The devil is in the details: Whole Slide Image acquisition and processing for artifacts detection, color variation, and data augmentation. A review.

NEEL KANWAL[1,*], **FERNANDO PÉREZ-BUENO**[2,*], **ARNE SCHMIDT**[2], **RAFAEL MOLINA**[2], (SENIOR MEMBER, IEEE), and **KJERSTI ENGAN**[1], (SENIOR MEMBER, IEEE)

[1] Department of Electrical Engineering and Computer Science, University of Stavanger, 4021 Stavanger, Norway
[2] Department of Computer Science and Artificial Intelligence, University of Granada, 18071 Granada, Spain
[*] These authors have contributed equally.

Corresponding author: Neel Kanwal (e-mail: neel.kanwal@uis.no).

ABSTRACT Whole Slide Images (WSI) are widely used in histopathology for research and the diagnosis of different types of cancer. The preparation and digitization of histological tissues leads to the introduction of artifacts and variations that need to be addressed before the tissues are analyzed. WSI preprocessing can significantly improve the performance of computational pathology systems and is often used to facilitate human or machine analysis. Color processing techniques are usually the main concern, while other areas are frequently ignored. In this paper, we present a detailed study of the state-of-the-art in three different areas of WSI preprocessing: Artifacts detection, color variation, and the emerging field of pathology-specific data augmentation. We include a summary of evaluation techniques along with a discussion of possible limitations and future research directions for new methods.

INDEX TERMS Artifacts Detection, Computational Pathology, Histopathological Images, Image Augmentation, Preprocessing, Stain Normalization

I. INTRODUCTION

In [1], the author describes five examples, from Tesla's fatal car crash to false facial recognition matches, where Artificial Intelligence (AI) failed to deliver. Talking about examples of where AI went wrong is not, as indicated in [2], intended to put down AI or minimize AI research. The idea is to take a look at where and how it went wrong, with the hope of creating better AI frameworks in the future.

In a 2021 interview [3], Andrew Ng explains that: "Those of us in machine learning are really good at doing well on a test set, but unfortunately deploying a system takes more than doing well on a test set". He gave the following very interesting example: "when we collect data from Stanford Hospital, then we train and test on data from the same hospital, indeed, we can publish papers showing [the algorithms] are comparable to human radiologists in spotting

certain conditions. It turns out [that when] you take that same model, that same AI system, to an older hospital down the street, with an older machine, and the technician uses a slightly different imaging protocol, that data drifts to cause the performance of AI system to degrade significantly. In contrast, any human radiologist can walk down the street to the older hospital and do just fine.". As explained in [4], the American College of Radiology survey is in agreement with A. Ng, "A large majority of the FDA-cleared algorithms have not been validated across a large number of sites, raising the possibility that patient and equipment bias could lead to inconsistent performance".

The author of [5] discusses ten mistakes often made in machine learning. He groups them into three sections based on the type of issue at hand: Data Issues, Modeling Issues, and Process Issues. For the data issue type he describes two

problems: not looking at the data and not looking for data leakage, see [5]. There is not just one reason that cause AI systems to fail, nor is there a clear solution to any of them. Nonetheless, getting to know the data to be dealt with in depth is of crucial importance. As D. Spiegelhalter explains in an interview [6] about his interesting book [7], "I think the ability to deal with data critically and to realize its strengths and limitations is the most important skill in the future world and it's an extremely marketable skill as well". This paper provides an in depth study of Whole Slide Image (WSI) acquisition and processing for artifacts detection, color variation, and data augmentation because, as indicated in the title, the devil is in the details and we should get to know them better.

WSIs play an important role in cancer diagnostics, among other things. Cancer is one of the leading causes of death worldwide, with nearly ten million deaths in 2020 [8]. The gold standard for the diagnosis of many cancer types is the examination of histopathological images by pathologists [9], traditionally under a microscope and digitized in recent years thanks to the advances in Digital Pathology (DP). Among other advantages, the digitization of the slides as WSIs makes it possible to create a digital archive of images and to develop Computer-aided Diagnosis (CAD) and prognosis systems. Machine learning based systems designed to assist pathologist, by predicting diagnosis, prognosis, segmenting, extracting Region of Interest (ROI), visualization etc., can be referred to as computational pathology (CPATH). Throughout the paper, we will use CPATH as an umbrella term for such systems. CPATH can be defined as a branch of pathology that involves computational analysis to digitized pathology images in combination with their associated meta-data, typically using AI methods such as Deep Learning (DL) [10]. CPATH systems can surpass the human eye in the assessment of smaller tissue characteristics in reasonable time and with considerable accuracy [10], and are providing necessary automation to mitigate the burden of the projected rise of cancer rates. DL techniques drive CPATH systems to perform faster and more accurate diagnostics in complex scenarios [11]. Some recent works in this area like [12] have attracted the interest of researchers and the media. In fact, research and the development of CPATH systems have observed a five-fold increase in the last five years [13].

DL based CPATH systems depend on the WSIs used to train them. Inappropriate training data can hamper the performance of CPATH systems and make them useless in unseen scenarios, for example WSIs from different laboratories [14], [15]. For that reason, obtaining and preparing data for the development of new systems often requires a great amount of time and effort. Assuming that the data is already gathered, the preprocessing stage is considered to require more than the 50% of the total effort [16]. Here, preprocessing is understood to be any analysis, cleaning or transformation process applied to the data before it is fed to a CPATH system, including image processing techniques when image data is used.

To focus on WSI preprocessing we must first look at how those images were obtained. The routine of acquiring histopathological glass slides often introduces different unintentional artifacts and variations due to manual tissue preparation, staining, and scanning hardware [15]. Artifacts such as folds, knife marks, creases, and tears add irrelevant morphological features that do not contain any histological information [17]. In addition, health systems manage huge collections of digital glass slides in central repositories from distant laboratories. WSIs collected from different laboratories may exhibit vast differences in a clinical nature of cancer, type of biopsy, color, age of the slide, tissue placement, and file formats. These abnormalities and variations are known to affect the performance of CPATH systems. By using preprocessing, however, it is possible to ensure sufficient relevant tissue patches and thus improve the overall performance of automated diagnosis [18].

This study starts with the WSI acquisition procedure in order to review the causes for WSI variations and to provide an overview of the crucial preprocessing steps for histological images. It explains how to handle a WSI and how the literature deals with the presence of unintentional artifacts in a WSI. Approaches to dealing with WSIs that contain significant cauterized, folded, or blurred areas that must be identified and removed before being fed to DL models are presented. Then, we explain how color discrepancies are resolved in the literature by using color deconvolution, color normalization or color augmentation. Finally, data augmentation techniques applied to histological images are also studied, including morphological, color, and generative approaches. Figure 1 presents a graphical overview of this study.

A. RELATED WORK

In recent years, several studies have compared histological image preprocessing techniques. They can be roughly classified in two branches: image processing applied to histological images and how this affects CPATH systems. In the first branch, Roy *et al.* [19] compared the performance of color normalization methods using similarity metrics. Tosta *et al.* [15] provided a more extensive comparative analysis, focusing on H&E images. In these articles, special emphasis is placed on methods that deal with color variation and other issues such as artifacts detection or augmentation techniques are ignored. Smith *et al.* [20] discussed preprocessing workflow by limiting artifacts detection to quality check and not specifying their histological occurrence during the WSI acquisition phase. Salvi *et al.* [18] reviewed pre- and post-processing in a more general way, broadly including tissue segmentation, artifacts detection, color normalization, and patch selection techniques but not specifying the effects of preprocessing on diagnosis performance. In this sense, the work by Tellez *et al.* [14] compared the effect of color normalization and augmentation methods on the performance of Convolutional Neural Networks (CNN) in several diagnostic tasks.

IEEE *Access*
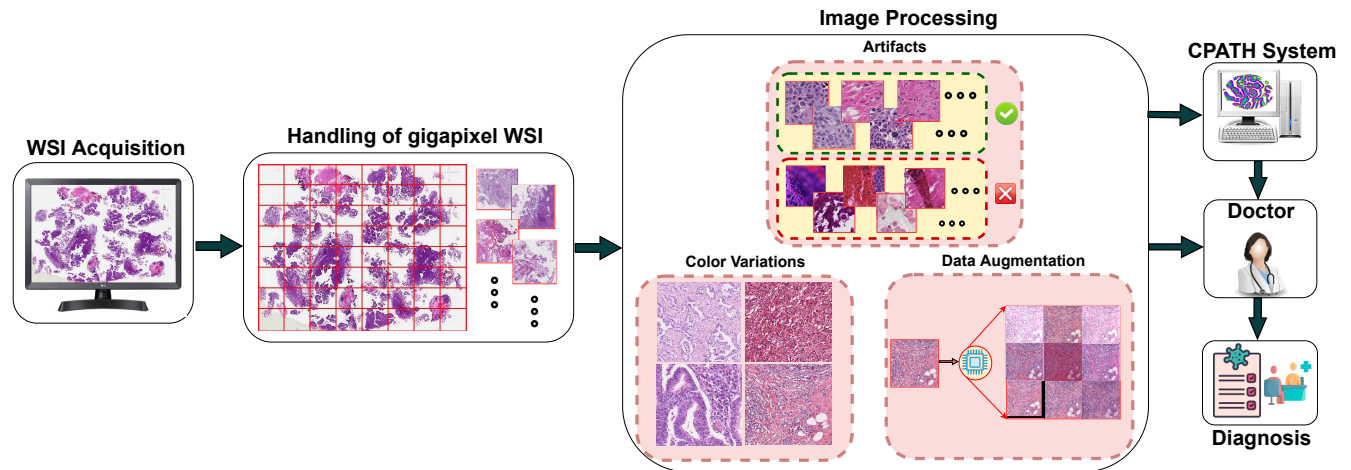
FIGURE 1. An introductory overview to this review. Gigapixel images acquired from WSI acquisition process are handled and split before applying processing methods. These processed histological images are used later by CPATH systems and doctors for diagnosis.

Works in the second branch are more focused on the performance of different CPATH systems but also mention preprocessing. Srinidhi *et al.* [21] comprehensively surveyed different learning strategies for common segmentation and classification tasks for various cancer types. They also included a brief discussion on domain adaptation and color normalization techniques. Dimitriou *et al.* [22] published digitization processes and annotation methods for patching but did not explain how to deal with preprocessing tasks such as artifacts or chromatic variability. Saxena *et al.* [23] reviewed feature extraction and data augmentation methods for breast-cancer datasets. Komura *et al.* [24] investigated applications of DL for histological image analysis and reviewed typical problems to be addressed in order to work with histopathological images, including color variation and artifacts. Gurcan *et al.* [25] presented a methodological review for detection and segmentation tasks with preprocessing steps that were limited to the normalization of illumination. The work by Tosta *et al.* [26] focuses on reviewing validation techniques for segmentation of lymphoma lesions where preprocessing and postprocessing were limited to color-space conversion and morphological operations respectively. Huang *et al.* [27] focused on the applications of CPATH systems in medicine, mentioning the need for preprocessing and feature extraction before carrying out diagnostic tasks. Morales *et al.* [9] identified the standardization of the WSIs in terms of artifacts and color variation as a challenge within current computational pathology.

The works discussed above show the relevance of WSI preprocessing. However, their main focus is to compare CPATH systems and learning approaches. The only work from the first branch that is more focused on preprocessing is that of Salvi *et al.* [18] and includes an overview of the different tasks required. In many cases, WSI preprocessing is reduced to color normalization [15], [19] and other relevant tasks such as artifacts detection or different approaches to dealing with

color variation [14], [28] are ignored.

In this work, we aim to focus on the different areas of WSI acquisition and preprocessing, and review the state-of-the-art approaches for histopathological images. First, Section II introduces WSI image acquisition, which is required to understand the WSI-specific preprocessing techniques. Then, Section III discusses WSI handling techniques, as usually it is not possible to work with the complete images due to their massive size. Next, we cover three WSI-specific image processing techniques: artifacts detection, color variation, and data augmentation. Section IV focuses on dealing with several types of histological artifacts that affect the performance of CPATH systems. Section V describes color variation and the different approaches to avoiding color generalization errors. Section VI introduces data augmentation methods for histological images. Finally, Section VII analyzes the challenges in WSI preprocessing and Section VIII concludes the paper.

II. WSI ACQUISITION
In this section we review the different steps for WSI acquisition, from acquiring the tissue sample to the digitization on the scanner, and explain the impact they have on the obtained images. Artifacts and variations are introduced during the acquisition process; therefore it is vital to understand the acquisition steps in order to apply WSI-specific image processing techniques.

The process of creating a fine quality histopathological glass slide for diagnostics requires competence and skills in both surgical and laboratory techniques [29], [30]. Although the steps in the WSI acquisition procedure are fixed, they are sensitive to a wide range of variables. Variations in chemicals, time, and temperature, among others, make it almost impossible to establish a standard routine among laboratories. Although the staining of the tissue is usually the biggest difference between laboratories, the final appearance

of the sample is affected by every step in the procedure [31]. Some artifacts might be minimized with expertise and precautionary measures [30], while minor variations cannot be entirely controlled. The steps in the acquisition sequence are described below and illustrated in Figure 2. Note that some artifacts are mentioned here to better illustrate the effect of the step, and more detailed artifacts information is provided in Section IV.

1) Biopsy

First, a tissue sample is obtained from the patient's body with a tissue-specific biopsy procedure performed by surgeons and paramedic assistants. The biopsy can be carried out by scraping or brushing the surface of the tissue or by removing the whole tumor. Some of the most commonly used techniques are: Needle biopsy, to take samples from inside the body (e.g. liver or prostate), punch biopsy, to remove cylindrical skin sections, and endoscopy or cystoscopy, to obtain samples from areas that are hard to reach (see [32] for a complete list). The response to the chemicals in subsequent steps will be affected by the type and size of the sample. In addition, some types of biopsies may introduce artifacts. Blood hemorrhages are a common complication [33] when using scalpels. Tissue can also be damaged due to the surgical tools or heat used during the extraction [31]. Also, a sample might be contaminated with coloring agents used to identify the area to be removed [34], or even by tattoo ink.

2) Fixation

Once the biopsy has been obtained, fixation is carried out as soon as possible to preserve the tissue and cellular structure and avoid deterioration. The fixative that is used (e.g. Phosphate formalin, Picric acid, B-5 fixative, Bouin's solution [35]) depends on the tissue type. Fixation time varies with the size of the tissue and between laboratories, lying in the range of 24-48 hours. Both the choice of fixative and fixation time affect how stains will bind to the tissue. An improper fixation affects the details in the sample, reduces the contrast and differences between dyes, and might even create undesired pigments [31]. Large samples or fixatives with poor penetration rates might produce uneven color during the staining step. Fixation carried out with freeze-drying methods forms ice crystals which may cause tissue distortion.

3) Dehydration

Dehydration is performed next to remove aqueous fixative fluid from tissue using alcohol. Different alcohol concentrations are used for varying time intervals. Water drops or excessive time in alcohol will affect the staining quality and may cause tissue shrinkage [36].

4) Clearing

Removing any dehydrating agent or alcohol left in the tissue using a xylene immersion is required for the subsequent steps. Alcohol residues might avoid posterior staining of certain areas, while excessive clearing times might cause

tissue brittleness, crystallization and crumbling during sectioning [34].

5) Embedding

Next comes tissue embedding, which is essential for preserving the structural appearance for the sectioning process. Specimens are enclosed in a supporting medium using a mold. Paraffin wax is most commonly used for embedding, and other mediums (e.g. Acrylic resins, Paraplast, or Polyfin) might be used depending on the tissue type and the sectioning tool, also affecting sectioning and appearance when using light microscopy [37]. The orientation of the tissue in the block is critical. Incorrect placement may damage diagnostic elements during sectioning or obscure them from further analysis.

6) Sectioning

The embedded tissue is chilled at a suitable temperature and then sectioned into thin slices using a microtome. Ideally, successive sections will stick edge to edge, forming a ribbon. Proper sectioning results in uniform thickness and depends on many factors including temperature, knife angle and cutting speed, and requires good handling experience and high-grade equipment [37]. The slicing typically varies from 2-10 μm. A thick section may be opaque and get heavily stained compared to a thin section, thus uniform thickness is highly desirable. Insufficient dehydration, clearing or improper embedding in previous steps or temperature can cause excessive hardening, leading to cracking during sectioning. Similarly, uncalibrated microtomy machines or dull medical-grade blades may tear or stretch the slices [29].

7) Flotation

A thermostatically controlled water bath is then used to flatten the ribbon and to place the sections onto the slide. Wrinkles and folds in the tissue may occur during sectioning and placing [36], [37]. Folded regions are useless for diagnosis and are twice as thick, and consequently absorb more stain. Excessive time in the water causes excessive expansion and thus distorts the tissue [37]. This step is often considered to be part of the sectioning process.

8) Staining

Staining is the process of adding chemical compounds (dyes) in order to highlight structural components of the tissue and enhance the contrast of specific cell types that provide important information for diagnosis. The stains react to the pH or specific proteins in the tissue, giving each element a distinctive color that pathologists can read. Different staining protocols may be selected according to the pathologist's requirements. The most common staining is the Hematoxylin and Eosin (H&E) staining, where Hematoxylin highlights DNA and RNA in purple, whereas Eosin highlights cytoplasm and proteins in pink [37]. Special stains such as immunohistochemical (IHC) are used to highlight different
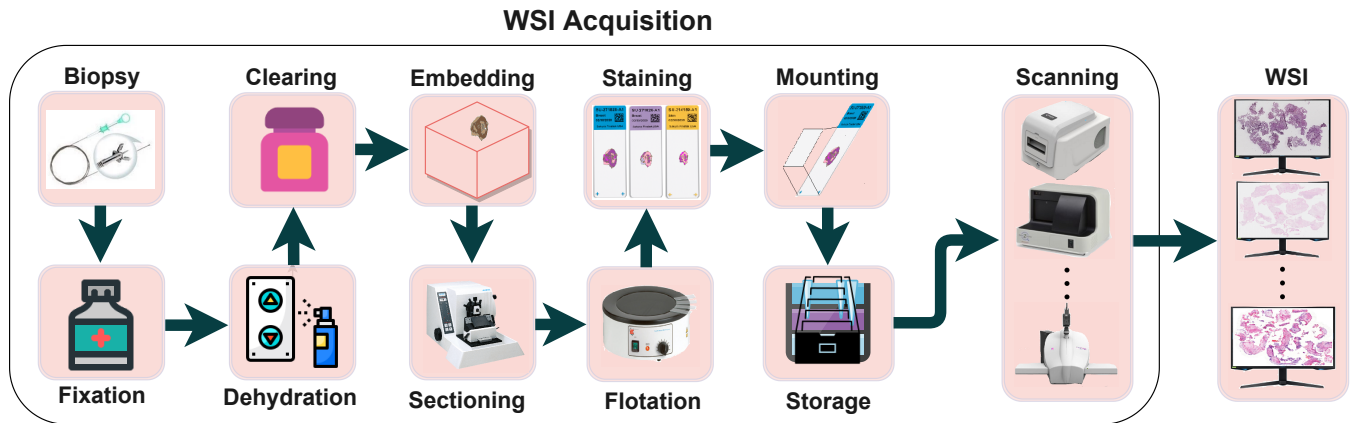
WSI Acquisition

FIGURE 2. A flow diagram of the WSI acquisition procedure. The procedure is subjective to material and human errors that interfere with tissue structure and formation. Different stages are involved in adding artifacts and variations to the final WSI.

proteins in brown. Most color differences are introduced during staining. The previous steps, stain manufacturer, concentration of the mix, mordant ratio, pH, oxidation, temperature, tissue thickness, and staining time are some of the variables that affect the final color. In addition, there is no consensus on the staining protocol, as pathologist might have different preferences over the appearance of the slide [37]. Staining artifacts [31] such as blotching or unstained areas may appear due to wax or alcohol residues from previous steps.

9) Mounting

In the final step in slide, the slides are covered with a mounting media before being protected with cover-glass. Common artifacts that might be introduced during mounting are air bubbles, dust or microorganism contamination.

10) Storage

In some cases, the mounted slides might be stored or even transported between laboratories [9] before scanning or re-scanning. Slides suffer a natural discoloration over time that might render the slides useless. The storage conditions need to be controlled, if slides are not stored in the dark, light might cause the stains to become bleached [31]. Many current archives contain glass slides that were collected over several years.

11) Scanning

Finally, slides are scanned to produce WSIs. Digital microscopy scanners vary widely and have a noticeable impact on the observed color due to scanner-specific illumination of the sample, sensors, and image processing carried out during the image acquisition [15]. Scans might operate with either bright-field, fluorescence illumination, or both [37]. Scanning occurs at different scales or magnification levels of the slide, typically $10\times$, $20\times$, and $40\times$. A vendor-defined pyramidal format stores different zoom levels as a WSI. Metadata such as the storage format, focal profile and other

technical and administrative parameters is often stored within the WSI. It is essential to choose the right focal profile and focal map to avoid blurring artifacts.

At the end of this procedure, the WSI file can be used by pathologists for annotation or diagnosis instead of using the microscope. Usually, scanning system vendors provide specific software to view WSI stored in their proprietary format. The final image perceived by pathologists is also affected by their display system. Feeding WSI images to CPATH systems requires additional steps that are covered in the next section. In addition, these digitized slides are likely to require the application of image processing techniques before they are used in CPATH systems, otherwise the diagnosis could be affected by the undesired variations and artifacts introduced during acquisition.

III. WSI HANDLING

The acquired WSIs contain all the information required to emulate the navigation of a glass slide on a microscope [38]. Several resolutions are available within the same file and thousands of individual images are stitched together, rendering files of gigapixel order. Therefore, the computational cost of analyzing a complete WSI with CPATH systems is usually very high, and it is almost impossible to analyze it all in one go due to memory restrictions [9], [39]. Furthermore, WSIs do not even fit in a GPU, which renders the tasks of processing and automated diagnosis almost impossible. The most common strategy is to analyze WSIs by breaking them down into smaller patches. The patching workflow is usually adapted for the following WSI processing or automatic diagnosis [9], [40]. In general, the process of WSI patching consists of two sub-problems: *what-to-patch*, and *when-to-patch*.

What-to-patch: Processing and analysis will be affected by how the WSI is split in sub-images [40]. Square patches of different sizes are commonly used. Smaller sizes are often used for DL models [41] to reduce the computational burden
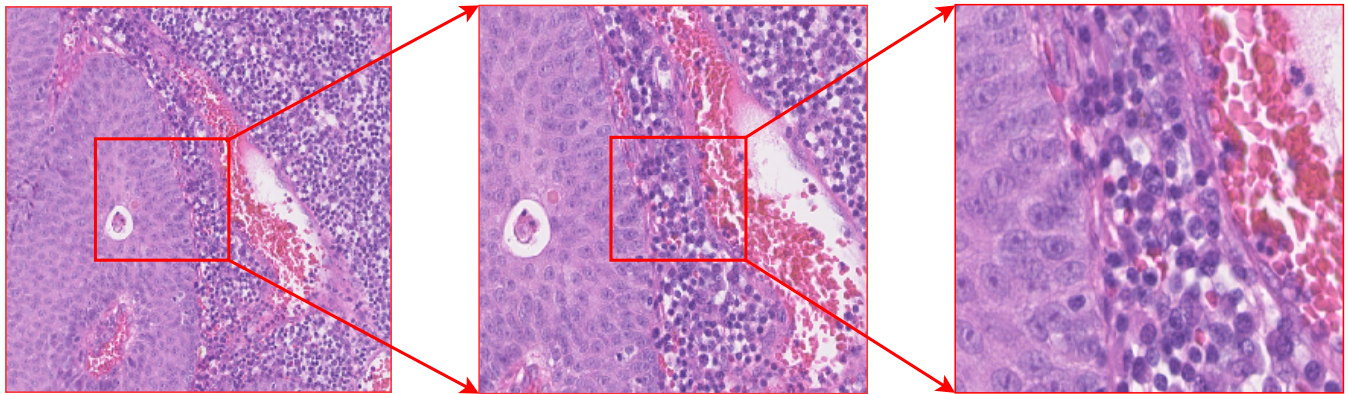
FIGURE 3. Patches extracted from the same WSI at a fixed 256×256 patch size and different magnification levels of 10x, 20x, and 40x (from left to right).

and to avoid excessive information within the patch. In some cases, bigger patch sizes might be needed to capture whole histopathological areas such as complete glands [40]. What is captured in the patch is also affected by the magnification used (see Figure 3). Patches can be extracted at one or multiple magnification levels. Using several levels at once gives a multi-resolution dataset where the same number of pixels in different magnification levels correspond to different fields of view. While this technique can mimic how the pathologists work when zooming in and out to get details or context, it requires complex models in order to be able to handle the different levels [38], [42]. To cover the entire WSI, patches are usually extracted using a sliding window with or without overlapping between patches [41]. If annotation masks are available, the patching can be performed only within ROI [43] or labeled areas, reducing the amount of irrelevant patches. Patching the background is usually avoided by automatically generated tissue masks using Otsu or other thresholding methods [40].

When-to-patch: Patching itself is a time-consuming process that usually is performed in advance and stored for subsequent processing. This pre-patching approach requires the patch settings to be fixed in advance, and requires extra storage for each patching configuration, e.g. if patching with 256×256 resolution and 512×512 resolution is to be tested, the complete patched dataset must be stored for each of the two setups separately. The extracted patches often become the actual dataset, substituting the use of the WSIs [44]. An alternative approach is patching "on-the-fly" [9], where a WSI-specific list of patch coordinates is stored [41], [42]. The extra storage required for different patching configurations is reduced because the patches must not be saved separately, and the flexibility for posterior processing is more flexible in terms of size, resolution, and overlapping. However, patching "on-the-fly" might imply an increase of the processing time as the WSI needs to be loaded and processed during training each time.

Patching makes it possible to load, process, and analyze WSIs, yet it also implies contextual information loss [9].

The best patching option should be chosen according to the task, model, memory, and computational constraints, and is a trade-off between these requirements [9]. Patching details are often briefly mentioned in research papers, but any subsequent steps will be affected by the patching procedure, making reproducible patching critical for reproducible research [41].

At this point, the WSI patches might be used to feed a CPATH system. However, good quality WSI is critical for CPATH systems. The following sections describe different preprocessing techniques that can be used to improve WSI quality and CPATH system performance.

IV. DETECTION OF ARTIFACTS
During the acquisition procedure (see Section II), undesired artifacts might be introduced in the slides. Artifacts are alterations of tissue or artificial structures introduced by extraneous factors [31] that may be present in some parts or even the whole WSI [33] and might hamper the diagnostic procedure. There is a wide range of possible histological artifacts [31], and they can be roughly divided into: 1) Tissue-level artifacts, 2) Slide-level artifacts, and 3) Scanner-level artifacts [45]. Figure 4 depicts some of the artifacts that are introduced in each step of the acquisition procedure.

Tissue-level artifacts: These artifacts are produced during the acquisition and processing of the tissue, from the biopsy to the staining steps (see (II-1) to (II-8) in Section II), or often in various steps. Tissue-level artifacts are often hard or impossible to rectify as this would require repeating the tissue acquisition process or even a new biopsy. Tissue can be damaged in the biopsy (II-1) (cauterized tissue, curling, squeezing, and hemorrhage) often eliminating the diagnostic value of the damaged areas [29], [33]. Several types of tissue ruptures can be produced during sectioning (II-6) as a result of the preceding steps (e.g., ice crystals due to inappropriate fixation (II-2), brittle tissue due to excessive clearing (II-4), or using a hard embedding (II-5)) [34], [36]. Improper orientation during embedding causes tangential sections that might not be of interest. Sectioning (II-6) might also cause

artifacts to occur (e.g., tears caused by a dull knife, chatters and cracks due to knife vibration, or uneven tissue thickness). Special care has to be taken during sectioning and flotation (II-7) to avoid tissue overlapping, referred to as tissue folds [34]. Staining (II-8) might also produce artifacts that are influenced by previous steps, such as blotching and unstained areas caused by embedding (II-5) and clearing (II-4) residues, respectively [34].

Slide-level artifacts: These artifacts are associated with the final pathology workflow steps, such as mounting and storage (see (II-9) to (II-10) in Section II), and can be resolved by repeating just these steps. Some of these artifacts are types of contamination such as dirt, fungi or microorganisms before mounting, or air bubbles produced when placing the cover slip [34]. Pen markings from previous manual analysis or damage due to improper storage are also considered slide-level artifacts. Slides with pen markings and dirt can be cleaned prior to scanning.

Scan-level artifacts: These artifacts are caused during scanning (see (II-11) in Section II) and do not appear on the glass slides. They can be easily solved by re-scanning if necessary. Blur is the most common scanner-level artifact that diminishes the overall sharpness of a WSI. It is produced by uneven tissue thickness or improper focal calibration. Modern microscopy scanners try to avoid blurring artifacts by selecting multiple focal points to adjust the focus to tissue height [46], [47], but having more focus points usually means longer scanning times. Other scanning artifacts appear due to hardware limitations. The glass slides often need to be scanned in separated pieces that are latter stitched together to create the WSI. The most common approaches are line and grid scanning, which may cause a strip-like or grid-like appearance, respectively, if not well illuminated [17].

The presence of these artifacts directly affects the performance of the CPATH systems; thus, it is crucial to detect WSIs or patches containing artifacts [48]. Having provided a general classification of the different artifacts that can occur during the WSI acquisition process, we will now focus on further detailing scan-level artifacts, blur and the tissue-level artifacts, folded tissue, blood, and damaged areas because those artifacts can have a major impact on WSI analysis and have been somewhat explored in the literature. At the end of this section, we will also address the general WSI quality assessment techniques.

A. BLURRED AREAS

Blur is often considered the most critical quality issue in WSI [49]. Methods that can objectively quantify the presence of blurry patches can be classified in No-reference, partial-reference, and full-reference methods [50]. Full and partial-reference methods require a non-blurred reference image. Unfortunately, references are not usually available, thus, no-reference metrics are usually preferred [51]. No-reference metrics assume that the distribution of the blur metric is different in sharp and blurry patches [50]. Using no-reference metrics, Wu *et al.* [50] proposed a workflow to classify

blurry and sharp regions in endomyocardial WSIs by determining pixel-level information and bin distribution. Local and global features were compared using several classifiers, where higher accuracy was achieved using the local features.

Gao *et al.* [52] detected in-focus and out-of-focus WSI regions by extracting 44 extensive features (e.g. neighborhood contrast, gradient and Laplacian features, local statistics, and wavelets) and training an AdaBoost classifier. Deep-Focus [47] uses a CNN to analyze blur in four different stains (H&E, Ki67, CD10, and CD21) with categorical cross-entropy loss. The approach used data augmentation (see Section VI) and was evaluated in a limited test set in terms of accuracy. The work by Albuquerque *et al.* [53] compared seven CNN architectures to classify blur for different focus levels. Their work detailed benefits of data-driven methods over knowledge-driven methods in terms of performance metrics. The method compared ordinal loss with nominal cross-entropy loss for multi-class focus assessment. Campanella *et al.* [54] quantified different blur levels using sharpness-based features along with a random forest model and residual network. Kohlberger *et al.* [55] proposed ConvFocus CNN architecture, to quantify and localize out-of-focus areas in a WSI. Their focus quality evaluator was trained on semi-synthetic data to learn discriminative features and was validated on limited real data. Similarly, Ang *et al.* [56] proposed FocusLiteNN, a data-driven method used to evaluate focus quality in various stains. FocusLiteNN uses a shallow CNN layer to transform features with transferability and relatively low complexity.

Once blur areas are detected, they are often discarded or re-scanned if possible. Deblurring and Super-Resolution (SR) of histological images can also be found in the literature. Zhao *et al.* [57] proposed a residual dense convolutional network for image deblurring in optical microscopic systems. Mukherjee *et al.* [58] built a recurrent SR network in order to use the intermediate resolutions available in the WSI to reconstruct a high-resolution image. Chen *et al.* [59] extended this work by linking a multi-scale SR network and diagnostic network. Singh *et al.* [60] proposed the idea of using a dark channel algorithm designed for haze removal in natural images to enhance medical images.

B. TISSUE FOLDS

Tissue folds are tissue-level artifacts that occur during the flotation step when a layer of tissue is placed over itself. In the folded areas, overlapping tissue might introduce morphological aberrations (e.g., overlapped nuclei) that may cause misinterpretation [61]. The tissue thickness is also increased by the additional layer, adsorbing more of the stain than the rest of the tissue on the glass slide.

The color difference in folded tissue has been used to identify these areas [61]–[63] by using color-space transformation. Palokangas *et al.* [61] developed an unsupervised approach using differences in the Hue, Saturation, and Intensity (HSI) channels as shown in Figure 5. Then, folded tissue was identified by the use of clustering. This method evaluates
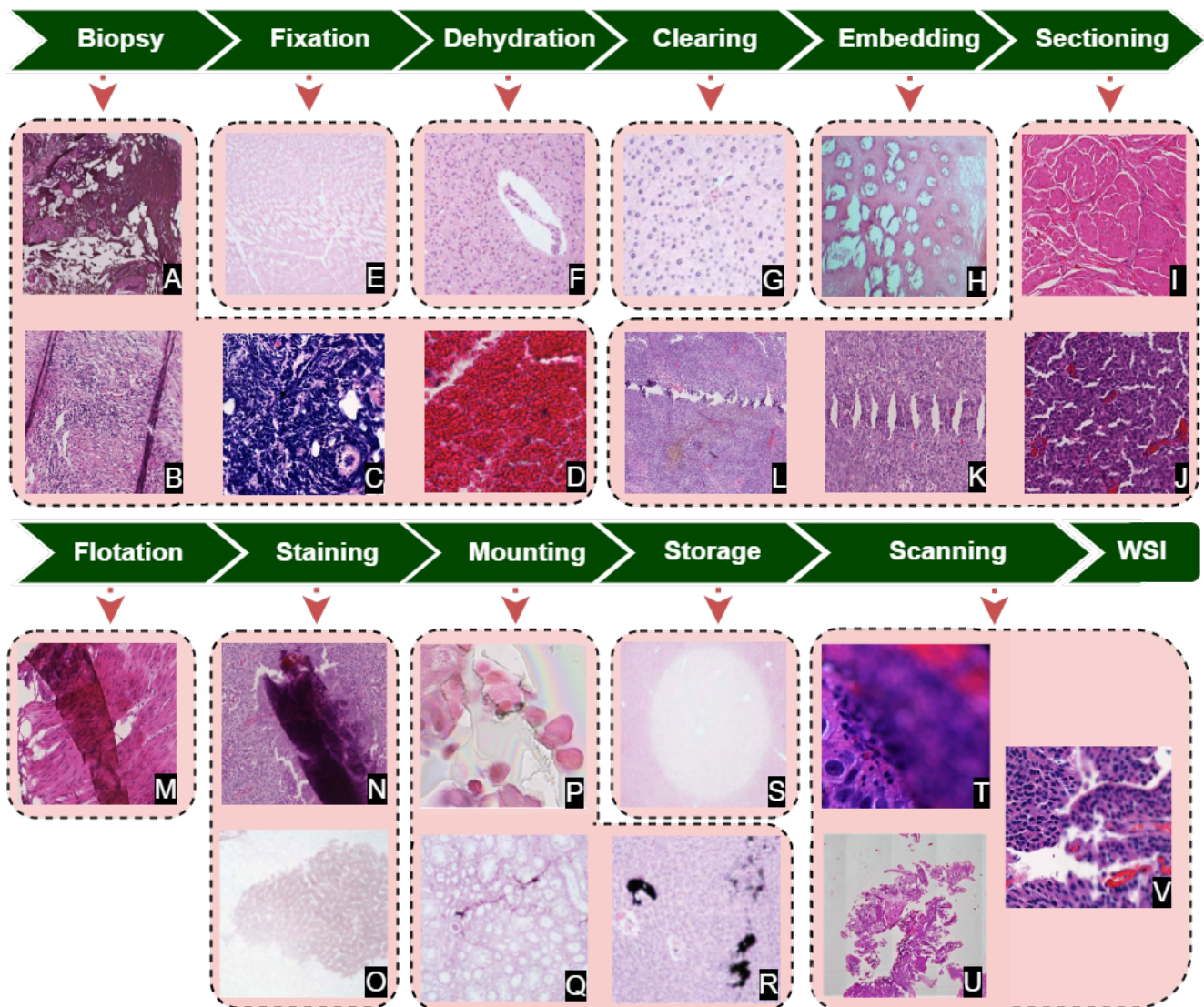
FIGURE 4. Artifacts can be introduced during the WSI acquisition (see Section II). (A),(B), and (C) are damaged tissue artifacts caused during biopsy due to heat, curling or squeezing, respectively. (D) represents a blood hemorrhage with no diagnostic value. (E) shows tearing of the tissue due to ice crystals caused by freeze-drying fixation methods. The shrinking of tissue due to excessive alcohol dehydration is shown in (F). (G) shows crystallized brittle tissue due to excessive cleaning. (H) shows a tangential section due to improper orientation of the tissue within the embedding block. During sectioning, tearing (I) and cracks (J) may occur due to flaws in the embedding procedure or improper temperature, while chatters (K) and tears (L) are caused by a loose and dull knife, respectively. Folded tissue is shown in (M), caused by an incorrect slide placement during flotation. Staining artifacts in previous steps (e.g., blotching (N) and unstained areas (O), might be due to residual wax or xylene, respectively. Air bubbles (P) and contamination due to microorganisms (Q), or dirt (R) might occur during mounting. The natural discoloration over time is accelerated by light exposure as shown in (S). Finally, blur artifact (T), strip-like appearance (U), or stitching (V) can be produced by the scanner due to improper calibration.

WSI based on the assumption that folds are present and will result in false positives in the absence of folds. Kothari *et al.* [63] combined RGB, HSI, CIELUV and CIELAB features with texture features such as Gray-level Co-occurence Matrix (GLCM) to discard tissue folds and pen marks.

Bautista and Yagi [62] detected folds at low magnification to try and to avoid these areas in the selection of focal points for the scanner. They used the RGB shift with an adaptive factor depending on saturation and luminance values to distinguish between tissue folds and the rest of the tissue. Although the authors recognize that their method could ig-

nore small tissue folds due to low magnification, it was not tested on higher magnifications. Wang *et al.* [64] extended the work by Palokangas *et al.* [61] by adding connectivity properties of tissue structures to detect tissue folds in low magnification WSIs. Although the method adapts the fold detection thresholds based on neighboring pixels, it needs to be optimized according to the dataset.

The use of color-based features might be affected by varying staining protocols. To overcome this issue, Shakhawat *et al.* [48] proposed the use of data-driven features trained with heterogeneous datasets. They explored the use of GLCM to

IEEE *Access*

RGB to HSI Transformation
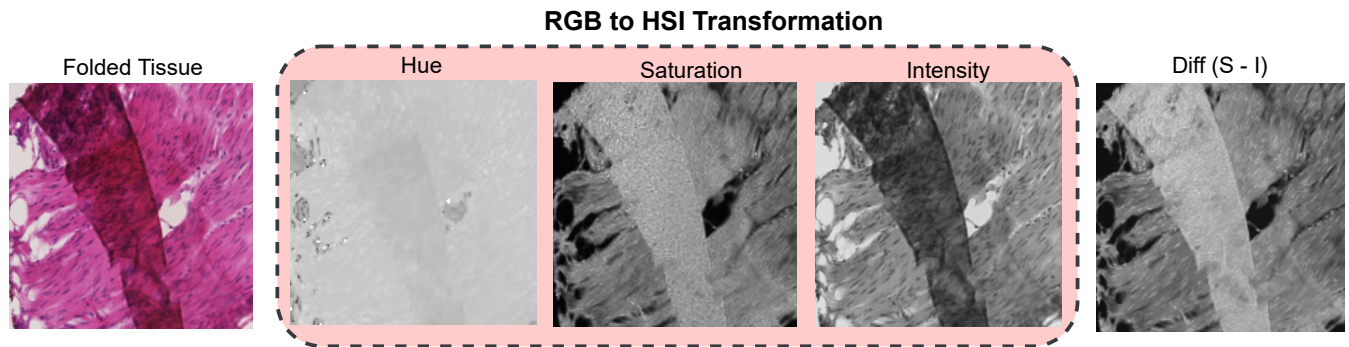
FIGURE 5. HSI transformation of a patch with folded tissue artifacts. Altered regions can be highlighted by different components of HSI color-space due to their colorimetric properties. An enhanced observation can be formed by subtracting of two prominent components [61].

feed a binary Support Vector Machine (SVM) classifier and detect folds at low magnification as a quality check step. Babaie *et al.* [65] proposed the use of five well-known pre-trained CNN to extract deep features that were then used to classify tissue folds with different classifiers (Decision trees, SVM and KNN).

C. DAMAGED AND BLOOD AREAS

Damaged tissue (e.g. cauterized, squeezed, etc.) and blood hemorrhages result from complicated specimen collection procedures such as trans-urethral resection in bladder cancer (see (II-1) in Section II). These regions are considered to be tissue-level artifacts and are often ignored [66] due to the lack of information relevant to diagnosis or prognosis [38]. Similar to folded tissue, damaged and blood areas differ in terms of stain absorption and thus can be separated with color histograms and texture features [67]. Despite the diagnostic irrelevance, there are not many publications which have focused on finding damaged tissue or blood. Some research focused on finding diagnostically relevant tissue include them as a class to discard. The method by Bahlmann *et al.* [68] flagged irrelevant patches using the percentile of the stain channels by [69] and a linear SVM. Mercan *et al.* [67] used k-means to find a dictionary and represent the WSI as a bag-of-words. The patches were classified into clusters using combination of Local Binary Patterns (LBP) extracted from the stain channels provided by [69] with L*a*b histograms. Blood was identified as one of the clusters. Wetteland *et al.* [66] presented a segmentation CNN to find several tissue classes, including blood and damaged tissue with the aim of finding relevant tissue. This work was extended to multi-scale in [38] using global and local context from different magnification levels, and combined with clustering to include low-probability patches in [70]. In Chadaj *et al.* [71] proposed a U-Net model to detect damaged tissue. Although the technique was only tested on IHC stained brain tissues, the authors considered its possible use for WSI analysis of other tissues and stain protocols.

Although blood is usually non-informative, in some cases it is critical for diagnosis. Chadaj *et al.* [72] tackled the problem of differentiating blood vessels (informative) from hemorrhage (uninformative) using the CMYK color-space and mathematical morphology to feed a decision tree. Blood detection is also a critical step in the diagnosis pipeline presented by Clymer *et al.* [73], where a RetinaNet model is used to detect blood vessels at low resolution, which were subsequently classified using an Xception CNN.

D. OVERALL QUALITY CHECK

We conclude this section on artifacts detection by describing the overall Quality Check (QC) models that have been proposed in the literature. Some of them have already been mentioned in Sections IV-A to IV-C.

WSI analysis is computationally expensive. The objective of a QC is to quickly identify faulty WSI containing significantly distorted features in order to discard or re-scan them before carrying out any further analysis [17]. QC approaches try to evaluate each WSI and provide a quality metric, usually by looking at lower magnifications in the multi-resolution pyramid [48], [49], [74]. The overall quality metrics are designed according to what pathologists consider to be the most relevant features and often include sharpness of the image, amount of artifacts or noise, and contrast metrics [49]. Most of the time, it is not possible to obtain an ideal high quality reference image. Thus, no-reference (blind) quality assessment methods are preferred for robust QC [51]. A general QC pipeline is shown in Figure 6. QC methods are a coarse approach to artifacts detection. Notable artifacts such as folded tissue, air bubbles and blur are often treated as a single class in QC approaches.

The method presented by Hashimoto *et al.* [51] uses a combination of image sharpness and noise measurements to derive a linear regression model in order to provide a quality metric. This work was extended by Shakhawat *et al.* [48] to distinguish whether the low quality of an image was caused by scanning or other artifacts. Ameisen *et al.* [74] proposed a set of QC metrics using blurriness, color separation, brightness and contrast assessments to evaluate different scanners. They discussed the trade-off between using lower magnification for quick QC or higher magnification for a more com-
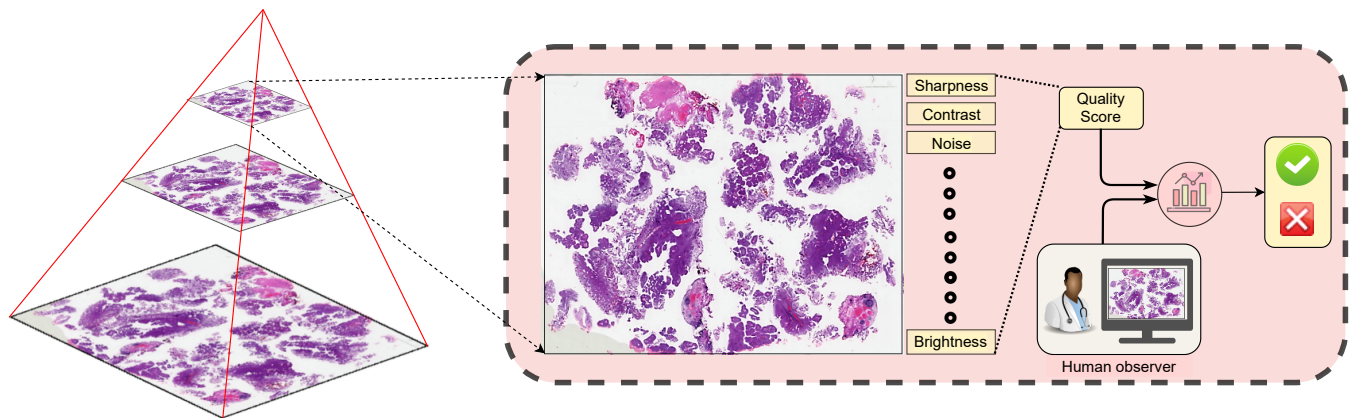
FIGURE 6. Overview of an automated QC at lower magnification. Various image parameters are used to set a quality score. Agreement between final quality score and the pathologists' subjective evaluation is often used to validate the effectiveness of the QC method.

prehensive analysis. Shrestha *et al.* [49] evaluated the quality and reproducibility of different scanners over time using a weighted average of five parameters: sharpness, contrast, brightness, uniform illumination and color separation. The QC by Avanaki *et al.* [75] compared quality estimators based on the availability of the gold standard reference. Structural Similarity (SSIM) is used when a reference image is available for evaluation. For reference-less estimation, a blind feature-enriched estimator was trained on artifacts free WSIs. The HistoQC [76] tool performs quality estimations by making a content-based evaluation for outliers in a cohort of WSIs. It uses image metrics, features and supervised classifiers to distinguish altered areas in the images. It also takes into account the WSI metadata to catalogue and contextualize visual information. Jimenez *et al.* [77] used full-reference and blind metrics from general image quality assessment and used them for WSI QC.

V. PROCESSING OF COLOR VARIATIONS

Once our WSIs are artifacts free, one might think that they are ready to be used in DL related tasks. However, even when using the same staining protocol (e.g. H&E, IHC), the color observed in WSIs strongly varies between different laboratories. The final color of the sample is affected by every step described in Section II. This makes it impossible to avoid color variation during WSI acquisition.

Although color variation does not usually affect the analysis and diagnosis of images by doctors, it hampers the performance of CPATH systems. The impact of color can be especially severe when working with data from several laboratories or when testing systems on data from new laboratories [14]. It is probably the most studied phenomenon in histopathological image preprocessing [14], [15], [18]. Therefore, addressing color variation is one of the main preprocessing tasks required to obtain reliable data that can be used with transferable CPATH systems.

To reduce the effect of color variation in the WSI analysis, several approaches can be found in the literature: Grayscale

conversion, Blind Color Deconvolution (BCD), Color Normalization (CN), and Color Augmentation (CA). Although most works dealing with color variation focus only on CN, it is interesting to provide an overview of all the approaches. We introduce a brief description of them in the following.

Grayscale conversion: Discarding the color information and using grayscale images is the most naive approach. It is supported by the hypothesis that color information is redundant since the diagnosis relies on morphological and structure patterns [14]. Although grayscale images are not commonly used, they can reduce the CPATH system generalization error for unseen colors [14]. However, it has been shown that discarding color information results in lower classification performance than other preprocessing techniques [14].

Blind Color Deconvolution: The basis of the staining procedure is to be able to differentiate the structure of the tissue according to the distribution of each stain [78]. BCD techniques aim to separate the observed multi-stained image into single-stain images. The separation is performed by estimating the color of the stains in the image, and the amount (concentration) of each stain for each pixel. A graphical representation of the BCD procedure is depicted in Figure 7. The amount of each dye absorbed by the tissue, which has been separated from the color information, can be used to feed CPATH systems instead of using the RGB channels directly [79]. This approach reduces the impact of color variation and tries to mimic how pathologists analyze the image as they identify the different tissue structures in the image by the amount of each stain.

Color Normalization: Nowadays CN is the most popular procedure for dealing with color variation (see the review in [15]). CN methods aim to adjust the color in the WSIs as if they were obtained according to the same staining and scanning procedure. Although BCD is the major first step in most CN methods, some of them achieve normalization without stain separation, e.g. by using style transfer [80] or global normalization [81]. Particularly, most recent methods based on deep generative models, i.e. variational autoen-

coders and generative adversarial networks, perform CN without BCD [80], [82].

Color Augmentation: Unlike BCD and CN, which aim to reduce color variation, CA [14] aims to generate color variations in training data, reducing the generalization error of classifiers on future test data acquired with different color properties.

In the following we will describe the BCD and CN techniques in detail and postpone the analysis of CA techniques (on account of their different nature) to Section VI, where it will be discussed together with other augmentation techniques.

A. STAIN SEPARATION USING BCD

Differential staining lies at the basis of pathology, providing information about the distribution of the structures within the tissue [78]. Classical BCD works [78], [83] were designed to help pathologists during the manual diagnosis. The use of BCD to deal with color variation came with the development of complex CPATH systems that use more than just shape features [84]. As BCD separates staining structure from its color information, it is ideal for reducing color variation while preserving structural information. The separated stain channels can be used for CN, CA, to obtain channel specific features [40], or directly for classification [85]–[87] which seems to improve the classification performance of the tested systems.

The work by Ruifrok *et al.* [78] introduced the use of the Beer-Lambert law and the *Optical Density* (OD) space, producing a linear representation of the combination of stains. The OD image can be separated into the concentration matrix and the color-vector matrix. Later works use the OD space and propose different approaches to find the color-vector matrix, which is considered to be unknown due to the color variation. As the staining procedure is additive, [83] proposed Non-Negative Matrix Factorization (NMF). This work was extended in [88] using regularization and sparsity terms. The work by Vahadane *et al.* [89] only used the sparsity term, with the assumption that a type of stain is only bound to certain structures. The sparsity term was revisited in [90] where the authors estimated the sparsity parameter using a fuzzy set method. Independent Component Analysis (ICA) was also explored in [83] and further developed in [91] with a stain vector correction step. Alsubaie *et al.* [92], [93] explored its use in the wavelet domain where the independence condition among stains is relaxed. Macenko *et al.* [84] proposed the use of Singular Value Decomposition (SVD) to separate H&E channels. Although the method in [84] is still commonly used, it was extended in [94] by taking outliers and the interaction between dyes into account. The authors of [95] used SVD in a linearly inverted RGB-space instead of the usual logarithmically inverted OD space. Clustering techniques have also been explored to obtain the color-vector matrix: In [96] the authors introduce the use of priors for the color vectors and use k-means to estimate the actual color. This work was extended in [97] by using the Maxwellian

chromacity plane to identify the reference colors and in [98] with k-means and considering a possible imbalance of the stains. The work in [99] adapts the deconvolution proposed in [78] by including a prior knowledge based optimization problem. Recently, Salvi *et al.* [100] proposed an adaptive refinement of the method in [84] using Gabor-filters and k-means to detect nuclei and stroma. A segmentation Gaussian Mixture Model (GMM) method was proposed in [101] to estimate the color-vector matrix and then extended in [102] with an image-specific color descriptor and a more robust color segmentation framework. Bayesian inference was applied by Hidalgo-Gavira *et al.* [103] introducing the use of a similarity prior on the color-vectors and a smoothness prior model on the concentrations. The Bayesian approach was also utilized by Pérez-Bueno *et al.* [85] with the use of a Total Variation (TV) prior. The work in [28] uses the high-pass filtered domain to set sparse general super Gaussian priors on the concentrations. Then BCD problem is approached as a dictionary learning problem in [104], implementing Bayesian K-SVD for BCD of histological images. DL has hardly been applied in BCD, but there are some examples. Duggal *et al.* [87] implements a stain deconvolution layer for CNN based in the use of [84] to provide a stain separated input to CNN-classifiers. Zheng *et al.* [86] use a Capsule Network that produces multiple stain separation candidates using 1×1 convolution operators and finally assembles the output based on a sparse constraint.
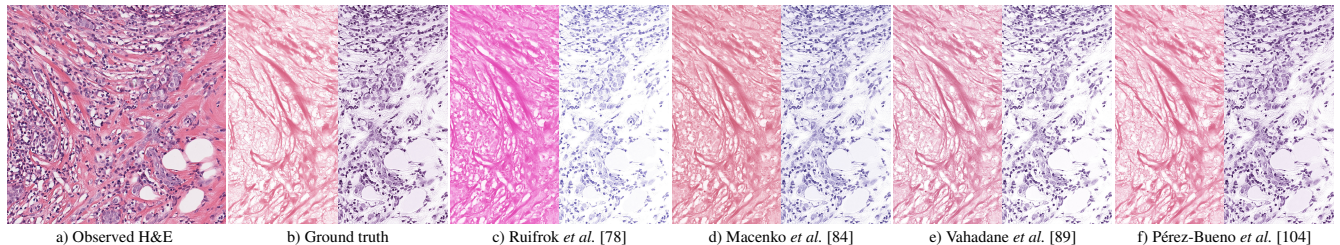
Figure 8 depicts the stain separation obtained by some of the different methods in the literature. Different estimations of the color-vector matrix and stain concentration will have an impact on posterior applications of the BCD results, like CN or CA.

B. COLOR NORMALIZATION

When CPATH systems started to use color-based features rather than only morphological features [84], CN was proposed in [84] for its use in WSIs. CN aims to obtain standardized images that mimic a chosen staining procedure, often obtained from a reference image. Then, the normalized images can be used as input to reduce the generalization error for CPATH systems trained without considering color variations. Most CNN based CPATH systems use the RGB image as input instead of previously obtained features [27]. The popularity of CNNs has also increased interest in CN. The main concern is that color correction needs to be done while preserving the histological structure. For this reason, most works concerning CN include a previous BCD step. In [105] CN methods are divided into color modification and color separation, where the latter are those that include a BCD step. In [15] the authors classify CN methods as histogram matching, color transfer, and spectral matching [106]. Histogram matching ignores the stain separation. Color transfer might include a segmentation or deconvolution step, it modifies the color using statistical correspondences between histological regions. Finally, spectral matching is a complete BCD approach where stain concentrations and color properties
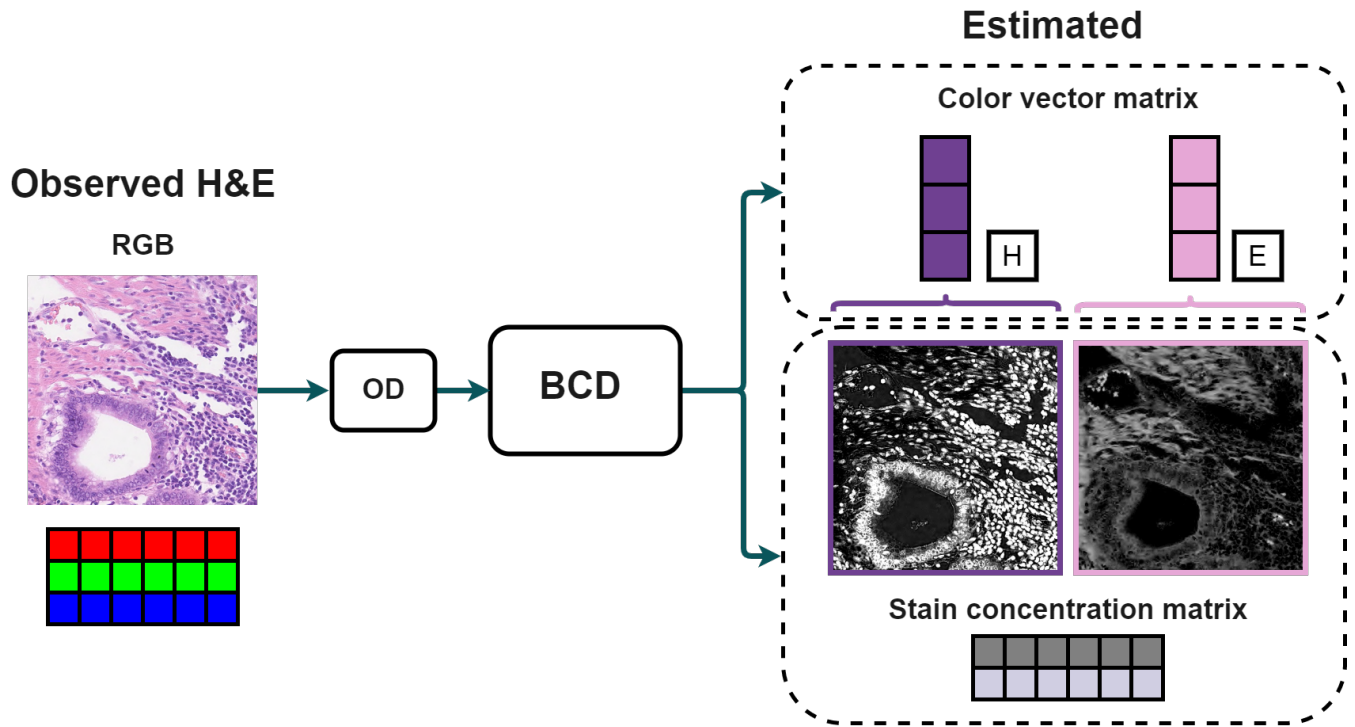
FIGURE 7. Usual pipeline for stain separation using BCD. First, the RGB image which can be seen as a matrix with one row for each RGB channel, is transformed to the logarithmically inverted OD space. Then, different methods are used to separate color from structure. The outputs are the estimated color vector matrix and stain concentration matrix, which contain the color information for each stain (H&E in the example) and the concentration of each stain for each pixel, respectively. The concentration matrix can be seen in the figure with one row for each stain as well as a grayscale image for each stain.

| a) Observed H&E | b) Ground truth | c) Ruifrok *et al.* [78] | d) Macenko *et al.* [84] | e) Vahadane *et al.* [89] | f) Pérez-Bueno *et al.* [104] |

FIGURE 8. a) Ground truth separated E-only (left) and H-only (right) images from a Breast image of the dataset in [93] and the results of several methods in the literature. The differences in the estimated color vector matrix can be appreciated.

are estimated. A more intuitive classification was introduced in [18], where methods are considered as being *i)* global CN, *ii)* CN after stain separation, and *iii)* color transfer using deep networks. We will follow this classification in the following discussion.

Global CN includes methods that do not separate stains before tackling color variations. The work in [107] includes two histogram based approaches to match the original and reference colors. The first one performs quantile normalization of the RGB channels to match original and reference color distributions. The second creates a color map with every unique RGB triplet and employs a mapping function to transform the values to the reference color map. Although it was not proposed for histopathological images, the work by Reinhard *et al.* [81] is commonly cited in the CN literature [14], [18]. In [81] the authors use the $l\alpha\beta$ color space to separate the chromacity channels and then adjust the mean

and deviation of each channel to match the reference image.

CN after Stain separation: Once the color-vector matrix and the stain concentrations are estimated using BCD, it is possible to obtain normalized images. Note that most works discussed in Section V-A were proposed for CN. In [84] the color-vector was replaced by a standard one, and the concentrations are scaled to have the same pesudo-maximum (99th percentile) as the reference image. The same CN procedure has been used by more recent BCD works [79], [89], [94] where the main differences can be found in each method's estimation of the color-vector matrix. The overview of this procedure is depicted in Figure 9, where we can be observe how the BCD estimation affects CN. While preserving the replacement of the color-vector matrix, Khan *et al.* [102] proposed a nonlinear (B-Spline) mapping of the concentrations to the reference image. In [99], the authors calculate the transformation matrix between source and reference image

using an optimization function. The work in [44] did not use BCD but instead separated stains and background classes using the HSV color space, and then scaled the mean and variation of each class separately. Recently, [108] presented a multiscale Retinex model that estimates and corrects the reflectance and illumination map for pixels of both stains separately.

Color transfer using DL: Many of the recent works presented regarding CN use DL techniques. One of the first applications DL to histopathological images was presented by Janowczyk *et al.* [109] using Sparse AutoEncoders. Bentaieb *et al.* [80] used a Generative Adversarial Network (GAN) to combine the normalization and classification of WSIs. The generator is considered to be a stain transfer network, while the discriminator simultaneously separates real and normalized images and also positive and negative classes. The StainGAN model [110] uses cycleGAN architecture to map unpaired images between two different scanners (Aperio and Hamatsu). In [82] Zanjani *et al.* use three different CNN models for CN. First a VAE model is used, in which the latent variable aims to encode K tissue classes, and the decoder obtains the normalized images. Second, a GAN model is used. It receives the lightness channel in the CIELAB color system as input and generates the chromatic channels. Third, a Deep Convolutional Gaussian Mixture Model (DCGMM) that jointly optimizes the combined CNN and GMM models. The GAN approach was used for CN cytological imaging by Chen *et al.* [111] and includes an intermediate style removal step.

Tellez *et al.* [14] proposed a CN network fed with heavily augmented images and trained to reconstruct its original appearance. Other popular CNN architectures have been adapted to stain normalization, such as the Pix2pix conditional GAN framework [112] or CycleGAN [113]. Zhou *et al.* [114] combined a Cycle-Consistent GAN with the color-vector obtained in [89]. Extending the unpaired CycleGAN architecture, Invertible Neural Networks were used by Lan *et al.* [115] to reduce the computational cost by means of parameter sharing. Patil *et al.* [116] proposed a lightweight fully-CNN that is attached to DL-based pipelines like a preprocessing block. Moghadan *et al.* [117] explored the disentanglement of style and content using a VAE architecture. Ideally, the style space represents the color information while the content space represent the structure of a histological image. However, the fidelity of the latent content space in terms of the structure of the image was not assessed. A conditional GAN was used by Ke *et al.* [118] and combined with federated learning to normalize the images to an interpolation of the stain styles in the data clusters.

C. METRICS FOR THE EVALUATION OF COLOR RELATED TECHNIQUES

After having discussed the relevance of reducing color variation in the preprocessing of histological images, we must now evaluate the effects that different approaches have on the images. The preservation of the tissue structure is often considered to be the most important feature, but it is not measured in all publications [80], [84], [110]. Tosta *et al.*in [15] reviewed the literature in terms of the evaluation techniques used in each work, but did not discuss the use of the different metrics. In this section, we introduce and discuss the most common metrics in the literature.

1) Quantitative metrics

Due to the expertise required to visually evaluate histological images and the fact that different pathologists may disagree on the quality of an image, the use of objective quantitative analysis is highly recommended.

Structure preservation: When the ground truth (GT) is available, metrics such as the Peak Signal to Noise Ratio (PSNR) or SSIM can be used to compare the results with the expected output. PSNR and SSIM are commonly used in BCD approaches [79]. Pathologists can easily identify the true stain colors in the image, making it possible to obtain a GT for the stain concentration [93]. Then the structure preservation of the BCD separation can be assessed. When the metric is calculated on the reconstructed separation (e.g. H-only and E-only images), the use of the Quaternion Structural Similarity (QSSIM) [28] is recommended to account for color similarities. Other approaches use the Euclidean distance between the GT and the BCD concentrations or the Normalized Mean Squared Error (NMSE). Note that the structure preservation is measured before using the CN techniques, and it is guaranteed by modifying only the color-related information. For those methods that do not rely on BCD and directly obtain CN, measuring the use of these metrics is only possible when the expected normalization result is available [89] (e.g. On the Mitos-Atypia dataset previously mentioned). Measuring PSNR, SSIM or other structural measures using original and CN images should be avoided [110], as better values will be obtained by not modifying the original image.

Color variation: When using BCD, it is possible to measure the difference between color vector matrices by using Euclidean distance or NMSE [44]. Color evaluation can be performed by comparing RGB or $l\alpha\beta$ median values [89] with the reference values or values from other images. This comparison does not consider different distributions of the stains in the images, so it is often performed on previously identified regions such as nuclei, cytoplasm or red blood cells [89]. The most popular metric is the Normalized Median Intensity (NMI) [28], [44], [99], where the median intensity value is divided by the 95th percentile. The NMI value is calculated for the entire dataset in which the color variance is measured, and then the standard deviation (SD) and the coefficient of variation (CV) (standard deviation divided by mean) are used as metrics. Lower NMI SD and NMI CV values indicate that the color distribution of the images are very similar.

CPATH system performance: The use of CPATH systems is the main reason for preprocessing techniques. There-
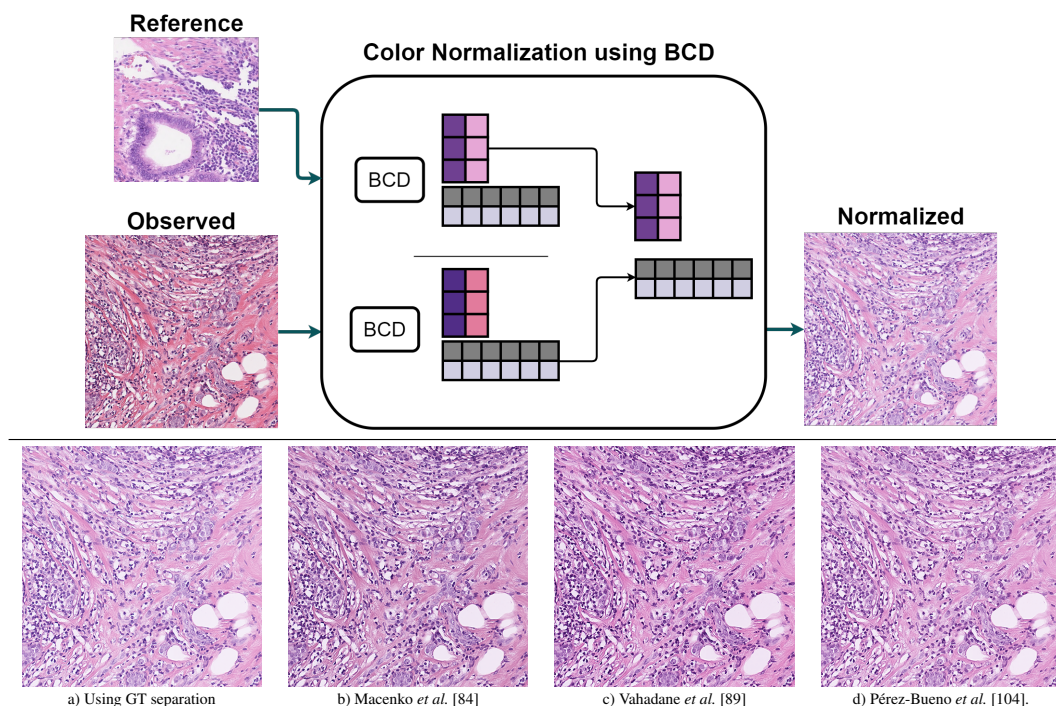
FIGURE 9. Top: Frequently used pipeline for CN after BCD (see Figure 7). First, the Reference and Observed images are deconvolved. Then, the color matrix (column vectors depicted with estimated colors for the image) is replaced by the reference and the stain concentration (row vectors depicted in gray) is preserved. The exact pipeline depends on each BCD method. **Bottom:** CN obtained with the same pipeline, using different BCD methods.

fore, it is important to see how preprocessing affects the final performance of the CPATH system [14] as compared with the original non-corrected images. Tumor segmentation [102], cell nuclei segmentation [44] or mitosis detection are common CPATH tasks. In some cases the performance of the CPATH system is tested and compared with systems previously designed by the authors [102]. Tellez *et al.* [14] presented an extensive evaluation on the effect preprocessing has on convolutional neural networks. The work in [28] assesses two different scenarios. In the first performance is tested using stain-specific features [40] on four classifiers. In the second scenario a VGG-19 is used and the performance is compared using the original images, CN images (RGB), and the OD concentrations. The use of RGB versus concentrations was also assessed in [87].

Computational complexity: Execution time and complexity alone will not determine the quality of a method. However, the massive size of WSI implies that it is important to consider the computational requirements of a method. For this reason, most authors include a time or complexity comparison with other methods.

2) Visual and qualitative analysis
CPATH systems are designed as a tool for pathologists. The human-machine collaboration scenario requires an input that is suitable for both human and machine. It therefore requires color-processed images to be to be visually evaluated together with the quantitative metrics. It is important to note the differences between the analyses performed by

pathologist and non-pathologist observers. As previously discussed, the WSI analysis requires considerable expertise. A visual analysis by pathologists is usually preferred but is often not included in the studies [15] since an expert pathologist might not be available. When included [43], [89], the pathologist's analysis is often reduced to evaluating the quality of small patches or ROI. Analysis performed by non-pathologist observers is often included, where a general assessment of image quality and color similarity can be made. Non-pathologists, however, cannot assess the diagnosis value of color-processed images. How CN or other preprocessing methods affect the pathologist diagnosis is still an open issue.

VI. IMAGE DATA AUGMENTATION
CPATH systems for cancer classification are normally based on data-driven DL models [9], [21]. Their performance can be considerably improved with image augmentation and it has recently gained more and more attention [119]. One augmentation technique, color augmentation (CA), was already mentioned in Section V in the context of color variation processing. The field of image augmentation, however, is much broader and includes other types of variations, and therefore we have devoted a whole section to it. The idea of image augmentation is to apply random transformations to the training images such that the model learns possible variations in the data, making CPATH systems more robust against unseen images. It can be used in addition to, or in some cases as an alternative to the previously presented preprocessing methods.

In other areas of DL, image augmentation already plays an important role [119] and it is to be expected that in the future it will become more and more important in the medical field. Apart from a better robustness to data variations, it can help avoid overfitting in small datasets [120] and to tackle class imbalance [121]. The augmented images can be seen as an artificial extension of the training images, such that the size of the dataset increases. It is important to note that some augmentation techniques for histopathological images are similar to those in other DL areas, while others are specific to the problem, such as the BCD-based methods that use stain separation for CA.

The image augmentation techniques can be divided into three categories: Transformations that aim to manipulate the morphology of the image, color augmentation and generative approaches, as described in the following subsections.

A. MORPHOLOGICAL AUGMENTATION

By morphological augmentation we mean all image transformations that aim to change the shape, structure or field of view of the input images. Typical basic augmentations include 90 degree rotations and vertical and horizontal mirroring. Further width and height shifts have been adapted [122], see Figure 10. To fill the 'free' areas of the patch, either a constant value can be used (black in our case) or the image content, mirrored at the boundary. Note that the augmentations with a padding of constant values can lead to unrealistic images, because black or white stripes appear at the image border. To circumvent this problem, 'random coordinate perturbation' was presented in [123], which means that the patches are extracted from the WSI with a random offset of the patch center. This can be seen as a width and height shift, and fills the 'free' space with the actual content of the neighbor patches.

These transformations can be extended by additive Gaussian noise and Gaussian blurring as proposed in [14]. In the case of Gaussian noise, a random value is added to the image RGB values that is drawn from a Gaussian distribution. Gaussian blurring describes the application of a Gaussian filter to the image and leads to fuzzy contours, see figure 10. In the case of elastic deformations and image scaling, the effect remains unclear in the existing literature: while Xiao *et al.* [124] do not recommend using these two techniques for image segmentation to reserve the original tissue features, Tellez *et al.* [14] successfully applied them to image classification.

B. COLOR AUGMENTATION

CA aims to systematically manipulate the color distribution of a given input image while preserving the structure. CA was previously introduced in Section V, as it can be used as an alternative to other color processing techniques [14]. Instead of standardizing the images, basic CA includes the random change of the brightness, contrast, hue or saturation of the image [14], [123], [124]. This can help to make the model invariant to different factors such as lightning conditions,

color intensity or other color variation introduced during acquisition (see Section II). Khan *et al.* [125] proposed to further modify, shuffle and shift the channels of the image in RGB or HSV space to obtain more color variations. CA can also be applied in addition to other techniques, such as CA after CN [86], [104]. More dedicated methods are based on BCD and try to mimic variations in the stains of the microscope images, see Figure 10. BCD-based augmentation strategies are tailored to histopathological images and have shown promising results [14], [104], [125]. Tellez *et al.* [126] proposed a method for H&E images consisting of three steps: First, BCD is performed to decompose the RGB image into one Hematoxylin and one Eosin channel. Then, the H&E channels are individually multiplied with a random value and finally the image is converted back to RGB color space. Different techniques can be used for BCD, as described in Section V-A, leading to different outcomes. The final performance is quite sensitive to a good deconvolution, as mentioned in [86], [104]. Xiao *et al.* [127] followed a different strategy: the images were transformed into the CIE-Lab color space [128]. For each channel, color transfer was applied with respect to a randomly chosen target patch: the mean of the channel distribution was shifted to the mean of the target patch. Faryna *et al.* [122] presented an approach that tailors the RandAugment strategy [129] to histopathological images by extending it (e.g. with BCD-based augmentation).

C. GENERATIVE APPROACHES

As a third category, generative approaches are outlined. Here, new image content is generated instead of just modifying the existing image. One technique of growing interest in this area is GANs, which are able to generate synthetic images that follow the same data distribution as the real images. In other DL areas, GAN-based approaches are becoming more and more popular for augmentation [119], however the literature of GAN-based augmentation for histopathological images remains scarce. Wei *et al.* [130] used cycleGAN [131] architecture and adapted it for data augmentation such that synthetic images of underrepresented classes were generated. The approach of Brock *et al.* [132] relies on the architecture of Biggan to generate artificial cancer tissue images [133]. An overview of the usage of GANs in the medical image domain, including different augmentation strategies can be found in [134]. Apart from GAN-based approaches, image generation can also be understood in a broader sense: In the work of [135], new images are generated by fusing the left half of one training image and the right half of another training image with pyramid pooling to avoid creating a sharp edge in the middle.

D. EVALUATION OF IMAGE AUGMENTATION TECHNIQUES

We conclude this section by describing the qualitative and quantitative approaches to evaluate image data augmentation. The goal of image data augmentation is to improve classification performance, however, if a qualitative evaluation is
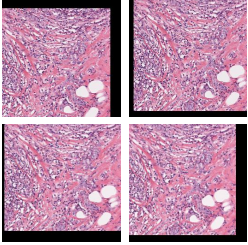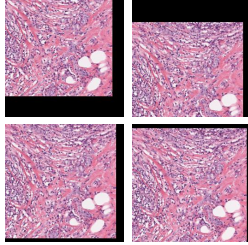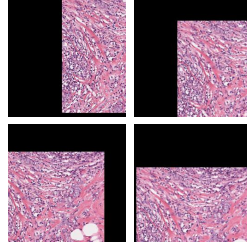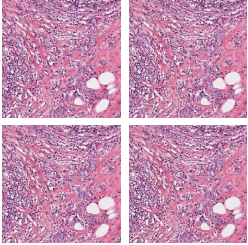
FIGURE 10. Examples of weak, intermediate and strong augmentation of selected transformations.

carried out by assessing examples of augmented images it is important to check if the transformations are too weak (no visual difference to the original images) or too strong (image content hardly be recognized). In Figure 10 we show a qualitative comparison of selected image augmentations for an example patch of a breast cancer classification task. For each transformation type, random samples of weak, intermediate and strong transformations are depicted. While weak transformations can be validated by non-experts, intermediate or strong image augmentation techniques require the assessment of pathologists to know if the augmented images are realistic or class-preserving. Quantitative evaluation usually requires a ground truth that is not available for image augmentation. Therefore, augmentation techniques are commonly evaluated by the final performance of the CPATH system, depending on the task, e.g. image classification, semantic segmentation or object detection. For augmentation, the evaluation is similar to the evaluation of color normalization previously described in Subsection V-C1 under 'CPATH system performance': a CPATH model is trained for each of the augmentation techniques. Then, the test performances of the models are compared to determine the best augmentation strategy [14], [104].

VII. CHALLENGES IN WSI PREPROCESSING AND FUTURE RESEARCH DIRECTIONS

In this work we have provided an extensive review of the WSI related preprocessing techniques, starting from data acquisition and dealing with the problems of artifacts detection, color variation, and image augmentation. The preprocessing required for WSI analysis is complex and often tissue-specific, disease-specific, and task-specific. The first challenge here is to identify and choose the right preprocessing pipeline. Due to the massive size of WSIs, the required steps and order should be chosen carefully to avoid redundancy. Handling, as discussed in Section III, is usually performed at the beginning, although some techniques using multi-resolution or segmentation might define their own procedure. It is not clear what order of preprocessing should be used. Artifacts detection is often needed, but the detection of specific artifacts depends on the tissue, biopsy, or even on the laboratory procedures. Whether to do it before color processing or not is also an open question. The presence of artifacts might hamper color processing, at the same time, previous color assessment could be used to improve the outcome of artifacts detection. When it comes to image augmentation, the discussion is similar. Are other techniques required if augmentation is used? In many cases, applying augmentation to clean, standardized images could lead to more controlled scenarios, however, considering artifacts and variation could create a wider range of plausible images during training. Therefore, a preprocessing pipeline should consider how different WSI preprocessing techniques interact with others, something which has not yet been explored in depth.

Each WSI preprocessing area discussed in this paper is at a different stage of development and has different challenges

to overcome. We will conclude this work with a discussion on the limitations and the challenges they present.

A. CHALLENGES IN ARTIFACTS DETECTION

Quality control evaluation has shown how artifacts detection and data curation affect the performance of the CPATH systems [38], [76]. Artifacts detection, however, is often ignored in the preprocessing pipeline. Current approaches often rely on low magnification analysis to discard complete WSIs. It is a well-known issue that QC approaches need to be extended to higher magnification in order to effectively detect some of the artifacts [48]. This, however, would require new computational efficient techniques to be implemented that can deal with the massive size of the WSIs. While some notable artifacts such as folded tissue, damaged tissue and blur are often mentioned as being critical, few works explore them separately. In many cases, artifacts detection is avoided by using automatic methods to separate non-informative tissue regions from ROIs. Focusing on ROIs may slightly speed up the preprocessing pipeline but might not remove all possible artifacts. Future research needs to address the presence of artifacts and measure how they affect the performance of CPATH systems.

Pathologists often consider blur to be the most critical defect in digital pathology. In addition, blur is a downside of the digitization process. It is not present when the slide is manually studied in the microscope, and as such, it is a new problem for pathologists. While blur is caused by a known focus problem and can be corrected using deblurring techniques, discarding blurred patches is often the most common approach. The structure fidelity of deblurring methods is a concern when working with medical images. This issue has been addressed in natural images and needs to be explored with histopathological images as well.

Most of the current methods used to detect specific artifacts (folded tissue, damaged tissue, and blood) rely on color differences and use color space transformation. Although most of these methods allow results to be influenced by color variation between images, this effect is usually not measured. Some of them include adaptive thresholding to deal with small variations, but it is not clear how inter-laboratory color changes will affect them. The range of possible artifacts is extremely broad, however, preprocessing techniques have only focused on finding and removing a few of them. Other than that, there is not enough literature available on other artifacts such as air bubbles, tissue tearing, contamination, etc. While pathologists need to be aware of the different types of artifacts, CPATH systems are far from being able to recognize one. Artifacts detection is not only a positive preprocessing step for CPATH, but one that is required in more complex, informative and interpretable systems. In this sense, research on new methods that are capable of identifying patches with different artifacts is needed.

Finally, artifacts might have an impact on other preprocessing areas such as color normalization and data augmentation.

How artifacts affect these areas has not been explored as of yet, and needs to be addressed in future research.

B. THE FUTURE OF COLOR PROCESSING

Despite the great impact color has on CPATH systems and the advances achieved in the color processing field, the latest color techniques are not often used in works concerning classification. As for the different approaches, CN is more popular than directly using the BCD stain separation, and CA is quickly gaining popularity. In the MICCAI 2018 conference Multi-Organ Nucleus Segmentation challenge [136], only half of the 32 teams used pathology-specific CN techniques. Vahadane *et al.* [89] and Macenko *et al.* [84] were the most popular. The other half used pixel intensity and RGB color transformations (pathology-unspecific). The work by Tellez *et al.* [14] also tested several CN approaches but only included the work of Macenko *et al.* [84] as BCD-based instead of more recent techniques. Neither the participants in the Multi-Organ challenge [136] nor Tellez *et al.* [14] reported the use of the deconvolved H&E channels. However, BCD can be found in classification studies [40], [137], usually using Ruifrok *et al.* [78] even when it is well known that it does not consider color variation. The potential performance boost of modern BCD [28], [79], [87] needs to be transferred to other classification approaches. In contrast, CA has quickly been adopted for color processing, as we will discuss later on this section.

Apart from a better transference to CPATH systems, there are several challenges that color processing needs to tackle. As they are closely related to the lack of research in finding artifacts, deviations from the desired staining schema are ignored in almost every color-related work. The $95^{th}/99^{th}$ percentile is used only in a few steps in the CN pipeline. As previously discussed, many artifacts can have a deep impact on the stain color-vector matrix estimation. This is usually avoided by identifying the ROIs [28], [93], [99] before tackling the color. If artifacts are not taken into account during BCD or CN, they could end up being confused with other histopathological features after standardization. For example, dust might be confused with cell nuclei or blood with Eosin.

Similarly, guaranteeing structure preservation is critical in color processing, and this is the main concern in BCD techniques. However, recent DL-based approaches [14] are more focused on classification performance. In [79] it was mentioned that both objectives can be conflicting. More research is needed on this interesting topic, which is also related to the interpretability of the systems for pathologists. Directly combining color processing with classification was carried out in [80], [87], but needs to be further explored in future work.

DL is of interest when dealing with color variation but several issues need to be addressed in future research. Its use for BCD needs to be explored, as well as the application of the stain separation using DL to CN, CA and classification. DL for CN often skips the BCD step and uses more complex and less interpretable latent intermediate spaces. It

is common in DL-based CN studies [80], [113] to see the lack of a reference image as an advantage, while training with images with a fixed staining protocol can be seen as using a reference laboratory instead. This often means that DL-based CN cannot deal with intra-laboratory color variations.

A fair comparison between the color processing methods in the literature is another challenge to be tackled. Review papers [15], [18] relate them in a theoretical way, but more works concerning quantification [14] are needed. A standardized protocol for measuring the quality of color processing methods has not been proposed so far.

Finally, the computational cost of color processing needs to be considerably improved due to the size of WSIs. The low computational cost of [78] is probably the reason why it is still commonly used, and can probably explain the popularity of pathology-unspecific techniques in [136]. Some approaches use unbiased pixel sampling [43], [84] to reduce the computational cost of finding the color matrix. The reduced time required by DL approaches once they are trained is one of these models' advantage, but their training cost is usually considerably high in terms of data, time and computational resources.

C. DATA AUGMENTATION POTENTIAL

Data augmentation is widely used in other areas of image classification, segmentation and object detection [119]. It is gaining popularity as a way of increasing the generalization capability of DL models in the medical field. While simple morphological transformations such as image rotation are widely used, more complex techniques still require further scientific analysis. One explanation could be that, for histopathological images, it is more difficult to tell which transformation preserves the class: For an image of a dog, for example, it is easy to see if the dog is still recognizable after a transformation. In the medical domain, this evaluation is more complex and often requires expert knowledge.

Further research in data augmentation could have huge potential, as shown in [14]. Here, more dedicated methods clearly outperform basic transformations in the final model classification performance: The application of color-based methods in addition to morphological transformations leads to a clear improvement and the best performing techniques include stain-based augmentations. An open question in this context is if the color transformations have to be realistic. While some approaches (e.g. [127]) aim to obtain augmented images with realistic color variations, others are so strong that the colors are clearly unnatural (see for example strong augmentations in [14]). To the best of the authors' knowledge, the risks of training with unrealistic augmentations in the medical domain has not yet been studied systematically.

Generative data augmentation for histopathological images is becoming more and more popular and will probably play an important role in the future, as it already does in other areas of DL [119]. GANs have the potential to realistically augment images or even generate completely new, artificial image data. This is especially interesting in the medical

domain, where labeled data is costly and access is limited due to privacy reasons.

Despite the huge potential of data augmentation for histopathological imaging, advanced augmentation techniques have not yet been widely applied in this area. Although some risks and limitations of data augmentation require further scientific studies, data augmentation promises to provide better generalizability of deep learning models and helps to overcome data shortage and class imbalance. Furthermore, stain-based augmentation methods are a powerful alternative to tackling color variations. Based on the current literature, it is not possible to draw a final conclusion on whether CA can replace CN methods. While in [14], [125], CA methods are stronger, [86] report that CN methods show the best results. Further research in this area is necessary to determine the best strategies for different use cases.

VIII. CONCLUSIONS

In this work we have described in depth the different steps of WSI-specific preprocessing and reviewed state-of-the-art techniques. The proper preprocessing of histopathological images is highly important because data-driven CPATH systems have shown promising results, yet are very sensitive to the data they are trained on. Depending on the existing data and task at hand, each preprocessing step must be carefully selected and evaluated. For this purpose, we provided an overview that helps researchers and practitioners obtain the best possible results. Starting with the WSI acquisition procedure, we have further explained the problems and existing methods for histological image preprocessing. First, approaches for dealing with the massive size of the WSIs and splitting them in patches were presented. Then, we explored how to perform artifacts detection to remove undesired structures or artificial structures from the images, such as blur, fold, blood or damaged areas. We reviewed the approaches for dealing with color variation between different centers: color deconvolution, color normalization and color augmentation, and the metrics to evaluate the color changes in the image. In addition, the latest data augmentation techniques applied to WSIs were presented, covering morphological transformation, color augmentation and generative approaches.

Finally, we discussed the challenges and future research directions for WSI preprocessing and the potential of DL techniques in this field. The size of the gigapixel WSIs, the amount of artifacts and possible variations in the images, and task-specific problem characteristics for different types of cancer are all challenges that show the complexity of WSI preprocessing and the need for specialized methods. The huge impact that preprocessing has on the development of accurate and reliable CPATH systems makes on thing clear: In automatic diagnosis systems, the devil is in the details. We cannot ignore them if we want to build sound, reliable, robust and widely used CPATH systems.

REFERENCES

[1] R. Ravi, "Ai gone wrong: 5 biggest ai failures of all time," https://www. jumpstartmag.com/ai-gone-wrong-5-biggest-ai-failures-of-all-time/, 2021 (accessed: January 2022).

[2] P. Dialani, "Famous ai ai gone wrong examples in the real world we need to know," https://www.analyticsinsight.net/ famous-ai-gone-wrong-examples-in-the-real-world-we-need-to-know/, 2021 (accessed: January 2022).

[3] T. S. Perry, "Andrew ng x-rays the ai hype," IEEE Spectrum, https: //spectrum.ieee.org/andrew-ng-xrays-the-ai-hype, May, 2021 (accessed: January 2022).

[4] Quartz, "Faulty image: Ai has a long way to go before doctors can trust it with your life," https://qz.com/2016153/ ai-promised-to-revolutionize-radiology-but-so-far-its-failing/, June, 2021 (accessed: January 2022).

[5] E. Mueller, "10 common machine learning mistakes and how to avoid them," https://www.capitalone.com/tech/machine-learning/ 10-common-machine-learning-mistakes/, 2021 (accessed: January 2022).

[6] S. W. Page, "David spiegelhalter on tempering the hype around ai," Hello World, https://helloworld.raspberrypi.org/articles/ hw-12-tempering-the-hype-around-ai, 2020 (accessed: January 2022).

[7] D. Spiegelhalter, *The Art of Statistics. Learning from Data.* Pellican, 2020.

[8] N. C. Institute, "The burden of cancer worldwide," https://www.cancer.gov/about-cancer/understanding/statistics, (accessed: August 19, 2021).

[9] S. Morales, K. Engan, and V. Naranjo, "Artificial intelligence in computational pathology – challenges and future directions," *Digital Signal Processing*, vol. 119, p. 103196, 2021.

[10] E. Abels, L. Pantanowitz, F. Aeffner, M. D. Zarella, J. Van Der Laak, M. M. Bui, V. N. Vemuri, A. V. Parwani, J. Gibbs, E. Agosto-Arroyo *et al.*, "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, 2019.

[11] A. Schmidt, J. Silva-Rodríguez, R. Molina, and V. Naranjo, "Efficient cancer classification by coupling semi supervised and multiple instance learning," *IEEE Access*, vol. 10, pp. 9763–9773, 2022.

[12] S. M. McKinney, M. Sieniek, V. Godbole, J. Godwin, N. Antropova, H. Ashrafian, T. Back, M. Chesus, G. S. Corrado, A. Darzi, M. Etemadi *et al.*, "International evaluation of an AI system for breast cancer screening," *Nature*, vol. 577, no. 7788, pp. 89–94, Jan. 2020.

[13] D. analytics, "Digital science and research solutions inc," https://app.dimensions.ai/analytics/publication/overview/timeline? search_mode=content&search_text=(%22Computer%20Aided% 20Diagnosis%22%20OR%20CAD%20system%20AND%20% 22Whole%20Slide%20Image%22)%20OR%20%22Computational% 20Pathology%22%20OR%20%20Digital%20Pathology%22&search_ type=kws&search_field=full_search&year_from=2016&year_to=2021, query: ("Computer Aided Diagnosis" OR CAD system AND "Whole Slide Image") OR "Computational Pathology" OR "Digital Pathology" (accessed: January 2022).

[14] D. Tellez, G. Litjens, P. Bándi, W. Bulten, J.-M. Bokhorst, F. Ciompi, and J. 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, p. 101544, 2019.

[15] T. A. A. Tosta, P. R. de Faria, L. A. Neves, and M. Z. do Nascimento, "Computational normalization of H&E-stained histological images: Progress, challenges and future potential," *Artificial Intelligence in Medicine*, vol. 95, pp. 118 – 132, 2019.

[16] S. Ramírez-Gallego, B. Krawczyk, S. García, M. Woźniak, and F. Herrera, "A survey on data preprocessing for data stream mining: Current status and future directions," *Neurocomputing*, vol. 239, pp. 39–57, May 2017.

[17] A. I. Wright, C. M. Dunn, M. Hale, G. G. Hutchins, and D. 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, 2020.

[18] M. Salvi, U. R. Acharya, F. Molinari, and K. M. Meiburger, "The impact of pre- and post-image processing techniques on deep learning frameworks: A comprehensive review for digital pathology image analysis," p. 104129, 2021.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI
10.1109/ACCESS.2022.3176091, IEEE Access

Kanwal *et al.*:The devil is in the details: A review

[19] S. Roy, A. kumar Jain, S. Lal, and J. Kini, "A study about color normalization methods for histopathology images," *Micron*, vol. 114, pp. 42–61, 2018.

[20] B. Smith, M. Hermsen, E. Lesser, D. Ravichandar, and W. Kremers, "Developing image analysis pipelines of whole-slide images: Pre-and post-processing," *Journal of Clinical and Translational Science*, vol. 5, no. 1, 2021.

[21] C. L. Srinidhi, O. Ciga, and A. L. Martel, "Deep neural network models for computational histopathology: A survey," *Medical Image Analysis*, p. 101813, 2020.

[22] N. Dimitriou, O. Arandjelović, and P. D. Caie, "Deep Learning for Whole Slide Image Analysis: An Overview," *Frontiers in Medicine*, vol. 6, no. November, pp. 1–7, 2019.

[23] S. Saxena and M. Gyanchandani, "Machine learning methods for computer-aided breast cancer diagnosis using histopathology: a narrative review," *Journal of medical imaging and radiation sciences*, vol. 51, no. 1, pp. 182–193, 2020.

[24] D. Komura and S. Ishikawa, "Machine learning methods for histopathological image analysis," *Computational and structural biotechnology journal*, vol. 16, pp. 34–42, 2018.

[25] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener, "Histopathological image analysis: A review," *IEEE reviews in biomedical engineering*, vol. 2, pp. 147–171, 2009.

[26] T. A. Azevedo Tosta, L. A. Neves, and M. Z. do Nascimento, "Segmentation methods of h&e-stained histological images of lymphoma: A review," *Informatics in Medicine Unlocked*, pp. 35–43, 2017.

[27] S. Huang, J. Yang, S. Fong, and Q. Zhao, "Artificial intelligence in cancer diagnosis and prognosis: Opportunities and challenges," *Cancer letters*, vol. 471, pp. 61–71, 2020.

[28] F. Pérez-Bueno, M. Vega, M. Sales, J. Aneiros-Fernández, V. Naranjo, R. Molina, and A. Katsaggelos, "Blind color deconvolution, normalization and classification of histological images using general super gaussian priors and bayesian inference," *Computer Methods and Programs in Biomedicine*, October 2021.

[29] V. Rastogi, N. Puri, S. Arora, G. Kaur, L. Yadav, and R. Sharma, "Artefacts: a diagnostic dilemma–a review," *Journal of clinical and diagnostic research: JCDR*, vol. 7, no. 10, p. 2408, 2013.

[30] H. O. Lyon, A. P. De Leenheer, R. W. Horobin, W. E. Lambert, E. K. Schulte, B. Van Liedekerke, and D. H. Wittekind, "Standardization of reagents and methods used in cytological and histological practice with emphasis on dyes, stains and chromogenic reagents," *The Histochemical Journal*, vol. 26, no. 7, pp. 533–544, 1994.

[31] A. Woods and R. Ellis, *Laboratory Histopathology: A Complete Reference*. Churchill Livingstone, 1994, no. v. 1.

[32] I. B. Dimenstein, "Grossing biopsies: an introduction to general principles and techniques," *Annals of Diagnostic Pathology*, vol. 13, no. 2, pp. 106–113, 2009.

[33] P. R. Bindhu, R. Krishnapillai, P. Thomas, and P. Jayanthi, "Facts in artifacts," *Journal of Oral and Maxillofacial Pathology*, vol. 17, no. 3, pp. 397–401, 2013.

[34] G. O. Rolls, N. J. Farmer, and J. B. Hall, *Artifacts in histological and cytological preparations*. Leica Microsystems., 2008.

[35] M. Srinivasan, D. Sedmak, and S. Jewell, "Effect of fixatives and tissue processing on the content and integrity of nucleic acids," *The American journal of pathology*, vol. 161, no. 6, pp. 1961–1971, 2002.

[36] S. A. Taqi, S. A. Sami, L. B. Sami, and S. A. Zaki, "A review of artifacts in histopathology," *Journal of oral and maxillofacial pathology: JOMFP*, vol. 22, no. 2, p. 279, 2018.

[37] S. K. Suvarna, C. Layton, and J. D. Bancroft, *Bancroft's Theory and Practice of Histological Techniques*, 8th ed., S. K. Suvarna, C. Layton, and J. D. Bancroft, Eds. Elsevier, 2019.

[38] R. Wetteland, K. Engan, T. Eftestøl, V. Kvikstad, and E. A. Janssen, "A multiscale approach for whole-slide image segmentation of five tissue classes in urothelial carcinoma slides," *Technology in Cancer Research & Treatment*, vol. 19, p. 1533033820946787, 2020.

[39] M. Khened, A. Kori, H. Rajkumar, G. Krishnamurthi, and B. Srinivasan, "A generalized deep learning framework for whole-slide image segmentation and analysis," *Scientific reports*, vol. 11, no. 1, pp. 1–14, 2021.

[40] A. E. Esteban, M. Lopez-Perez, A. Colomer, M. A. Sales, R. Molina, and V. Naranjo, "A new optical density granulometry-based descriptor for the classification of prostate histological images using shallow and deep Gaussian processes," *Computer Methods and Programs in Biomedicine*, vol. 178, pp. 303–317, 2019.

[41] R. Wetteland, K. Engan, and T. Eftesol, "Parameterized extraction of tiles in multilevel gigapixel images," in *2021 12th International Symposium on Image and Signal Processing and Analysis (ISPA)*, 2021, pp. 78–83.

[42] R. Wetteland, V. Kvikstad, T. Eftestøl, E. Tøssebro, M. Lillesand, E. A. Janssen, and K. Engan, "Automatic diagnostic tool for predicting cancer grade in bladder cancer patients using deep learning," *IEEE Access*, vol. 9, pp. 115 813–115 825, 2021.

[43] Y. Zheng, Z. Jiang, F. Xie, H. Zhang, Y. Ma, H. Shi, and Y. Zhao, "Feature extraction from histopathological images based on nucleus-guided convolutional neural network for breast lesion classification," *Pattern Recognition*, vol. 71, pp. 14–25, 2017.

[44] B. E. Bejnordi, M. Veta, P. J. Van Diest, B. Van Ginneken, N. Karssemeijer, G. Litjens, J. A. Van Der Laak, M. Hermsen, Q. F. Manson, M. Balkenhol *et al.*, "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," *Jama*, vol. 318, no. 22, pp. 2199–2210, 2017.

[45] Y. Chen, J. Zee, A. Smith, C. Jayapandian, J. Hodgin, D. Howell, M. Palmer, D. Thomas, C. Cassol, A. B. Farris III *et al.*, "Assessment of a computerized quantitative quality control tool for whole slide images of kidney biopsies," *The Journal of Pathology*, vol. 253, no. 3, pp. 268–278, 2021.

[46] X. Zhou, R. Molina, Y. Ma, T. Wang, and D. Ni, "Parameter-free gaussian psf model for extended depth of field in brightfield microscopy," *IEEE Transactions on Image Processing*, vol. 29, pp. 3227–3238, 2019.

[47] C. Senaras, M. K. K. Niazi, G. Lozanski, and M. N. Gurcan, "Deepfocus: detection of out-of-focus regions in whole slide digital images using deep learning," *PloS one*, vol. 13, no. 10, p. e0205387, 2018.

[48] H. M. Shakhawat, T. Nakamura, F. Kimura, Y. Yagi, and M. 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.

[49] P. 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, 2016.

[50] H. Wu, J. H. Phan, A. K. Bhatia, C. A. Cundiff, B. M. Shehata, and M. D. Wang, "Detection of blur artifacts in histopathological whole-slide images of endomyocardial biopsies," in *2015 37th annual international Conference of the IEEE Engineering in Medicine and biology society (EMBC)*. IEEE, 2015, pp. 727–730.

[51] N. Hashimoto, P. A. Bautista, M. Yamaguchi, N. Ohyama, and Y. Yagi, "Referenceless image quality evaluation for whole slide imaging," *Journal of pathology informatics*, vol. 3, 2012.

[52] D. Gao, D. Padfield, J. Rittscher, and R. McKay, "Automated training data generation for microscopy focus classification," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2010, pp. 446–453.

[53] T. Albuquerque, A. Moreira, and J. S. Cardoso, "Deep ordinal focus assessment for whole slide images," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 657–663.

[54] G. Campanella, A. R. Rajanna, L. Corsale, P. J. Schüffler, Y. Yagi, and T. 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.

[55] T. Kohlberger, Y. Liu, M. Moran, P.-H. C. Chen, T. Brown, J. D. Hipp, C. H. Mermel, and M. C. Stumpe, "Whole-slide image focus quality: Automatic assessment and impact on ai cancer detection," *Journal of pathology informatics*, vol. 10, 2019.

[56] Z. Wang, M. S. Hosseini, A. Miles, K. N. Plataniotis, and Z. Wang, "Focuslitenn: High efficiency focus quality assessment for digital pathology," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020, pp. 403–413.

[57] H. Zhao, Z. Ke, N. Chen, S. Wang, K. Li, L. Wang, X. Gong, W. Zheng, L. Song, Z. Liu *et al.*, "A new deep learning method for image deblurring in optical microscopic systems," *Journal of biophotonics*, vol. 13, no. 3, p. e201960147, 2020.

[58] L. Mukherjee, H. D. Bui, A. Keikhosravi, A. Loeffler, and K. W. Eliceiri, "Super-resolution recurrent convolutional neural networks for learning with multi-resolution whole slide images," *Journal of biomedical optics*, vol. 24, no. 12, p. 126003, 2019.

[59] Z. Chen, X. Guo, P. Y. Woo, and Y. Yuan, "Super-resolution enhanced medical image diagnosis with sample affinity interaction," *IEEE Transactions on Medical Imaging*, vol. 40, no. 5, pp. 1377–1389, 2021.

[60] A. Singh, A. Chandra, R. Kumar, K. Singh, and N. Dey, "Dark channel processing for medical image enhancement," in *2019 IEEE International*

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3176091, IEEE Access

Kanwal *et al.*: The devil is in the details: A review.

IEEE Access

WIE Conference on Electrical and Computer Engineering (WIECON-ECE). IEEE, 2019, pp. 1–6.

[61] S. Palokangas, J. Selinummi, and O. Yli-Harja, "Segmentation of folds in tissue section images," in *2007 29th annual international Conference of the IEEE Engineering in Medicine and biology society.* IEEE, 2007, pp. 5641–5644.

[62] P. A. Bautista and Y. Yagi, "Detection of tissue folds in whole slide images," *Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society: Engineering the Future of Biomedicine, EMBC 2009*, pp. 3669–3672, 2009.

[63] S. Kothari, J. H. Phan, T. H. Stokes, and M. D. Wang, "Pathology imaging informatics for quantitative analysis of whole-slide images," *Journal of the American Medical Informatics Association*, vol. 20, no. 6, pp. 1099–1108, 08 2013.

[64] M. Wang, S. Kothari, and J. Phan, "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, p. 22, 2013.

[65] M. Babaie and H. R. Tizhoosh, "Deep features for tissue-fold detection in histopathology images," in *European Congress on Digital Pathology.* Springer, 2019, pp. 125–132.

[66] R. Wetteland, K. Engan, T. Eftestøl, V. Kvikstad, and E. A. Janssen, "Multiclass tissue classification of whole-slide histological images using convolutional neural networks," *ICPRAM 2019 - Proceedings of the 8th International Conference on Pattern Recognition Applications and Methods*, pp. 320–327, 2019.

[67] E. Mercan, S. Aksoy, L. G. Shapiro, D. L. Weaver, T. Brunye, and J. G. Elmore, "Localization of diagnostically relevant regions of interest in whole slide images," in *2014 22nd International Conference on Pattern Recognition.* IEEE, 2014, pp. 1179–1184.

[68] C. Bahlmann, A. Patel, J. Johnson, J. Ni, A. Chekkoury, P. Khurd, A. Kamen, L. Grady, E. Krupinski, A. Graham *et al.*, "Automated detection of diagnostically relevant regions in h&e stained digital pathology slides," in *Medical Imaging 2012: Computer-Aided Diagnosis*, vol. 8315, 2012, p. 831504.

[69] A. C. Ruifrok and D. A. Johnston, "Quantification of histochemical staining by color deconvolution," *Analytical and Quantitative Cytology and Histology*, vol. 23, no. 4, pp. 291–299, 2001.

[70] O. N. Dalheim, R. Wetteland, V. Kvikstad, E. A. M. Janssen, and K. Engan, "Semi-supervised tissue segmentation of histological images," in *Colour and visual computing symposium/CEUR Workshop Proceedings*, 2020.

[71] Z. Swiderska-Chadaj, T. Markiewicz, J. Gallego, G. Bueno, B. Grala, and M. Lorent, "Deep learning for damaged tissue detection and segmentation in ki-67 brain tumor specimens based on the u-net model," *Bulletin of the Polish Academy of Sciences: Technical Sciences*, pp. 849–856, 2018.

[72] Z. Swiderska-Chadaj, T. Markiewicz, S. Cierniak, and R. Koktysz, "Automatic quantification of vessels in hemorrhoids whole slide images," in *2016 17th International Conference Computational Problems of Electrical Engineering (CPEE).* IEEE, 2016, pp. 1–4.

[73] D. Clymer, S. Kostadinov, J. Catov, L. Skvarca, L. Pantanowitz, J. Cagan, and P. LeDuc, "Decidual vasculopathy identification in whole slide images using multiresolution hierarchical convolutional neural networks," *The American Journal of Pathology*, vol. 190, no. 10, pp. 2111–2122, 2020.

[74] D. Ameisen, C. Deroulers, V. Perrier, F. Bouhidel, M. Battistella, L. Legrès, A. Janin, P. Bertheau, and J. B. Yunès, "Towards better digital pathology workflows: Programming libraries for high-speed sharpness assessment of Whole Slide Images," *Diagnostic Pathology*, vol. 9, no. 1, pp. 1–7, 2014.

[75] A. R. N. Avanaki, K. S. Espig, A. Xthona, C. Lanciault, and T. R. L. Kimpe, "Automatic image quality assessment for digital pathology," in *Breast Imaging*, A. Tingberg, K. Lång, and P. Timberg, Eds. Cham: Springer International Publishing, 2016, pp. 431–438.

[76] A. Janowczyk, R. Zuo, H. Gilmore, M. Feldman, and A. Madabhushi, "Histoqc: an open-source quality control tool for digital pathology slides," *JCO clinical cancer informatics*, vol. 3, pp. 1–7, 2019.

[77] A. Jiménez, G. Bueno, G. Cristóbal, O. Déniz, D. Toomey, and C. Conway, "Image quality metrics applied to digital pathology," in *Optics, Photonics and Digital Technologies for Imaging Applications IV*, vol. 9896. International Society for Optics and Photonics, 2016, p. 98960S.

[78] A. C. Ruifrok and D. A. Johnston, "Quantification of histochemical staining by color deconvolution," *Analytical and quantitative cytology and histology*, vol. 23, pp. 291–299, 2001.

[79] F. Pérez-Bueno, M. López-Pérez, M. Vega, J. Mateos, V. Naranjo, R. Molina, and A. K. Katsaggelos, "A TV-based image processing framework for blind color deconvolution and classification of histological images," *Digital Signal Processing*, vol. 101, p. 102727, 2020.

[80] A. Bentaieb and G. Hamarneh, "Adversarial stain transfer for histopathology image analysis," *IEEE Transactions on Medical Imaging*, vol. 37, no. 3, pp. 792–802, 3 2018.

[81] E. Reinhard, M. Ashikhmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Computer Graphics and Applications*, vol. 21, no. 5, pp. 34–41, 2001.

[82] F. G. Zanjani, S. Zinger, B. E. Bejnordi, J. A. van der Laak, and P. H. de With, "Stain normalization of histopathology images using generative adversarial networks," in *2018 IEEE 15th International symposium on biomedical imaging (ISBI 2018).* IEEE, 2018, pp. 573–577.

[83] A. Rabinovich, S. Agarwal, C. Laris, J. H. Price, and S. J. Belongie, "Unsupervised color decomposition of histologically stained tissue samples," in *Advances in Neural Information Processing Systems*, 2004, pp. 667–674.

[84] M. Macenko, M. Niethammer *et al.*, "A method for normalizing histology slides for quantitative analysis," in *International Symposium on Biomedical Imaging (ISBI)*, 2009, pp. 1107–1110.

[85] F. Perez-Bueno, M. Vega, V. Naranjo, R. Molina, and A. K. Katsaggelos, "Super Gaussian Priors for Blind Color Deconvolution of Histological Images," *Proceedings - International Conference on Image Processing, ICIP*, vol. 2020-Octob, pp. 3010–3014, 2020.

[86] Y. Zheng, Z. Jiang, H. Zhang, F. Xie, D. Hu, S. Sun, J. Shi, and C. Xue, "Stain standardization capsule for application-driven histopathological image normalization," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 2, pp. 337–347, 2020.

[87] R. Duggal, A. Gupta, R. Gupta, and P. Mallick, "Sd-layer: stain deconvolutional layer for cnns in medical microscopic imaging," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2017, pp. 435–443.

[88] J. Xu, L. Xiang, G. Wang, S. Ganesan, M. Feldman, N. N. Shih, H. Gilmore, and A. Madabhushi, "Sparse non-negative matrix factorization (SNMF) based color unmixing for breast histopathological image analysis," *Computerized Medical Imaging and Graphics*, vol. 46, pp. 20–29, 2015.

[89] A. Vahadane, T. Peng, A. Sethi, S. Albarqouni, L. Wang, M. Baust, K. Steiger, A. M. Schlitter, I. Esposito, and N. Navab, "Structure-preserving color normalization and sparse stain separation for histological images," *IEEE Transactions on Medical Imaging*, vol. 35, pp. 1962–1971, 2016.

[90] T. A. Tosta, P. R. de Faria, J. P. S. Servato, L. A. Neves, G. F. Roberto, A. S. Martins, and M. Z. do Nascimento, "Unsupervised method for normalization of hematoxylin-eosin stain in histological images," *Computerized Medical Imaging and Graphics*, vol. 77, p. 101646, 2019.

[91] N. Trahearn, D. Snead, I. Cree, and N. Rajpoot, "Multi-class stain separation using independent component analysis," in *Medical Imaging 2015: Digital Pathology*, 2015, p. 94200J.

[92] N. Alsubaie, S. E. A. Raza, and N. Rajpoot, "Stain deconvolution of histology images via independent component analysis in the wavelet domain," in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, 2016, pp. 803–806.

[93] N. Alsubaie, N. Trahearn, S. E. Raza, D. Snead, and N. M. Rajpoot, "Stain deconvolution using statistical analysis of multi-resolution stain colour representation," *PLoS ONE*, vol. 12, no. 1, pp. 1–15, 2017.

[94] M. T. McCann, J. Majumdar *et al.*, "Algorithm and benchmark dataset for stain separation in histology images," in *International Conference on Image Processing (ICIP)*, 2014, pp. 3953–3957.

[95] L. Astola, "Stain separation in digital bright field histopathology," in *2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, 2016, pp. 1–6.

[96] M. Niethammer, D. Borland, J. S. Marron, J. Woosley, and N. E. Thomas, "Appearance Normalization of Histology Slides," in *Machine Learning in Medical Imaging.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 58–66.

[97] M. Gavrilovic, J. C. Azar *et al.*, "Blind color decomposition of histological images," *IEEE Transactions on Medical Imaging*, vol. 32, pp. 983–994, 2013.

[98] J. Vicory, H. D. Couture, N. E. Thomas, D. Borland, J. Marron, J. Woosley, and M. Niethammer, "Appearance normalization of histology slides," *Computerized Medical Imaging and Graphics*, vol. 43, pp. 89–98, 2015.

[99] Y. Zheng, Z. Jiang, H. Zhang, F. Xie, J. Shi, and C. Xue, "Adaptive color deconvolution for histological WSI normalization," *Computer Methods and Programs in Biomedicine*, vol. 170, pp. 107–120, 2019.

[100] M. Salvi, N. Michielli, and F. Molinari, "Stain color adaptive normalization (scan) algorithm: Separation and standardization of histological stains in digital pathology," *Computer Methods and Programs in Biomedicine*, vol. 193, p. 105506, 2020.

[101] D. Magee, D. Treanor, D. Crellin, M. Shires, K. Smith, K. Mohee, and P. Quirke, "Colour normalisation in digital histopathology images," *Proc Optical Tissue Image analysis in Microscopy, Histopathology and Endoscopy*, pp. 100–111, 2009.

[102] A. M. Khan, N. Rajpoot, D. Treanor, and D. Magee, "A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution," *IEEE Transactions on Biomedical Eng.*, vol. 61, no. 6, pp. 1729–1738, 2014.

[103] N. Hidalgo-Gavira, J. Mateos, M. Vega, R. Molina, and A. K. Katsaggelos, "Variational Bayesian blind color deconvolution of histopathological images," *IEEE Transactions on Image Processing*, vol. 29, no. 1, pp. 2026–2036, 2020.

[104] F. Pérez-Bueno, J. Serra, M. Vega, J. Mateos, R. Molina, and A. K. Katsaggelos, "Bayesian k-svd for h&e blind color deconvolution. applications to stain normalization, data augmentation, and cancer classification." *Computerized Medical Imaging and Graphics*, 2022.

[105] D. Onder, S. Zengin, and S. Sarioglu, "A Review on Color Normalization and Color Deconvolution Methods in Histopathology," *Applied Immunohistochemistry & Molecular Morphology*, vol. 22, no. 10, pp. 713–719, Dec. 2014.

[106] T. A. A. Tosta, P. R. de Faria, L. A. Neves, and M. Z. do Nascimento, "Color normalization of faded h&e-stained histological images using spectral matching," *Computers in Biology and Medicine*, vol. 111, p. 103344, 2019.

[107] R. A. Hoffman, S. Kothari, and M. D. Wang, "Comparison of normalization algorithms for cross-batch color segmentation of histopathological images," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2014, pp. 194–197.

[108] M. Z. Hoque, A. Keskinarkaus, P. Nyberg, and T. Seppänen, "Retinex model based stain normalization technique for whole slide image analysis," *Computerized Medical Imaging and Graphics*, vol. 90, p. 101901, 2021.

[109] A. Janowczyk, A. Basavanhally, and A. Madabhushi, "Stain normalization using sparse autoencoders (stanosa): Application to digital pathology," *Computerized Medical Imaging and Graphics*, vol. 57, pp. 50–61, 2017.

[110] M. T. Shaban, C. Baur, N. Navab, and S. Albarqouni, "Staingan: Stain style transfer for digital histological images," in *2019 Ieee 16th international symposium on biomedical imaging (Isbi 2019)*. IEEE, 2019, pp. 953–956.

[111] X. Chen, J. Yu, S. Cheng, X. Geng, S. Liu, W. Han, J. Hu, L. Chen, X. Liu, and S. Zeng, "An unsupervised style normalization method for cytopathology images," *Computational and Structural Biotechnology Journal*, vol. 19, pp. 3852–3863, 2021.

[112] P. Salehi and A. 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.

[113] M. Runz, D. Rusche, S. Schmidt, M. R. Weihrauch, J. Hesser, and C.-A. Weis, "Normalization of HE-stained histological images using cycle consistent generative adversarial networks," *Diagnostic Pathology*, vol. 16, no. 1, p. 71, Aug. 2021.

[114] N. Zhou, D. Cai, X. Han, and J. Yao, "Enhanced cycle-consistent generative adversarial network for color normalization of h&e stained images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 694–702.

[115] J. Lan, S. Cai, Y. Xue, Q. Gao, M. Du, H. Zhang, Z. Wu, Y. Deng, Y. Huang, T. Tong, and G. Chen, "Unpaired stain style transfer using invertible neural networks based on channel attention and long-range residual," *IEEE Access*, vol. 9, pp. 11 282–11 295, 2021.

[116] A. Patil, M. Talha, A. Bhatia, N. C. Kurian, S. Mangale, S. Patel, and A. Sethi, "Fast, self supervised, fully convolutional color normalization of h amp;e stained images," in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, 2021, pp. 1563–1567.

[117] A. Z. Moghadam, H. Azarnoush, S. A. Seyyedsalehi, and M. Havaei, "Stain transfer using generative adversarial networks and disentangled features," *Computers in Biology and Medicine*, vol. 142, p. 105219, 2022.

[118] J. Ke, Y. Shen, and Y. Lu, "Style normalization in histology with federated learning," in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, 2021, pp. 953–956.

[119] C. Shorten and T. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, pp. 1–48, 2019.

[120] A. Mikołajczyk and M. Grochowski, "Data augmentation for improving deep learning in image classification problem," in *2018 International Interdisciplinary PhD Workshop (IIPhDW)*, 2018, pp. 117–122.

[121] M. Sebai, X. Wang, and T. Wang, "MaskMitosis: a deep learning framework for fully supervised, weakly supervised, and unsupervised mitosis detection in histopathology images," *Medical & Biological Engineering & Computing*, vol. 58, no. 7, pp. 1603–1623, 2020.

[122] K. Faryna, J. van der Laak, and G. Litjens, "Tailoring automated data augmentation to h&e-stained histopathology," in *Medical Imaging with Deep Learning*, 2021.

[123] M. Khened, A. Kori, H. Rajkumar, G. Krishnamurthi, and B. Srinivasan, "A generalized deep learning framework for whole-slide image segmentation and analysis," *Scientific Reports*, vol. 11, no. 1, p. 11579, 2021.

[124] Q.-E. Xiao, P.-C. Chung, H.-W. Tsai, K.-S. Cheng, N.-H. Chow, Y.-Z. Juang, H.-H. Tsai, C.-H. Wang, and T.-A. Hsieh, "Hematoxylin and eosin (h&e) stained liver portal area segmentation using multi-scale receptive field convolutional neural network," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 9, no. 4, pp. 623–634, 2019.

[125] A. Khan, M. Atzori, S. Otálora, V. Andrearczyk, and H. Müller, "Generalizing convolution neural networks on stain color heterogeneous data for computational pathology," in *Medical Imaging 2020: Digital Pathology*, vol. 11320. International Society for Optics and Photonics, 2020, p. 113200R.

[126] D. Tellez, M. Balkenhol, I. Otte-Höller, R. van de Loo, R. Vogels, P. Bult, C. Wauters, W. Vreuls, S. Mol, N. Karssemeijer *et al.*, "Whole-slide mitosis detection in h&e breast histology using phh3 as a reference to train distilled stain-invariant convolutional networks," *IEEE transactions on medical imaging*, vol. 37, pp. 2126–2136, 2018.

[127] Y. Xiao, E. Decencière, S. Velasco-Forero, H. Burdin, T. Bornschlögl, F. Bernerd, E. Warrick, and T. Baldeweck, "A new color augmentation method for deep learning segmentation of histological images," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE, 2019, pp. 886–890.

[128] F. W. Billmeyer, "Color science: Concepts and methods, quantitative data and formulae, 2nd ed., by gunter wyszecki and w. s. stiles, john wiley and sons, new york, 1982, 950 pp." *Color Research & Application*, vol. 8, no. 4, pp. 262–263, 1983.

[129] E. D. Cubuk, B. Zoph, J. Shlens, and Q. V. Le, "Randaugment: Practical automated data augmentation with a reduced search space," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 702–703.

[130] J. Wei, A. Suriawinata, L. Vaickus, B. Ren, X. Liu, J. Wei, and S. Hassanpour, "Generative image translation for data augmentation in colorectal histopathology images," *Proceedings of machine learning research*, vol. 116, p. 10, 2019.

[131] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[132] A. Brock, J. Donahue, and K. Simonyan, "Large scale GAN training for high fidelity natural image synthesis," in *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019.

[133] A. C. Quiros, R. Murray-Smith, and K. Yuan, "Pathology {GAN}: Learning deep representations of cancer tissue," in *Medical Imaging with Deep Learning*, 2019.

[134] X. Yi, E. Walia, and P. Babyn, "Generative adversarial network in medical imaging: A review," *Medical Image Analysis*, vol. 58, p. 101552, 2019.

[135] S. T. M. Ataky, J. de Matos, A. d. S. Britto, L. E. Oliveira, and A. L. Koerich, "Data augmentation for histopathological images based on gaussian-laplacian pyramid blending," in *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2020, pp. 1–8.

[136] N. Kumar, R. Verma, D. Anand, Y. Zhou, O. F. Onder, E. Tsougenis, H. Chen, P.-A. Heng, J. Li, Z. Hu, Y. Wang *et al.*, "A multi-organ nucleus segmentation challenge," *IEEE Transactions on Medical Imaging*, vol. 39, no. 5, pp. 1380–1391, 2020.

[137] Y. Hong, Y. J. Heo, B. Kim, D. Lee, S. Ahn, S. Y. Ha, I. Sohn, and K.-M. Kim, "Deep learning-based virtual cytokeratin staining of gastric

IEEE *Access*

carcinomas to measure tumor–stroma ratio," *Scientific Reports*, vol. 11, no. 1, p. 19255, Sep. 2021.

NEEL KANWAL received his bachelor's degree in Electronics Engineering from the National University of Science & Technology in 2015 and M.Sc in Computer Networks Engineering from Politecnico di Torino in 2020. He is currently doing his Ph.D. at the University of Stavanger (UiS), Norway, under the supervision of Prof. Kjersti Engan. He is a member of the Biomedical Data Analysis Laboratory at the Department of Electrical Engineering and Computer Science UiS. He worked as a Laboratory Engineer at FAST University, Karachi. The Ph.D. is part of CLARIFY (European Marie-Curie Program), which aims to develop a robust diagnostic environment for digital pathology. His research interest includes preprocessing, segmentation, and anonymization of histological whole slide images.

FERNANDO PÉREZ BUENO received the degree in Telecommunications Engineering from the Universidad de Granada in 2015 and the M.S. degree in Data Science and Computer Engineering from the University of Granada in 2019. He started the Ph.D. degree in 2018 at the University of Granada under the supervision of Prof. Molina, being a member of the Visual Information Processing Group in the Department of Computer Science and Artificial Intelligence. His research interests focus on the use of Bayesian modeling and inference to solve different problems related to image restoration and machine learning. Centered in cancer histopathological images improvement and classification. During his research, he has addressed several problems, such as blind image deconvolution, image denoising and image classification.

ARNE SCHMIDT received the bachelor's degree in mathematics from the Freie Universität Berlin, in 2015, and the master's degree in mathematics from the Technische Universität Berlin, in 2018. He is currently pursuing the Ph.D. degree with the University of Granada, under the supervision of Prof. Rafael Molina. He worked as a Programmer specialized in deep learning with Astrofein GmbH, Fraunhofer Heinrich Hertz Institut, and TomTom. The Ph.D. is part of the CLARIFY project which focuses on digital pathology for cancer classification. His research interests include probabilistic deep learning, Gaussian processes, and crowdsourcing with applications to medical images.

RAFAEL MOLINA (Senior Member, IEEE) received the M.Sc. degree in mathematics (statistics) and the Ph.D. degree in optimal design in linear models from the University of Granada, Granada, Spain, in 1979 and 1983, respectively. He was the Dean of the Computer Engineering School, University of Granada, from 1992 to 2002, where he became a Professor of computer science and artificial intelligence, in 2000. He was the Head of the Computer Science and Artificial Intelligence Department, University of Granada, from 2005 to 2007. He has coauthored an article that received the runner-up prize at reception for early stage researchers at the House of Commons in 2007. He has coauthored an awarded Best Student Paper at the IEEE International Conference on Image Processing in 2007, the ISPA Best Paper in 2009, and the EUSIPCO 2013 Best Student Paper. His research interest focuses mainly on using Bayesian modeling and inference in image restoration (applications to astronomy and medicine), super-resolution of images and video, blind deconvolution, computational photography, source recovery in medicine, compressive sensing, low-rank matrix decomposition, active learning, fusion, supervised learning, and crowdsourcing. He has served as an Associate Editor for Applied Signal Processing, from 2005 to 2007, and the IEEE TRANSACTIONS ON IMAGE PROCESSING, from 2010 to 2014. Since 2011, he has been serving as an Area Editor for Digital Signal Processing.

KJERSTI ENGAN (Senior Member, IEEE) is a professor at the Electrical Engineering and Computer Science Department at the University of Stavanger (UiS), Norway. She received the BE degree in electrical engineering from Bergen University College in 1994 and the M.Sc. and Ph.D degrees in 1996 and 2000 respectively, in electrical engineering and information technology from the UiS. She is the leader of the Biomedical data analysis lab, BMDLab, at UiS. Her research areas include signal and image processing and machine learning with emphasis on medical applications and dictionary learning for sparse signal and image representation. She has a particular interest in AI for newborn survival, stroke detection from CTP imaging and AI in computational pathology. She is a senior member of IEEE. She has served as Associate editor and Senior Area editor for IEEE Signal Processing Letters and as a member of IEEE Image, Video, and Multidimensional Signal Processing Technical Committee (IVMSP), and as associate editor for SIAM Journal on Imaging Sciences (SIIMS).

• • •