic dilemma - a review.," J Clin Diagn Res, vol. 7, no. 10, pp. 2408–2413, Oct 2013.

[52] Syed. Taqi, Syed. Sami, Lateef. Sami, and Syed. Zaki, "A review of artifacts in histopathology," Journal of Oral and Maxillofacial Pathology, vol. 22, no. 2, pp. 279, 2018.

[53] Yukako Yagi and John R. Gilbertson, "A relationship between slide quality and image quality in whole slide imaging (WSI)," Diagnostic Pathology, vol. 3, no. 1, pp. S12, 2008.

[54] Khushboo. Sahay, Monica. Mehendiratta, Shweta. Rehani, Madhumani. Kumra, Rashi. Sharma, and Priyanka. Kardam, "Cytological artifacts masquerading interpretation," Journal of Cytology, vol. 30, no. 4, pp. 241–246, 2013.

[55] Roy. Lee, David. McClintock, Nora. Laver, and Yukako. Yagi, "Evaluation and optimization for liquid-based preparation cytology in whole slide imaging," Journal of Pathology Informatics, vol. 2, no. 1, pp. 46, 2011.

[56] Gijs Smit, Francesco Ciompi, Maria Cigéhn, Anna Bodén, Jeroen van der Laak, and Caner Mercan, "Quality control of whole-slide images through multi-class semantic segmentation of artifacts," in Medical Imaging with Deep Learning, 2021.

[57] Geoffrey O Rolls, Neville J Farmer, and John B Hall, Artifacts in histological and cytological preparations, Leica Microsystems., 2008.

[58] Nicholas. Wang, Jeremy. Kaplan, Joonsang. Lee, Jeffrey. Hodgin, Aaron. Udager, and Arvind. Rao, "Stress testing pathology models with generated artifacts," Journal of Pathology Informatics, vol. 12, no. 1, pp. 54, 2021.

[59] Neel Kanwal, Fernando Pérez-Bueno, Arne Schmidt, Kjersti Engan, and Rafael Molina, "The devil is in the details: Whole slide image acquisition and processing for artifacts detection, color variation, and data augmentation: A review," IEEE Access, vol. 10, pp. 58821–58844, 2022.

[60] Ali R. N. Avanaki, Kathryn S. Espig, Albert Xthona, Christian Lanciault, and Tom R. L. Kimpe, "Automatic Image Quality Assessment for Digital Pathology," in Breast Imaging - 13th International Workshop, IWDM 2016, Malmö, Sweden, June 19-22, 2016, Proceedings, Anders Tingberg, Kristina Lång, and Pontus Timberg, Eds. 2016, vol. 9699 of Lecture Notes in Computer Science, pp. 431–438, Springer.

[61] Sonal. Kothari, John. Phan, and May. Wang, "Eliminating tissue-fold artifacts in histopathological whole-slide images for improved image-based prediction of cancer grade," Journal of Pathology Informatics, vol. 4, no. 1, pp. 22, 2013.

[62] Pinky. Bautista and Yukako. Yagi, "Improving the visualization and detection of tissue folds in whole slide images through color enhancement," Journal of Pathology Informatics, vol. 1, no. 1, pp. 25, 2010.

[63] Sakari Palokangas, Jyrki Selinummi, and Olli Yli-Harja, "Segmentation of folds in tissue section images.," Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference, vol. 2007, pp. 5642–5645, 2007.

[64] Yijiang Chen, Jarcy Zee, Abigail Smith, Catherine Jayapandian, Jeffrey Hodgin, David Howell, Matthew Palmer, David Thomas, Clarissa Cassol, Alton B Farris III, Kathryn Perkinson, Anant Madabhushi, Laura Barisoni, and Andrew Janowczyk, "Assessment of a computerized quan-

titative quality control tool for whole slide images of kidney biopsies," The Journal of Pathology, vol. 253, no. 3, pp. 268–278, 2021.
[65] R. Rajesh Kumar, V. Ajith Kumar, P. N. Sharath Kumar, S. Sudhamony, and R. Ravindrakumar, "Detection and removal of artifacts in cervical Cytology images using Support Vector Machine," in 2011 IEEE International Symposium on IT in Medicine and Education, 2011, vol. 1, pp. 717–721.
[66] Hossain Md Shakhawat, Tomoya Nakamura, Fumikazu Kimura, Yukako Yagi, and Masahiro Yamaguchi, "Automatic Quality Evaluation of Whole Slide Images for the Practical Use of Whole Slide Imaging Scanner," ITE Transactions on Media Technology and Applications, vol. 8, no. 4, pp. 252–268, 2020, Publisher: The Institute of Image Information and Television Engineers.
[67] Md Shakhawat Hossain, Toyama Nakamura, Fumikazu Kimura, Yukako Yagi, and Masahiro Yamaguchi, "Practical image quality evaluation for whole slide imaging scanner," in Biomedical Imaging and Sensing Conference. 2018, vol. 10711, p. 107111S, International Society for Optics and Photonics.
[68] Adrien Foucart, Olivier Debeir, and Christine Decaestecker, "Artifact Identification in Digital Pathology from Weak and Noisy Supervision with Deep Residual Networks," in 2018 4th International Conference on Cloud Computing Technologies and Applications (Cloudtech), 2018, pp. 1–6.
[69] Adrien Foucart, Olivier Debeir, and Christine Decaestecker, "Snow Supervision in Digital Pathology: Managing Imperfect Annotations for Segmentation in Deep Learning," Research Square, 2021.
[70] Morteza Babaie and Hamid R Tizhoosh, "Deep features for tissue-fold detection in histopathology images," in European Congress on Digital Pathology. 2019, pp. 125–132, Springer.
[71] Sharib Ali, Nasullah Khalid Alham, Clare Verrill, and Jens Rittscher, "Ink Removal from Histopathology Whole Slide Images by Combining Classification, Detection and Image Generation Models," in 16th IEEE International Symposium on Biomedical Imaging, ISBI 2019, Venice, Italy, April 8-11, 2019. 2019, pp. 928–932, IEEE.
[72] Teng Zhang, Johanna Carvajal, Daniel F. Smith, Kun Zhao, Arnold Wiliem, Peter Hobson, Anthony Jennings, and Brian C. Lovell, "SlideNet: Fast and Accurate Slide Quality Assessment Based on Deep Neural Networks," in 2018 24th International Conference on Pattern Recognition (ICPR), 2018, pp. 2314–2319.
[73] Maryam Haghighat, Lisa Browning, Korsuk Sirinukunwattana, Stefano Malacrino, Nasullah Khalid Alham, Richard Colling, Ying Cui, Emad Rakha, Freddie C. Hamdy, Clare Verrill, and Jens Rittscher, "Pathprofiler: Automated quality assessment of retrospective histopathology whole-slide image cohorts by artificial intelligence - a case study for prostate cancer research," medRxiv, 2021.
[74] Maryam Haghighat, Lisa Browning, Korsuk Sirinukunwattana, Stefano Malacrino, Nasullah Khalid Alham, Richard Colling, Ying Cui, Emad Rakha, Freddie C. Hamdy, Clare Verrill, and Jens Rittscher, "Automated quality assessment of large digitised histology cohorts by artificial intelligence," Scientific Reports, vol. 12, no. 1, pp. 5002, 2022.
[75] Filippo Fraggetta, Yukako Yagi, Marcial Garcia-Rojo, Andrew J. Evans, J. Mark Tuthill, Alexi Baidoshvili, Douglas J. Hartman, Junya Fukuoka, and Liron Pantanowitz, "The importance of eslide macro images for primary diagnosis with whole slide imaging," Journal of Pathology Informatics, vol. 9, no. 1, pp. 46, 2018.
[76] Tomé Albuquerque, Ana Moreira, Beatriz Barros, Diana Montezuma, Sara P. Oliveira, Pedro C. Neto, João Monteiro, Liliana Ribeiro, Sofia Gonçalves, Ana Monteiro, Isabel M. Pinto, and Jaime S. Cardoso, "Quality control in digital pathology: Automatic fragment detection and counting," in 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2022, pp. 588–593.
[77] "Digital Imaging and Communications in Medicine (DICOM) Supplement 145: Whole Slide Microscopic Image IOD and SOP Classes," Tech. Rep., DICOM Standards Committee, Working Group 26, Pathology, 2010.
[78] Adam Goode, Benjamin Gilbert, Jan Harkes, Drazen Jukic, and Mahadev Satyanarayanan, "OpenSlide: A vendor-neutral software foundation for digital pathology," Journal of pathology informatics, vol. 4, pp. 27–27, Sept. 2013.
[79] Nita Mulliqi, Kimmo Kartasalo, Henrik Olsson, Xiaoyi Ji, Lars Egevad, Martin Eklund, and Pekka Ruusuvuori, "OpenPhi: an interface to access Philips iSyntax whole slide images for computational pathology," Bioinformatics, vol. 37, no. 21, pp. 3995–3997, Aug. 2021.
[80] Yinhai Wang, Kate E. Williamson, Paul J. Kelly, Jacqueline A. James, and Peter W. Hamilton, "SurfaceSlide: A Multitouch Digital Pathology Platform," PLOS ONE, vol. 7, no. 1, pp. 1–12, Jan. 2012.
[81] Anurag Sharma, Pinky Bautista, and Yukako Yagi, "Balancing Image Quality and Compression Factor for Special Stains Whole Slide Images," Analytical Cellular Pathology, vol. 35, pp. 960684, 2012, Publisher: IOS Press.
[82] Daniel Bug, Felix Bartsch, Nadine Sarah Schaadt, Mathias Wien, Friedrich Feuerhake, Julia Schüler, Eva Oswald, and Dorit Merhof, "Scalable HEVC for Histological Whole-Slide Image Compression," in Bildverarbeitung für die Medizin 2020 - Algorithmen - Systeme - Anwendungen. Proceedings des Workshops vom 15. bis 17. März 2020 in Berlin. 2020, Informatik Aktuell, pp. 315–321, Springer.
[83] Henrik. Helin, Teemu. Tolonen, Onni. Ylinen, Petteri. Tolonen, Juha. Napankangas, and Jorma. Isola, "Optimized JPEG 2000 compression for efficient storage of histopathological whole-Slide images," Journal of Pathology Informatics, vol. 9, no. 1, pp. 20, 2018.
[84] Thomas Kalinski, Ralf Zwönitzer, Florian Grabellus, Sien-Yi Sheu, Saadettin Sel, Harald Hofmann, and Albert Roessner, "Lossless Compression of JPEG2000 Whole Slide Images Is Not Required for Diagnostic Virtual Microscopy," American Journal of Clinical Pathology, vol. 136, no. 6, pp. 889–895, Jan. 2011.
[85] Farhad Ghazvinian Zanjani, Svitlana Zinger, Bastian Piepers, Saeed Mahmoudpour, Peter Schelkens, and Peter H. N. de With, "Impact of JPEG 2000 compression on deep convolutional neural networks for metastatic cancer detection in histopathological images," Journal of Medical Imaging, vol. 6, no. 2, pp. 1 – 9, 2019, Publisher: SPIE.
[86] Yijiang Chen, Andrew Janowczyk, and Anant Madabhushi, "Quantitative Assessment of the Effects of Compression on Deep Learning in Digital Pathology Image Analysis," JCO Clinical Cancer Informatics, , no. 4, pp. 221–233, 2020.
[87] Samuel Dodge and Lina Karam, "Understanding how image quality affects deep neural networks," in 2016 eighth international conference on quality of multimedia experience (QoMEX). 2016, pp. 1–6, IEEE.
[88] Emily L Clarke and Darren Treanor, "Colour in digital pathology: a review.," Histopathology, vol. 70, no. 2, pp. 153–163, Jan. 2017.
[89] "Guidelines for Hematoxylin & Eosin staining," Tech. Rep., National Society for Histotechnology, 2001.
[90] "A Quick Guide to Cytological Staining," Tech. Rep., Zeiss, 2018.
[91] "Technical Performance Assessment of Digital Pathology Whole Slide Imaging Devices: Guidance for Industry and Food and Drug Administration Staff," Tech. Rep., Food and Drug Administration, 2016.
[92] Yukako Yagi, "Color standardization and optimization in Whole Slide Imaging," Diagnostic Pathology, vol. 6, no. 1, pp. S15, 2011, ISBN: 1746-1596.
[93] Pinky Bautista, Noriaki Hashimoto, and Yukako Yagi, "Color standardization in whole slide imaging using a color calibration slide," Journal of Pathology Informatics, vol. 5, no. 1, pp. 4–4, Jan. 2014.
[94] Wei-Chung Cheng, Firdous Saleheen, and Aldo Badano, "Assessing color performance of whole-slide imaging scanners for digital pathology," Color Research & Application, vol. 44, no. 3, pp. 322–334, 2019.
[95] Prarthana Shrestha and Bas Hulsken, "Color accuracy and reproducibility in whole slide imaging scanners," Journal of Medical Imaging, vol. 1, no. 2, pp. 1 – 8, 2014.
[96] R. Jenoski, S. Mayer, R. Marks, and R. Salmon, "Color Calibration for Digital Cytology Scanner," Journal of Pathology Informatics, vol. 12, no. 1, pp. 19, 2021.
[97] Santanu Roy, Alok kumar Jain, Shyam Lal, and Jyoti Kini, "A study about color normalization methods for histopathology images," Micron, vol. 114, pp. 42–61, 2018.
[98] Thaina Aparecida Azevedo Tosta, Paulo Rogério de Faria, Leandro Alves Neves, and Marcelo Zanchetta do Nascimento, "Computational normalization of h&e-stained histological images: Progress, challenges and future potential," Artif. Intell. Medicine, vol. 95, pp. 118–132, 2019.
[99] Sonal Kothari, John H. Phan, Richard A. Moffitt, Todd H. Stokes, Shelby E. Hassberger, Qaiser Chaudry, Andrew N. Young, and May D. Wang, "Automatic batch-invariant color segmentation of histological cancer images," in Proceedings of the 8th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2011, March 30 - April 2, 2011, Chicago, Illinois, USA. 2011, pp. 657–660, IEEE.
[100] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley, "Color transfer between images," IEEE Computer Graphics and Applications, vol. 21, no. 5, pp. 34–41, 2001.

[101] D. Magee, D. Treanor, D. Crellin, M. Shires, K. Smith, K. Mohee, and P. Quirke, "Colour Normalisation in Digital Histopathology Images," in MICCAI Workshop - Optical Tissue Image Analysis in Microscopy, Histopathology and Endoscopy, 2009, pp. 100–111.
[102] 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, Aug. 2001.
[103] Marc Macenko, Marc Niethammer, J. S. Marron, David Borland, John T. Woosley, Xiaojun Guan, Charles Schmitt, and Nancy E. Thomas, "A method for normalizing histology slides for quantitative analysis," in 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2009, pp. 1107–1110.
[104] Marc Niethammer, David Borland, J. S. Marron, John Woosley, and Nancy E. Thomas, "Appearance Normalization of Histology Slides," in Machine Learning in Medical Imaging, Fei Wang, Pingkun Yan, Kenji Suzuki, and Dinggang Shen, Eds. 2010, pp. 58–66, Springer Berlin Heidelberg.
[105] Andreea Anghel, Milos Stanisavljevic, Sonali Andani, Nikolaos Papandreou, Jan Hendrick Rüschoff, Peter Wild, Maria Gabrani, and Haralampos Pozidis, "A High-Performance System for Robust Stain Normalization of Whole-Slide Images in Histopathology," Frontiers in Medicine, vol. 6, pp. 193, 2019.
[106] Yushan Zheng, Zhiguo Jiang, Haopeng Zhang, Fengying Xie, Dingyi Hu, Shujiao Sun, Jun Shi, and Chenghai Xue, "Stain standardization capsule for application-driven histopathological image normalization," IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 2, pp. 337–347, 2021.
[107] Anubha Gupta, Rahul Duggal, Shiv Gehlot, Ritu Gupta, Anvit Mangal, Lalit Kumar, Nisarg Thakkar, and Devprakash Satpathy, "GCTI-SN: Geometry-inspired chemical and tissue invariant stain normalization of microscopic medical images," Medical Image Analysis, vol. 65, pp. 101788, 2020.
[108] Adnan Mujahid Khan, Nasir Rajpoot, Darren Treanor, and Derek Magee, "A Nonlinear Mapping Approach to Stain Normalization in Digital Histopathology Images Using Image-Specific Color Deconvolution," IEEE Transactions on Biomedical Engineering, vol. 61, no. 6, pp. 1729–1738, 2014.
[109] Jakob Nikolas Kather, Cleo-Aron Weis, Alexander Marx, Alexander K. Schuster, Lothar R. Schad, and Frank Gerrit Zöllner, "New Colors for Histology: Optimized Bivariate Color Maps Increase Perceptual Contrast in Histological Images," PLOS ONE, vol. 10, no. 12, pp. 1–22, Dec. 2016.
[110] Xingyu Li and Konstantinos N. Plataniotis, "A Complete Color Normalization Approach to Histopathology Images Using Color Cues Computed From Saturation-Weighted Statistics," IEEE Trans. Biomed. Eng., vol. 62, no. 7, pp. 1862–1873, 2015.
[111] Abhishek Vahadane, Tingying Peng, Amit Sethi, Shadi Albarqouni, Lichao Wang, Maximilian Baust, Katja Steiger, Anna Melissa Schlitter, Irene Esposito, and Nassir Navab, "Structure-Preserving Color Normalization and Sparse Stain Separation for Histological Images," IEEE Transactions on Medical Imaging, vol. 35, no. 8, pp. 1962–1971, 2016.
[112] Gaoyi Lei, Yuanqing Xia, Di-Hua Zhai, Wei Zhang, Duanduan Chen, and Defeng Wang, "StainCNNs: An efficient stain feature learning method," Neurocomputing, vol. 406, pp. 267–273, 2020.
[113] Babak Ehteshami Bejnordi, Geert Litjens, Nadya Timofeeva, Irene Otte-Höller, André Homeyer, Nico Karssemeijer, and Jeroen AWM van der Laak, "Stain Specific Standardization of Whole-Slide Histopathological Images," IEEE Transactions on Medical Imaging, vol. 35, no. 2, pp. 404–415, 2016.
[114] Pegah Salehi and Abdolah Chalechale, "Pix2Pix-based Stain-to-Stain Translation: A Solution for Robust Stain Normalization in Histopathology Images Analysis," in 2020 International Conference on Machine Vision and Image Processing (MVIP), 2020, pp. 1–7.
[115] Hyungjoo Cho, Sungbin Lim, Gunho Choi, and Hyunseok Min, "Neural Stain-Style Transfer Learning using GAN for Histopathological Images," CoRR, vol. abs/1710.08543, 2017.
[116] M. Tarek Shaban, Christoph Baur, Nassir Navab, and Shadi Albarqouni, "Staingan: Stain Style Transfer for Digital Histological Images," in 16th IEEE International Symposium on Biomedical Imaging, ISBI 2019, Venice, Italy, April 8-11, 2019. 2019, pp. 953–956, IEEE.
[117] Hongtao Kang, Die Luo, Weihua Feng, Shaoqun Zeng, Tingwei Quan, Junbo Hu, and Xiuli Liu, "StainNet: A Fast and Robust Stain Normalization Network," Frontiers in Medicine, vol. 8, pp. 2002, 2021.
[118] Hanwen Liang, Konstantinos N. Plataniotis, and Xingyu Li, "Stain Style Transfer of Histopathology Images via Structure-Preserved Generative Learning," in Machine Learning for Medical Image Reconstruction, Farah Deeba, Patricia Johnson, Tobias Würfl, and Jong Chul Ye, Eds. 2020, pp. 153–162, Springer International Publishing.
[119] Aman Shrivastava, William Adorno, Yash Sharma, Lubaina Ehsan, S. Asad Ali, Sean R. Moore, Beatrice C. Amadi, Paul Kelly, Sana Syed, and Donald E. Brown, "Self-attentive Adversarial Stain Normalization," in Pattern Recognition. ICPR International Workshops and Challenges - Virtual Event, January 10-15, 2021, Proceedings, Part I, Alberto Del Bimbo, Rita Cucchiara, Stan Sclaroff, Giovanni Maria Farinella, Tao Mei, Marco Bertini, Hugo Jair Escalante, and Roberto Vezzani, Eds. 2020, vol. 12661 of Lecture Notes in Computer Science, pp. 120–140, Springer.
[120] Xihao Chen, Jingya Yu, Shenghua Cheng, Xiebo Geng, Sibo Liu, Wei Han, Junbo Hu, Li Chen, Xiuli Liu, and Shaoqun Zeng, "An unsupervised style normalization method for cytopathology images," Computational and Structural Biotechnology Journal, vol. 19, pp. 3852–3863, 2021.
[121] Jian Ren, Ilker Hacihaliloglu, Eric A. Singer, David J. Foran, and Xin Qi, "Unsupervised domain adaptation for classification of histopathology whole-slide images," Frontiers in Bioengineering and Biotechnology, vol. 7, pp. 102, 2019.
[122] Jiann-Shu Lee and Yao-Xian Ma, "Stain style transfer for histological images using s3cgan," Sensors, vol. 22, no. 3, 2022.
[123] Haseeb Nazki, Ognjen Arandjelovic, In Hwa Um, and David J. Harrison, "Multipathgan: Structure preserving stain normalization using unsupervised multi-domain adversarial network with perception loss," CoRR, vol. abs/2204.09782, 2022.
[124] Bingchao Zhao, Chu Han, Xipeng Pan, Jiatai Lin, Zongjian Yi, Changhong Liang, Xin Chen, Bingbing Li, Weihao Qiu, Danyi Li, Li Liang, Ying Wang, and Zaiyi Liu, "Restainnet: A self-supervised digital re-stainer for stain normalization," Computers and Electrical Engineering, vol. 103, pp. 108304, 2022.
[125] Sebastian Otalora, Niccolo Marini, Damian Podareanu, Ruben Hekster, David Tellez, Jeroen Van Der Laak, Henning Müller, and Manfredo Atzori, "stainlib: a python library for augmentation and normalization of histopathology h&e images," bioRxiv, 2022.
[126] Maximilian Ernst Tschuchnig, Gertie Janneke Oostingh, and Michael Gadermayr, "Generative Adversarial Networks in Digital Pathology: A Survey on Trends and Future Potential," Patterns, vol. 1, no. 8, pp. 100144, 2020.
[127] Marlen Runz, Daniel Rusche, Stefan Schmidt, Martin R. Weihrauch, Jürgen Hesser, and Cleo-Aron Weis, "Normalization of HE-stained histological images using cycle consistent generative adversarial networks," Diagnostic Pathology, vol. 16, no. 1, pp. 71, 2021.
[128] Patrick Leo, George Lee, Natalie N. C. Shih, Robin Elliott, Michael D. Feldman, and Anant Madabhushi, "Evaluating stability of histomorphometric features across scanner and staining variations: prostate cancer diagnosis from whole slide images," Journal of Medical Imaging, vol. 3, no. 4, pp. 1 – 11, 2016.
[129] Yiping Jiao, Junhong Li, and Shumin Fei, "Staining condition visualization in digital histopathological whole-slide images," Multimedia Tools and Applications, vol. 81, no. 13, pp. 17831–17847, 2022.
[130] Jocelyn Liu, Samuel Lam, Paul Lemaillet, and Wei-Chung Cheng, "Colorimetrical evaluation of color normalization methods for H& E-stained images," in Medical Imaging 2021: Digital Pathology, John E. Tomaszewski and Aaron D. Ward, Eds. 2021, vol. 11603, pp. 199 – 207, SPIE, Backup Publisher: International Society for Optics and Photonics.
[131] Dorsa Ziaei, Weizhe Li, Samuel Lam, Wei-Chung Cheng, and Weijie Chen, "Characterization of color normalization methods in digital pathology whole slide images," in Medical Imaging 2020: Digital Pathology, John E. Tomaszewski and Aaron D. Ward, Eds. 2020, vol. 11320, pp. 290 – 297, SPIE.
[132] Marc Aubreville, Christof Bertram, Mitko Veta, Robert Klopfleisch, Nikolas Stathonikos, Katharina Breininger, Natalie D. ter Hoeve, Francesco Ciompi, and Andreas Maier, "Quantifying the Scanner-Induced Domain Gap in Mitosis Detection," CoRR, vol. abs/2103.16515, 2021, arXiv: 2103.16515.
[133] Zeeshan Nisar, Jelica Vasiljević, Pierre Gançarski, and Thomas Lampert, "Towards measuring domain shift in histopathological stain translation in an unsupervised manner," in 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI), 2022, pp. 1–5.
[134] Justin Tyler Pontalba, Thomas Gwynne-Timothy, Ephraim David, Kiran Jakate, Dimitrios Androutsos, and April Khademi, "Assessing the Impact

of Color Normalization in Convolutional Neural Network-Based Nuclei Segmentation Frameworks," Frontiers in Bioengineering and Biotechnology, vol. 7, pp. 300, 2019.

[135] Zaneta Swiderska-Chadaj, Thomas de Bel, Lionel Blanchet, Alexi Baidoshvili, Dirk Vossen, Jeroen van der Laak, and Geert Litjens, "Impact of rescanning and normalization on convolutional neural network performance in multi-center, whole-slide classification of prostate cancer," Scientific Reports, vol. 10, no. 1, pp. 14398, 2020.

[136] Francesco Bianconi, Jakob N. Kather, and Constantino Carlos Reyes-Aldasoro, "Experimental Assessment of Color Deconvolution and Color Normalization for Automated Classification of Histology Images Stained with Hematoxylin and Eosin," Cancers, vol. 12, no. 11, 2020.

[137] Sonal Kothari, John H. Phan, Richard A. Moffitt, Todd H. Stokes, Shelby E. Hassberger, Qaiser Chaudry, Andrew N. Young, May D. Wang, Michael Gadermayr, Sean Steven Cooper, Barbara Mara Klinkhammer, Peter Boor, and Dorit Merhof, "A Quantitative Assessment of Image Normalization for Classifying Histopathological Tissue of the Kidney," in Pattern Recognition - 39th German Conference, GCPR 2017, Basel, Switzerland, September 12-15, 2017, Proceedings, Volker Roth and Thomas Vetter, Eds. 2017, vol. 10496 of Lecture Notes in Computer Science, pp. 3–13, Springer.

[138] David Tellez, Geert Litjens, Péter Bándi, Wouter Bulten, John-Melle Bokhorst, Francesco Ciompi, and Jeroen van der Laak, "Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology," Medical Image Analysis, vol. 58, pp. 101544, 2019.

[139] Zabit Hameed, Begonya Garcia-Zapirain, JoséJavier Aguirre, and Mario Arturo Isaza-Ruget, "Multiclass classification of breast cancer histopathology images using multilevel features of deep convolutional neural network," Scientific Reports, vol. 12, no. 1, pp. 15600, 2022.

[140] Francesco Ciompi, Oscar Geessink, Babak Ehteshami Bejnordi, Gabriel Silva de Souza, Alexi Baidoshvili, Geert Litjens, Bram van Ginneken, Iris Nagtegaal, and Jeroen van der Laak, "The importance of stain normalization in colorectal tissue classification with convolutional networks," CoRR, vol. abs/1702.05931, 2017.

[141] Jeffrey Boschman, Hossein Farahani, Amirali Darbandsari, Pouya Ahmadvand, Ashley Van Spankeren, David Farnell, Adrian B Levine, Julia R Naso, Andrew Churg, Steven JM Jones, Stephen Yip, Martin Köbel, David G Huntsman, C Blake Gilks, and Ali Bashashati, "The utility of color normalization for ai-based diagnosis of hematoxylin and eosin-stained pathology images," The Journal of Pathology, vol. 256, no. 1, pp. 15–24, 2022.

[142] Sebastian Otálora, Manfredo Atzori, Vincent Andrearczyk, Amjad Khan, and Henning Müller, "Staining Invariant Features for Improving Generalization of Deep Convolutional Neural Networks in Computational Pathology," Frontiers in Bioengineering and Biotechnology, vol. 7, pp. 198, 2019.

[143] Jelica Vasiljević, Friedrich Feuerhake, Cédric Wemmert, and Thomas Lampert, "Towards histopathological stain invariance by Unsupervised Domain Augmentation using generative adversarial networks," Neurocomputing, vol. 460, pp. 277–291, 2021.

[144] David Ameisen, Christophe Deroulers, Valérie Perrier, Jean-Baptiste Yunès, Fatiha Bouhidel, Maxime Battistella, Luc Legrès, Anne Janin, and Philippe Bertheau, "Stack or trash? quality assessment of virtual slides," Diagnostic Pathology, vol. 8, no. 1, pp. S23, 2013.

[145] David Ameisen, Christophe Deroulers, Valérie Perrier, Fatiha Bouhidel, Maxime Battistella, Luc Legrès, Anne Janin, Philippe Bertheau, and Jean-Baptiste Yunès, "Towards better digital pathology workflows: programming libraries for high-speed sharpness assessment of Whole Slide Images," Diagnostic Pathology, vol. 9, no. 1, pp. S3, 2014.

[146] Ana Jiménez, Gloria Bueno, Gabriel Cristóbal, Oscar Déniz, David Toomey, and Catherine Conway, "Image quality metrics applied to digital pathology," in Optics, Photonics and Digital Technologies for Imaging Applications IV, Peter Schelkens, Touradj Ebrahimi, Gabriel Cristóbal, Frédéric Truchetet, and Pasi Saarikko, Eds. 2016, vol. 9896, pp. 170 – 187, SPIE.

[147] Mahdi S. Hosseini, Yueyang Zhang, and Konstantinos N. Plataniotis, "Encoding visual sensitivity by maxpol convolution filters for image sharpness assessment," IEEE Transactions on Image Processing, vol. 28, no. 9, pp. 4510–4525, 2019.

[148] Mahdi S. Hosseini, Jasper A. Z. Brawley-Hayes, Yueyang Zhang, Lyndon Chan, Konstantinos N. Plataniotis, and Savvas Damaskinos, "Focus Quality Assessment of High-Throughput Whole Slide Imaging in Digital Pathology," IEEE Trans. Medical Imaging, vol. 39, no. 1, pp. 62–74, 2020.

[149] Noriaki Hashimoto, Pinky A Bautista, Masahiro Yamaguchi, Nagaaki Ohyama, and Yukako Yagi, "Referenceless image quality evaluation for whole slide imaging," Journal of pathology informatics, vol. 3, 2012, Publisher: Wolters Kluwer–Medknow Publications.

[150] Dashan Gao, Dirk Padfield, Jens Rittscher, and Richard McKay, "Automated training data generation for microscopy focus classification," in Medical Image Computing and Computer-Assisted Intervention (MICCAI). 2010, pp. 446–453, Springer Berlin Heidelberg.

[151] Bernd Lahrmann, Nektarios A. Valous, Urs Eisenmann, Nicolas Wentzensen, and Niels Grabe, "Semantic focusing allows fully automated single-layer slide scanning of cervical cytology slides," PLOS ONE, vol. 8, no. 4, pp. 1–10, 2013, Publisher: Public Library of Science.

[152] Xavier Moles Lopez, Etienne D'Andrea, Paul Barbot, Anne-Sophie Bridoux, Sandrine Rorive, Isabelle Salmon, Olivier Debeir, and Christine Decaestecker, "An Automated Blur Detection Method for Histological Whole Slide Imaging," PLOS ONE, vol. 8, no. 12, pp. 1–11, Dec. 2013, Publisher: Public Library of Science.

[153] Gabriele Campanella, Arjun R. Rajanna, Lorraine Corsale, Peter J. Schüffler, Yukako Yagi, and Thomas J. Fuchs, "Towards machine learned quality control: A benchmark for sharpness quantification in digital pathology," Computerized Medical Imaging and Graphics, vol. 65, pp. 142–151, 2018.

[154] T. Nair, A. Pour, and J. H. Chuang, "The effect of blurring on lung cancer subtype classification accuracy of convolutional neural networks," in 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2020, pp. 2987–2989.

[155] Caglar Senaras, M. Khalid Khan Niazi, Gerard Lozanski, and Metin N. Gurcan, "DeepFocus: Detection of out-of-focus regions in whole slide digital images using deep learning," PLOS ONE, vol. 13, no. 10, pp. 1–13, Oct. 2018, Publisher: Public Library of Science.

[156] Timo. Kohlberger, Yun. Liu, Melissa. Moran, Po-Hsuan. Chen, Trissia. Brown, Jason. Hipp, Craig. Mermel, and Martin. Stumpe, "Whole-slide image focus quality: Automatic assessment and impact on ai cancer detection," Journal of Pathology Informatics, vol. 10, no. 1, pp. 39, 2019.

[157] Zhongling Wang, Mahdi S. Hosseini, Adyn Miles, Konstantinos N. Plataniotis, and Zhou Wang, "FocusLiteNN: High Efficiency Focus Quality Assessment for Digital Pathology," in Medical Image Computing and Computer Assisted Intervention – MICCAI 2020, Anne L. Martel, Purang Abolmaesumi, Danail Stoyanov, Diana Mateus, Maria A. Zuluaga, S. Kevin Zhou, Daniel Racoceanu, and Leo Joskowicz, Eds. 2020, pp. 403–413, Springer International Publishing.

[158] Tomé Albuquerque, Ana Moreira, and Jaime S. Cardoso, "Deep Ordinal Focus Assessment for Whole Slide Images," in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops, Oct. 2021, pp. 657–663.

[159] Qiang Li, Xianming Liu, Kaige Han, Cheng Guo, Junjun Jiang, Xiangyang Ji, and Xiaolin Wu, "Learning to autofocus in whole slide imaging via physics-guided deep cascade networks," Opt. Express, vol. 30, no. 9, pp. 14319–14340, Apr 2022.

[160] Xiebo Geng, Xiuli Liu, Shenghua Cheng, and Shaoqun Zeng, "Cervical cytopathology image refocusing via multi-scale attention features and domain normalization," Medical Image Analysis, vol. 81, pp. 102566, 2022.

[161] Slawomir Walkowski and Janusz Szymas, "Quality evaluation of virtual slides using methods based on comparing common image areas," Diagnostic Pathology, vol. 6, no. 1, pp. S14, 2011.

[162] Norman Zerbe, Peter Hufnagl, and Karsten Schlüns, "Distributed computing in image analysis using open source frameworks and application to image sharpness assessment of histological whole slide images," Diagnostic pathology, vol. 6 Suppl 1, pp. S16–S16, 2011.

[163] Katarzyna Tomczak, Patrycja Czerwińska, and Maciej Wiznerowicz, "Reviewthe cancer genome atlas (tcga): an immeasurable source of knowledge," Contemporary Oncology/Współczesna Onkologia, pp. 68–77, 2015.

[164] Lawrence M. Firestone, Kelly Ann Cook, Kaitlyn Culp, Niharika Talsania, and K. Preston, "Comparison of autofocus methods for automated microscopy.," Cytometry, vol. 12, no. 3, pp. 195–206, 1991.

[165] Yu Sun, Stefan Duthaler, and Bradley J. Nelson, "Autofocusing in computer microscopy: Selecting the optimal focus algorithm," Microscopy Research and Technique, vol. 65, no. 3, pp. 139–149, 2004.

[166] Khosro Bahrami and Alex C. Kot, "A fast approach for no-reference image sharpness assessment based on maximum local variation," IEEE Signal Processing Letters, vol. 21, no. 6, pp. 751–755, 2014.

[167] Samuel J. Yang, Marc Berndl, D. Michael Ando, Mariya Barch, Arunachalam Narayanaswamy, Eric Christiansen, Stephan Hoyer, Chris Roat, Jane Hung, Curtis T. Rueden, Asim Shankar, Steven Finkbeiner, and Philip Nelson, "Assessing microscope image focus quality with deep learning," BMC Bioinformatics, vol. 19, no. 1, pp. 77, 2018.

[168] Gabriele Campanella, Matthew G. Hanna, Luke Geneslaw, Allen Miraflor, Vitor Werneck Krauss Silva, Klaus J. Busam, Edi Brogi, Victor E. Reuter, David S. Klimstra, and Thomas J. Fuchs, "Clinical-grade computational pathology using weakly supervised deep learning on whole slide images," Nature Medicine, vol. 25, no. 8, pp. 1301–1309, 2019.

[169] Zhigang Song, Shuangmei Zou, Weixun Zhou, Yong Huang, Liwei Shao, Jing Yuan, Xiangnan Gou, Wei Jin, Zhanbo Wang, Xin Chen, Xiaohui Ding, Jinhong Liu, Chunkai Yu, Calvin Ku, Cancheng Liu, Zhuo Sun, Gang Xu, Yuefeng Wang, Xiaoqing Zhang, Dandan Wang, Shuhao Wang, Wei Xu, Richard C. Davis, and Huaiyin Shi, "Clinically applicable histopathological diagnosis system for gastric cancer detection using deep learning," Nature Communications, vol. 11, no. 1, pp. 4294, 2020.

[170] Alexander I. Wright, Catriona M. Dunn, Michael Hale, Gordon G. A. Hutchins, and Darren E. Treanor, "The Effect of Quality Control on Accuracy of Digital Pathology Image Analysis," IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 2, pp. 307–314, 2021.

[171] Filippo Fraggetta, Vincenzo L'Imperio, David Ameisen, Rita Carvalho, Sabine Leh, Tim-Rasmus Kiehl, Mircea Serbanescu, Daniel Racoceanu, Vincenzo Della Mea, Antonio Polonia, Norman Zerbe, and Catarina Eloy, "Best practice recommendations for the implementation of a digital pathology workflow in the anatomic pathology laboratory by the european society of digital and integrative pathology (esdip)," Diagnostics, vol. 11, no. 11, 2021.

[172] D. Ameisen, J. Auger-Kantor, and E. Ameisen, "Towards a framework for continuous real-time image quality assurance," J Pathol Inform, vol. 11, no. 30, 2020.

[173] Prarthana Shrestha, Rik Kneepkens, Gijs van Elswijk, Jeroen Vrijnsen, Roxana Ion, Dirk Verhagen, Esther Abels, Dirk Vossen, and Bas Hulsken, "Objective and Subjective Assessment of Digital Pathology Image Quality," AIMS Medical Science, vol. 2, no. 1, pp. 65–78, 2015.

[174] Prarthana. Shrestha, R. Kneepkens, J. Vrijnsen, D. Vossen, E. Abels, and B. Hulsken, "A quantitative approach to evaluate image quality of whole slide imaging scanners," Journal of Pathology Informatics, vol. 7, no. 1, pp. 56, 2016.

[175] Andreas Holzinger, Randy Goebel, Michael Mengel, and Heimo Müller, Eds., Artificial Intelligence and Machine Learning for Digital Pathology - State-of-the-Art and Future Challenges, vol. 12090 of Lecture Notes in Computer Science, Springer, 2020.

[176] V. N. Vapnik and A. Ya. Chervonenkis, "On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities," Theory Probab. Appl., vol. 16, no. 2, pp. 264–280, 1971.

[177] Eric Baum and David Haussler, "What size net gives valid generalization?," in Advances in Neural Information Processing Systems, D. Touretzky, Ed. 1989, vol. 1, Morgan-Kaufmann.

[178] Peter L. Bartlett, Nick Harvey, Christopher Liaw, and Abbas Mehrabian, "Nearly-tight VC-dimension and Pseudodimension Bounds for Piecewise Linear Neural Networks," Journal of Machine Learning Research, vol. 20, no. 63, pp. 1–17, 2019.

[179] Philip M. Long and Hanie Sedghi, "Generalization bounds for deep convolutional neural networks," in 8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020. 2020, OpenReview.net.

[180] Syed Ashar Javed, Dinkar Juyal, Zahil Shanis, Shreya Chakraborty, Harsha Pokkalla, and Aaditya Prakash, "Rethinking machine learning model evaluation in pathology," CoRR, vol. abs/2204.05205, 2022.

[181] Noorul Wahab, Islam M Miligy, Katherine Dodd, Harvir Sahota, Michael Toss, Wenqi Lu, Mostafa Jahanifar, Mohsin Bilal, Simon Graham, Young Park, Giorgos Hadjigeorghiou, Abhir Bhalerao, Ayat G Lashen, Asmaa Y Ibrahim, Ayaka Katayama, Henry O Ebili, Matthew Parkin, Tom Sorell, Shan E Ahmed Raza, Emily Hero, Hesham Eldaly, Yee Wah Tsang, Kishore Gopalakrishnan, David Snead, Emad Rakha, Nasir Rajpoot, and Fayyaz Minhas, "Semantic annotation for computational pathology: multidisciplinary experience and best practice recommendations," The Journal of Pathology: Clinical Research, vol. 8, no. 2, pp. 116–128, 2022.

[182] Raphaël. Marée, "The need for careful data collection for pattern recognition in digital pathology," Journal of Pathology Informatics, vol. 8, no. 1, pp. 19, 2017.

[183] Mahdi S. Hosseini, Lyndon Chan, Gabriel Tse, Michael Tang, Jun Deng, Sajad Norouzi, Corwyn Rowsell, Konstantinos N. Plataniotis, and Savvas Damaskinos, "Atlas of digital pathology: A generalized hierarchical histological tissue type-annotated database for deep learning," in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 11739–11748.

[184] Rodrigo Escobar Díaz Guerrero, Lina Carvalho, Thomas Bocklitz, Juergen Popp, and José Luis Oliveira, "Software tools and platforms in digital pathology: a review for clinicians and computer scientists," Journal of Pathology Informatics, vol. 13, pp. 100103, 2022.

[185] Fabrice de Chaumont, Stéphane Dallongeville, Nicolas Chenouard, Nicolas Hervé, Sorin Pop, Thomas Provoost, Vannary Meas-Yedid, Praveen Pankajakshan, Timothée Lecomte, Yoann Le Montagner, Thibault Lagache, Alexandre Dufour, and Jean-Christophe Olivo-Marin, "Icy: an open bioimage informatics platform for extended reproducible research," Nature Methods, vol. 9, no. 7, pp. 690–696, 2012.

[186] Peter Bankhead, Maurice B. Loughrey, JoséA. Fernández, Yvonne Dombrowski, Darragh G. McArt, Philip D. Dunne, Stephen McQuaid, Ronan T. Gray, Liam J. Murray, Helen G. Coleman, Jacqueline A. James, Manuel Salto-Tellez, and Peter W. Hamilton, "Qupath: Open source software for digital pathology image analysis," Scientific Reports, vol. 7, no. 1, pp. 16878, 2017.

[187] Raphaël Marée, Benjamin Stévens, Loïc Rollus, Natacha Rocks, Xavier Moles Lopez, Isabelle Salmon, Didier Cataldo, and Louis Wehenkel, "A rich internet application for remote visualization and collaborative annotation of digital slides in histology and cytology," Diagnostic Pathology, vol. 8, no. Suppl 1, pp. S26–S26, 09 2013.

[188] Marc Aubreville, Christof Bertram, Robert Klopfleisch, and Andreas K. Maier, "Sliderunner - A tool for massive cell annotations in whole slide images," in Bildverarbeitung für die Medizin 2018 - Algorithmen - Systeme - Anwendungen. Proceedings des Workshops vom 11. bis 13. März 2018 in Erlangen, Andreas K. Maier, Thomas M. Deserno, Heinz Handels, Klaus Hermann Maier-Hein, Christoph Palm, and Thomas Tolxdorff, Eds. 2018, Informatik Aktuell, pp. 309–314, Springer Vieweg.

[189] Runtian Miao, Robert Toth, Yu Zhou, Anant Madabhushi, and Andrew Janowczyk, "Quick annotator: an open-source digital pathology based rapid image annotation tool," The Journal of Pathology: Clinical Research, vol. 7, no. 6, pp. 542–547, 2021.

[190] Jerry W. Wei, Arief A. Suriawinata, Bing Ren, Xiaoying Liu, Mikhail Lisovsky, Louis J. Vaickus, Charles Brown, Michael Baker, Mustafa Nasir-Moin, Naofumi Tomita, Lorenzo Torresani, Jason Wei, and Saeed Hassanpour, "Learn like a pathologist: Curriculum learning by annotator agreement for histopathology image classification," in IEEE Winter Conference on Applications of Computer Vision, WACV 2021, Waikoloa, HI, USA, January 3-8, 2021. 2021, pp. 2472–2482, IEEE.

[191] Christof A. Bertram, Mitko Veta, Christian Marzahl, Nikolas Stathonikos, Andreas K. Maier, Robert Klopfleisch, and Marc Aubreville, "Are pathologist-defined labels reproducible? comparison of the TUPAC16 mitotic figure dataset with an alternative set of labels," in Interpretable and Annotation-Efficient Learning for Medical Image Computing - Third International Workshop, iMIMIC 2020, Second International Workshop, MIL3ID 2020, and 5th International Workshop, LABELS 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4-8, 2020, Proceedings. 2020, vol. 12446 of Lecture Notes in Computer Science, pp. 204–213, Springer.

[192] Christian Marzahl, Marc Aubreville, Christof A. Bertram, Jennifer Maier, Christian Bergler, Christine Kröger, Jörn Voigt, Katharina Breininger, Robert Klopfleisch, and Andreas Maier, "Exact: a collaboration toolset for algorithm-aided annotation of images with annotation version control," Scientific Reports, vol. 11, no. 1, pp. 4343, 2021.

[193] Zeyad Emam, Andrew Kondrich, Sasha Harrison, Felix Lau, Yushi Wang, Aerin Kim, and Elliot Branson, "On the state of data in computer vision: Human annotations remain indispensable for developing deep learning models," CoRR, vol. abs/2108.00114, 2021.

[194] Yves-Rémi Van Eycke, Adrien Foucart, and Christine Decaestecker, "Strategies to reduce the expert supervision required for deep learning-based segmentation of histopathological images," Frontiers in Medicine, vol. 6, pp. 222, 2019.

[195] Connor Shorten and Taghi M. Khoshgoftaar, "A survey on image data augmentation for deep learning," Journal of Big Data, vol. 6, no. 1, pp. 60, 2019.

[196] Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He, "A comprehensive survey on transfer learning," Proceedings of the IEEE, vol. 109, no. 1, pp. 43–76, 2021.

[197] Mei Wang and Weihong Deng, "Deep visual domain adaptation: A survey," Neurocomputing, vol. 312, pp. 135–153, 2018.

[198] Dingwen Zhang, Junwei Han, Gong Cheng, and Ming-Hsuan Yang, "Weakly supervised object localization and detection: A survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1–1, 2021.

[199] Atsushi Teramoto, Ayumi Yamada, Yuka Kiriyama, Tetsuya Tsukamoto, Ke Yan, Ling Zhang, Kazuyoshi Imaizumi, Kuniaki Saito, and Hiroshi Fujita, "Automated classification of benign and malignant cells from lung cytological images using deep convolutional neural network," Informatics in Medicine Unlocked, vol. 16, pp. 100205, 2019.

[200] Jonas Annuscheit, Benjamin Voigt, Oliver Fischer, Patrick Baumann, Sebastian Lohmann, Christian Krumnow, and Christian Herta, "Systematic investigation of basic data augmentation strategies on histopathology images," in Artificial Intellgence – Application in Life Sciences and Beyond. The Upper Rhine Artificial Intelligence Symposium UR-AI 2021, 2021, pp. 39–48.

[201] Khrystyna Faryna, Jeroen van der Laak, and Geert Litjens, "Tailoring automated data augmentation to H& E-stained histopathology," in Proceedings of the Fourth Conference on Medical Imaging with Deep Learning, Mattias Heinrich, Qi Dou, Marleen de Bruijne, Jan Lellmann, Alexander Schläfer, and Floris Ernst, Eds. 07–09 Jul 2021, vol. 143 of Proceedings of Machine Learning Research, pp. 168–178, PMLR.

[202] Shallu Sharma and Sumit Kumar, "The xception model: A potential feature extractor in breast cancer histology images classification," ICT Express, vol. 8, no. 1, pp. 101–108, 2022.

[203] Xingyu Li and Konstantinos N. Plataniotis, "How Much Off-The-Shelf Knowledge Is Transferable From Natural Images To Pathology Images?," CoRR, vol. abs/2005.01609, 2020, arXiv: 2005.01609.

[204] Yash Sharma, Lubaina Ehsan, Sana Syed, and Donald E. Brown, "Histotransfer: Understanding transfer learning for histopathology," in IEEE EMBS International Conference on Biomedical and Health Informatics, BHI 2021, Athens, Greece, July 27-30, 2021. 2021, pp. 1–4, IEEE.

[205] Tayyab Aitazaz, Abdullah Tubaishat, Feras Al-Obeidat, Babar Shah, Tehseen Zia, and Ali Tariq, "Transfer learning for histopathology images: an empirical study," Neural Computing and Applications, 2022.

[206] Romain Mormont, Pierre Geurts, and Raphaël Marée, "Comparison of Deep Transfer Learning Strategies for Digital Pathology," in 2018 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2018, Salt Lake City, UT, USA, June 18-22, 2018. 2018, pp. 2262–2271, Computer Vision Foundation / IEEE Computer Society.

[207] Hyun-Jong Jang, In Hye Song, and Sung Hak Lee, "Generalizability of Deep Learning System for the Pathologic Diagnosis of Various Cancers," Applied Sciences, vol. 11, no. 2, 2021.

[208] Romain Mormont, Pierre Geurts, and Raphaël Marée, "Multi-Task Pre-Training of Deep Neural Networks for Digital Pathology," IEEE J. Biomed. Health Informatics, vol. 25, no. 2, pp. 412–421, 2021.

[209] Ryan Zhang, Jiadai Zhu, Stephen Yang, Mahdi S. Hosseini, Angelo Genovese, Lina Chen, Corwyn Rowsell, Savvas Damaskinos, Sonal Varma, and Konstantinos N. Plataniotis, "Histokt: Cross knowledge transfer in computational pathology," in ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2022, pp. 1276–1280.

[210] Aïcha BenTaieb and Ghassan Hamarneh, "Adversarial Stain Transfer for Histopathology Image Analysis," IEEE Trans. Medical Imaging, vol. 37, no. 3, pp. 792–802, 2018.

[211] Jian Ren, Ilker Hacihaliloglu, Eric A. Singer, David J. Foran, and Xin Qi, "Adversarial Domain Adaptation for Classification of Prostate Histopathology Whole-Slide Images," in Medical Image Computing and Computer Assisted Intervention - MICCAI 2018 - 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II, Alejandro F. Frangi, Julia A. Schnabel, Christos Davatzikos, Carlos Alberola-López, and Gabor Fichtinger, Eds. 2018, vol. 11071 of Lecture Notes in Computer Science, pp. 201–209, Springer.

[212] Seo Jeong Shin, Seng Chan You, Hokyun Jeon, Ji Won Jung, Min Ho An, Rae Woong Park, and Jin Roh, "Style transfer strategy for developing a generalizable deep learning application in digital pathology," Comput. Methods Programs Biomed., vol. 198, pp. 105815, 2021.

[213] Mahdi S. Hosseini, Lyndon Chan, Weimin Huang, Yichen Wang, Danial Hasan, Corwyn Rowsell, Savvas Damaskinos, and Konstantinos N. Plataniotis, "On Transferability of Histological Tissue Labels in Computational Pathology," in Computer Vision - ECCV 2020 - 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part XXIX, Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, Eds. 2020, vol. 12374 of Lecture Notes in Computer Science, pp. 453–469, Springer.

[214] Jérôme Rony, Soufiane Belharbi, Jose Dolz, Ismail Ben Ayed, Luke McCaffrey, and Eric Granger, "Deep weakly-supervised learning methods for classification and localization in histology images: a survey," CoRR, vol. abs/1909.03354, 2019.

[215] Lyndon Chan, Mahdi S. Hosseini, and Konstantinos N. Plataniotis, "A comprehensive analysis of weakly-supervised semantic segmentation in different image domains," International Journal of Computer Vision, vol. 129, no. 2, pp. 361–384, 2021.

[216] Fahdi Kanavati, Gouji Toyokawa, Seiya Momosaki, Michael Rambeau, Yuka Kozuma, Fumihiro Shoji, Koji Yamazaki, Sadanori Takeo, Osamu Iizuka, and Masayuki Tsuneki, "Weakly-supervised learning for lung carcinoma classification using deep learning," Scientific Reports, vol. 10, no. 1, pp. 9297, 2020.

[217] Atsushi Teramoto, Yuka Kiriyama, Tetsuya Tsukamoto, Eiko Sakurai, Ayano Michiba, Kazuyoshi Imaizumi, Kuniaki Saito, and Hiroshi Fujita, "Weakly supervised learning for classification of lung cytological images using attention-based multiple instance learning," Scientific Reports, vol. 11, no. 1, pp. 20317, 2021.

[218] Ming Y. Lu, Drew F. K. Williamson, Tiffany Y. Chen, Richard J. Chen, Matteo Barbieri, and Faisal Mahmood, "Data-efficient and weakly supervised computational pathology on whole-slide images," Nature Biomedical Engineering, vol. 5, no. 6, pp. 555–570, 2021.

[219] Narmin Ghaffari Laleh, Hannah Sophie Muti, Chiara Maria Lavinia Loeffler, Amelie Echle, Oliver Lester Saldanha, Faisal Mahmood, Ming Y. Lu, Christian Trautwein, Rupert Langer, Bastian Dislich, Roman D. Buelow, Heike Irmgard Grabsch, Hermann Brenner, Jenny Chang-Claude, Elizabeth Alwers, Titus J. Brinker, Firas Khader, Daniel Truhn, Nadine T. Gaisa, Peter Boor, Michael Hoffmeister, Volkmar Schulz, and Jakob Nikolas Kather, "Benchmarking weakly-supervised deep learning pipelines for whole slide classification in computational pathology," Medical Image Analysis, vol. 79, pp. 102474, 2022.

[220] Ozan Ciga, Tony Xu, and Anne Louise Martel, "Self supervised contrastive learning for digital histopathology," Machine Learning with Applications, vol. 7, pp. 100198, 2022.

[221] Yoni Schirris, Efstratios Gavves, Iris Nederlof, Hugo Mark Horlings, and Jonas Teuwen, "Deepsmile: Contrastive self-supervised pre-training benefits msi and hrd classification directly from h&e whole-slide images in colorectal and breast cancer," Medical Image Analysis, vol. 79, pp. 102464, 2022.

[222] Chetan L. Srinidhi, Seung Wook Kim, Fu-Der Chen, and Anne L. Martel, "Self-supervised driven consistency training for annotation efficient histopathology image analysis," Medical Image Analysis, vol. 75, pp. 102256, 2022.

[223] Diana Montezuma, Ana Monteiro, João Fraga, Liliana Ribeiro, Sofia Gonçalves, André Tavares, João Monteiro, and Isabel Macedo-Pinto, "Digital pathology implementation in private practice: Specific challenges and opportunities," Diagnostics, vol. 12, no. 2, 2022.

[224] Nehal M. Atallah, Michael S. Toss, Clare Verrill, Manuel Salto-Tellez, David Snead, and Emad A. Rakha, "Potential quality pitfalls of digitalized whole slide image of breast pathology in routine practice," Modern Pathology, vol. 35, no. 7, pp. 903–910, 2022.

[225] Yunlong Zhang, Yuxuan Sun, Honglin Li, Sunyi Zheng, Chenglu Zhu, and Lin Yang, "Benchmarking the robustness of deep neural networks to common corruptions in digital pathology," in Medical Image Computing and Computer Assisted Intervention – MICCAI 2022, Linwei Wang, Qi Dou, P. Thomas Fletcher, Stefanie Speidel, and Shuo Li, Eds., Cham, 2022, pp. 242–252, Springer Nature Switzerland.

[226] Byron Smith, Meyke Hermsen, Elizabeth Lesser, Deepak Ravichandar, and Walter Kremers, "Developing image analysis pipelines of whole-slide images: Pre- and post-processing," Journal of Clinical and Translational Science, vol. 5, no. 1, pp. e38, 2021.

ROMAIN BRIXTEL was graduated with a Ph.D. in 2011 at the university of Caen (France) in Computer Science, Artificial Intelligence and Natural Language Processing.

He pursued his work with some post-doctoral studies at the university of Caen and Lausanne from 2012 to 2016 in various fields driven by Machine Learning perspectives as multimodal and multiscale document studies, plagiarism detection and charisma analysis . He then have been introduced into Datexim, a company dedicated to medical image analysis where he became CSO in 2019 and build projects with Digital Pathology breakthrough in mind.

SÉBASTIEN BOUGLEUX received the M.Sc. degree (2002) and the Ph.D. degree (2007) in Computer Science from the University of Caen Normandie, France.

He was an Assistant Professor (2005-2007) with the University of Caen Normandie, ENSICaen and ENSTC Caen, and a Postdoctoral Researcher (2008) with CEREMADE at the University of Paris Dauphine, France. In 2008, he became an Associate Professor of Computer Science at the University of Caen Normandie and a member of the GREYC CNRS laboratory. He was director of graduate studies (2009-2013 and 2021-2025) and the training director of the D2PC Bachelor degree (2013-2017) at the Multimedia and Internet Department of the West Normandy Institute of Technology. His research interests lie in the area of statistical and structural pattern recognition, computer vision and image processing, with a focus on graph-based models.

OLIVIER LÉZORAY received the M.Sc., Ph.D and Habilitation degrees in Computer Science from the University of Caen Normandy, France, in 1996, 2000, and 2007, respectively.

From 1999 to 2000, he was an Assistant Professor with the Computer Science Department of the University of Caen. In 2000, he joined the West Normandy Institute of Technology as an Associate Professor of Computer Science in the Communication Networks and Services Department and was director of graduate studies (2000-2009). In 2010, he became a Full Professor at the West Normandy Institute of Technology in the Multimedia and Internet (MMI) Department. He was the head of the MMI Department from 2015 to 2021. Since 2022 he is deputy director of the GREYC CNRS research lab. He is the author of two books, 11 book chapters, 50 papers in international journals, and more than 120 in international conferences. His research interests are in graph-based signal processing, adaptive and multidimensional mathematical morphology, and machine learning. Applications of his works include computational photography, computer-aided diagnosis, and computer vision. He was an Associate Editor for the IEEE Transactions on Signal Processing and is for the IET Image Processing. He guest-edited for the following journals: Signal Processing: Image Communication, Journal of Electronic Imaging, Signal Image and Video Processing, Computerized Medical Imaging and Graphics, Signal Processing, and the EURASIP Journal on Advances in Signal Processing. He is a Senior Member of the IEEE and a member of SPIE, IAPR and EURASIP.

YANN CAILLOT received the Cytotechnology Diploma from the Dijon's Cytology school, B.Sc. in Biotechnology from the University of Bourgogne (France) and the International Academy of Cytology Certification in 2005, 2008 and 2011 respectively.

From 2005 to 2016, he worked as Cytotechnologist in a private laboratory. In 2016 he integrated the Datexim company where he is still working as Product Manager and Clinical Expert. Passionate with both gynecological and non-gynecological cytology, as well as histology, he also fulfilled his interest in new technologies by collaborating in the development of the first CE-marked software for cervical cancer screening.

BENOÎT LEMOINE received a M.Sc degree Degree in Machine Learning and Image Processing at Ensicaen, Caen Normandy (France), in 2020.

He then joined Datexim R&D team as junior developer, working to extend the first CE-marked software for cervical cancer screening. Machine learning enthusiast, he stays in touch with state of the art to propose innovative tools for digital pathology.

MATHIEU FONTAINE completed his Ph.D. thesis in 2013 at the university of Caen (France), in Computer Science, on Constraint Satisfaction Problem.

After spending 2 years working on facial recognition, he joined Datexim in 2014 as a R&D engineer to work on image similarities detection then on Artificial Intelligence in Digital Pathology, more specifically on cell nucleus detection and classification for cervical cancer screening.

DALAL NEBATI was graduated with a M.Sc degree from Oran University (Algeria) in microbiology in 2015.

In 2016, she successfully participated in the formation of cervico-vaginal cytopathology. From 2011 to 2017, she worked as an R&D engineer and participated in the conception of clinical trials and in the development of the first CE-marked software for bladder cancer. In 2022, she joined the Datexim company as a cytology engineer involved in various research projects on cervical cancer. Dalal is passionate about Digital Pathology and Artificial Intelligence and has spent years infusing her expert knowledge in this field of oncology and pathology into innovative digital applications.

ARNAUD RENOUF was graduated with a PhD in 2011 at University of Caen (France) in Computer Sciences with a specialization in Image processing and Artificial Intelligence.

He founded Datexim in 2011 where he started as the COO and CSO of the company. In 2018 he became the President of Datexim, building the international image of the company, leading the collaboration with other international industrial and academic partners, and bringing his expertise to the customers worldwide through his business development activities. From a science background, he quickly improved his skills in business development and company management thanks to the HEC Paris training program (Challenge+) and another training at EM Lyon.