

The impact of pre- and post-image processing techniques on deep learning frameworks: A comprehensive review for digital pathology image analysis

*Original*

The impact of pre- and post-image processing techniques on deep learning frameworks: A comprehensive review for digital pathology image analysis / Salvi, M.; Acharya, U. R.; Molinari, F.; Meiburger, K. M.. - In: COMPUTERS IN BIOLOGY AND MEDICINE. - ISSN 0010-4825. - 128:(2021). [[10.1016/j.combiomed.2020.104129](https://doi.org/10.1016/j.combiomed.2020.104129)]

*Availability:*

This version is available at: 11583/2858398 since: 2020-12-19T19:14:46Z

*Publisher:*

Elsevier Ltd

*Published*

DOI:[10.1016/j.combiomed.2020.104129](https://doi.org/10.1016/j.combiomed.2020.104129)

*Terms of use:*

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

*Publisher copyright*

(Article begins on next page)



Contents lists available at ScienceDirect

# Computers in Biology and Medicine

journal homepage: <http://www.elsevier.com/locate/complbiomed>

## The impact of pre- and post-image processing techniques on deep learning frameworks: A comprehensive review for digital pathology image analysis

Massimo Salvi<sup>a,\*</sup>, U. Rajendra Acharya<sup>b,c,d</sup>, Filippo Molinari<sup>a</sup>, Kristen M. Meiburger<sup>a</sup>

<sup>a</sup> Politecnico di Torino, PoliToBIOMed Lab, Biolab, Department of Electronics and Telecommunications, Corso Duca Degli Abruzzi 24, Turin, 10129, Italy

<sup>b</sup> Department of Electronics and Computer Engineering, Ngee Ann Polytechnic, Singapore

<sup>c</sup> Department of Biomedical Engineering, School of Science and Technology, SUSS University, Clementi, 599491, Singapore

<sup>d</sup> Department of Bioinformatics and Medical Engineering, Asia University, Taiwan

### ARTICLE INFO

#### Keywords:

Digital pathology  
Histology  
Deep learning  
Image analysis  
Pre-processing  
Post-processing

### ABSTRACT

Recently, deep learning frameworks have rapidly become the main methodology for analyzing medical images. Due to their powerful learning ability and advantages in dealing with complex patterns, deep learning algorithms are ideal for image analysis challenges, particularly in the field of digital pathology. The variety of image analysis tasks in the context of deep learning includes classification (e.g., healthy vs. cancerous tissue), detection (e.g., lymphocytes and mitosis counting), and segmentation (e.g., nuclei and glands segmentation). The majority of recent machine learning methods in digital pathology have a pre- and/or post-processing stage which is integrated with a deep neural network. These stages, based on traditional image processing methods, are employed to make the subsequent classification, detection, or segmentation problem easier to solve. Several studies have shown how the integration of pre- and post-processing methods within a deep learning pipeline can further increase the model's performance when compared to the network by itself. The aim of this review is to provide an overview on the types of methods that are used within deep learning frameworks either to optimally prepare the input (pre-processing) or to improve the results of the network output (post-processing), focusing on digital pathology image analysis. Many of the techniques presented here, especially the post-processing methods, are not limited to digital pathology but can be extended to almost any image analysis field.

### 1. Introduction & overview

Recently, computerized approaches have been rapidly developing in the field of medical image analysis with the aim of providing clinical information, integrating second opinions and minimizing human intervention. An exponentially growing field in computerized approaches are methods based on deep learning, with studies showing how deep neural networks have reached the performance of state-of-the-art methods in almost all medical imaging fields [1,2].

Deep neural networks are a type of artificial neural network that are many layers deep, meaning there are many layers in between the input and output layer. The deep architecture allows the learning of more complex models compared to shallow architectures, although it also increases the number of important training parameters, such as the number of layers and number of units per layer.

The recent increase in both dataset sizes and computing power have allowed the application of Convolutional Neural Networks (CNNs) to the

specific case of image analysis, which apply convolutions on the input image. CNNs are the most commonly used deep network, and they are trained on either the entire image or on image patches and the important features are learned by optimizing a specific loss function. During the training process, the weights for each neuron making up a neural layer are learned, and once the training phase is completed, the trained network is then used for inference on new images/image patches.

There are many advantages of CNNs compared to more traditional methods. Mainly, it provides the benefit of automatically learning high-level useful features directly without having to extract handcrafted features, and has a hierarchical feature representation, allowing these multilevel representations from a pixel to high-level semantic features which are learned automatically [3,4]. Moreover, CNNs can provide a semantic segmentation by associating each pixel of the input image/patch to a label or class, and there is the opportunity to jointly optimize numerous related tasks together, such as combining both classification and bounding box regression [5].

\* Corresponding author. Corso Duca degli Abruzzi, 24 - 10129, Torino, Italy.  
E-mail address: [massimo.salvi@polito.it](mailto:massimo.salvi@polito.it) (M. Salvi).

<https://doi.org/10.1016/j.complbiomed.2020.104129>

Received 17 October 2020; Accepted 13 November 2020

Available online 21 November 2020

0010-4825/© 2020 The Authors.

Published by Elsevier Ltd.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

While CNNs are a powerful tool for image analysis, there are also some drawbacks. In particular, deep networks are complex with an enormous amount of training parameters and it can be difficult to interact with any single layer within the deep network. Moreover, deep networks in general are sometimes viewed as a black-box that does not explain their predictions in a way that humans can understand [6]. Still, the advantages often outweigh the disadvantages and over the recent years CNNs have become the most commonly used method in image analysis.

A field of medical imaging where CNNs have been extensively used is digital pathology, in which histology slides are digitized to produce very high-resolution images typically of the whole slide thanks to whole slide digital scanners [7]. The analysis of histology slides is fundamental for cancer diagnosis and grading, typically done by an expert pathologist, and is becoming more and more complex due to the rise in cancer incidence and patient-specific treatment options, requiring the attentive analysis of a large number of slides for a complete diagnosis [8]. Moreover, pathologists must often extract a number of quantitative parameters that are required for commonly used grading systems (e.g., cell counting, area, length, percentage of a specific cell presence within slide). All of these issues bring about a very high inter and intra operator variability [9] and occasionally a quantitative measure, such as the percentage of a specific cell presence within a whole slide, really becomes only a qualitative assessment which can depend on the pathologist's expertise.

Due to these issues, many computerized techniques have been developed to process the digitized histology slides in all of the three main computer vision tasks, which are: classification, object detection, and segmentation. All three of these tasks have an image as an input but the output they provide is different. *Image classification* has the task of predicting the class or type of an object within the input image, so the provided output is simply a class label. *Object detection* must locate the presence of various objects with a bounding box, and subsequently classify the located objects within the input image. *Object segmentation*

extends upon object detection in that the recognized objects are located not by using a coarse bounding box but rather by highlighting the specific pixels of the object. Over the years, numerous deep learning network architectures have been proposed in the field of digital pathology for classification (e.g., cancer recognition), detection (e.g., mitosis counting), and segmentation (e.g., nuclei and glands identification) in digital histopathological images. Fig. 1 displays the overall framework of these networks for each of the three previously described tasks. A detailed description of the three network frameworks is provided at the beginning of each dedicated section (i.e., 4, 5, and 6).

Numerous CNN architectures have been proposed in the literature for digital pathology image analysis [10–12]. Over time, deeper networks have been proposed more to exponentially increase the capability of extracting high-level semantic features from histology images [13]. More recently, studies including integrated or fused methods between deep learning networks and more traditional methods of pre- and post-processing have been growing in number, where the pre-processing optimally prepares the input for the network and the post-processing improves the results of the network output. These hybrid frameworks allow a high-level feature extraction by a CNN with the accuracy of conventional techniques. Occasionally pre-processing techniques can be implemented to locate, manage, and reduce typical artifacts in histopathological images, whereas postprocessing methods are used to further reduce the prediction errors of the network. Network prediction errors occur either randomly or due to the intrinsic limitations of the neural network model. For example, in classification tasks, the spatial interactions between neighboring sub-images or patches can be employed to correct the prediction of the network. In detection and segmentation tasks, basic image processing techniques (connected component analysis, morphological operators) can be used along with more advanced or sophisticated methods (deep fusion models) to both exclude false positives and reduce pixel-level prediction errors. Many of the post-processing strategies reported in this review were initially proposed for general machine learning problems but, as deep learning

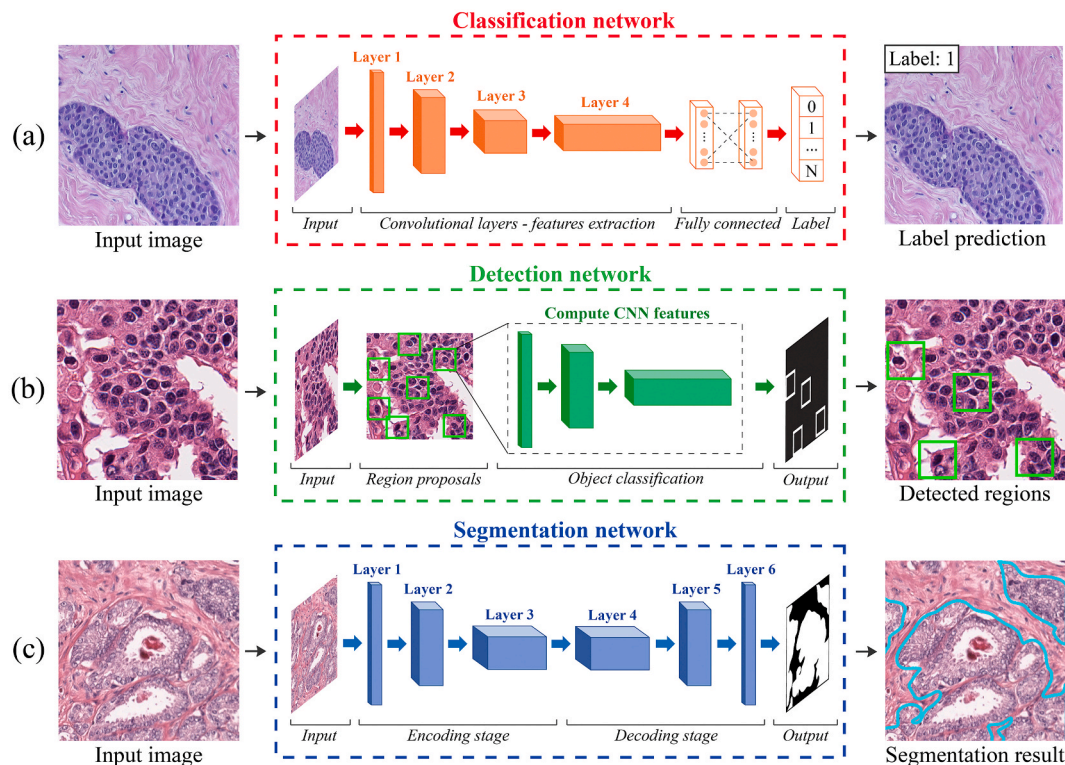


Fig. 1. Typical architectures used in deep learning frameworks. (a) Classification task: starting from the input image, the network predicts the class label. (b) Detection task: the network locates the objects of interest within the image. (c) Segmentation task: the deep network predicts the objects' contour.

became the main methodology for image analysis, they were increasingly integrated into CNNs to obtain better and more robust frameworks. Many studies have shown how the integration between pre- and post-processing methods within a deep learning pipeline can allow to further increase the performance when compared to the network by itself [14–16].

The aim of this review is to provide an overview on the types of pre- and post-processing methods that are used within deep learning frameworks for digital pathology image analysis. The pre-processing methods shown here are mainly specific to the field of digital pathology, whereas the majority of the post-processing methods can be extended to numerous other medical imaging fields.

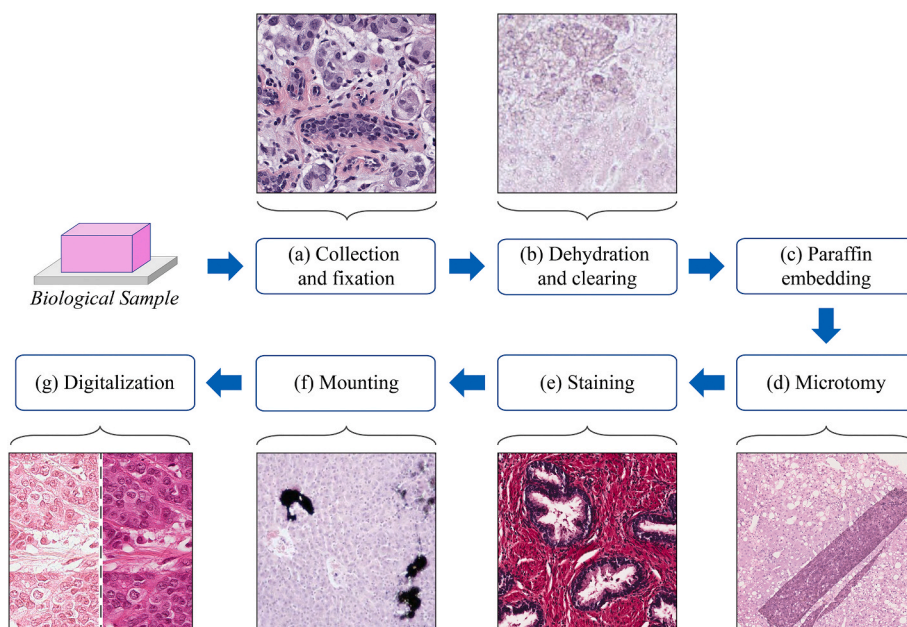
The outline of this review is as follows. Section 2 presents an overview of the characteristics and challenges in digital pathology. Section 3 describes the common pre-processing strategies adopted by deep learning frameworks for histological image analysis. An exhaustive description of the general post-processing approaches used for classification, detection and segmentation tasks is provided in Sections 4, 5 and 6, respectively. A discussion on the methods and future perspectives of hybrid deep learning frameworks concludes the paper.

## 2. Characteristics & challenges in digital pathology

In order to obtain digitalized histological images, specific sequential steps must be followed, which are typically carried out manually by laboratory technicians [17,18]. The histological tissue should be treated to preserve its internal architecture and to present an appearance similar to its aspect inside the living organism (Fig. 2). The following protocol is applied to the histological sample to obtain the corresponding digital image: (a) collection and fixation, (b) dehydration and clearing, (c) paraffin embedding, (d) microtomy, (e) staining, (f) mounting, and (g) digitalization [19]. All of these steps can generate different artifacts that can lower the quality of the histological image:

a) *Collection and fixation*: in order to preserve the tissue from bacteria or cellular enzymes, the specimen is treated with a fixative to prevent chemical and physical alterations [17]. The choice of fixative depends both on the tissue and the analysis to be carried out and it is fundamental for maintaining the appearance of the histological tissue (Fig. 2 - step a).

- b) *Dehydration and clearing*: the goal of this step is to remove the water from the tissue to facilitate the microtomy. If the dehydration is not adequately performed, water drops can be microscopically observed, which makes histological sample regions opaque. As a consequence, microscopic details may be lost and unexpected changes in staining patterns of cells and tissue structures may happen (Fig. 2 – step b).
- c) *Paraffin embedding*: in order to evaporate the solvent used in the previous step and to fill all the spaces within the tissue, the sample is embedded with heated paraffin. At the end of this process, a paraffin block containing the histological sample is obtained.
- d) *Microtomy*: in this step, the paraffin block is progressively sectioned using a microtome [20]. If the tissue section is not uniformly cut, the sample appearance could be compromised. The optimal thickness is 5  $\mu\text{m}$  as it can reveal both the tissue architecture and cell morphology (nucleus and cytoplasm). Cellular structures cannot be displayed correctly on thicker sections ( $>10 \mu\text{m}$ ) while thinner samples ( $<2 \mu\text{m}$ ) allow only the evaluation of the sub-cellular distributions (nucleoli) [19]. Another common artifact is the folding of tissue, which is caused by imprecise placing of the tissue sample on the microscope slide (Fig. 2 – step d).
- e) *Staining*: since the tissue sample becomes translucent after microtomy, specific dyes are applied to the histological slice to highlight the cellular components. The pH and the concentration of the solution as well as the staining time can influence the appearance of the histological slide [21]. The depth of coloration is related to the length of time the sample spends in contact with the dyes (Fig. 2 – step e).
- f) *Mounting*: in this step, the slices are enclosed with a transparent coverslip to protect the tissue from external agents. Coverslip placement can generate artifacts such as the presence of dust (Fig. 2 – step f) or air bubbles or contamination with microorganisms.
- g) *Digitalization*: in the context of digital pathology, the histological slides are also digitized using modern scanners. Sample digitalization can impose a color variation due to the different scanning platform (e.g., sensor chips, ambient illumination, bulbs) and acquisition technology (whiteness correction, image compression). The appearance of the same physical histological slide can vary widely using two different scanners (Fig. 2 – step g). In addition, image blurring can occur if the sample is not aligned with the focal plane of the scanner. The storage condition of the histological specimen can



**Fig. 2.** Protocol applied to a biological sample to produce the corresponding histological slide. During this manual process, various artifacts can be generated. (a) Collection and fixation: the choice of fixative is essential to maintain an optimal contrast between histological structures. (b) Dehydration and clearing: example of dehydration fault. (c) Paraffin embedding. (d) Microtomy: illustration of a tissue-fold artifact. (e) Staining: histological sample with too long staining time. (f) Mounting: specimen with dust contamination. (g) Digitalization: appearance of the same sample digitalized with two different scanners.

also alter the way in which the tissue interacts with the stain, in addition to its natural discoloration, causing faded samples [22].

Although the color variability of histological specimens only partially limits the interpretation of images by pathologists, it can dramatically affect the result of automatic image analysis algorithms [23]. Previous studies have shown that the performance of deep learning frameworks for segmentation/classification of histological images deteriorates in the presence of high color variations within the training dataset [22,24,25].

One of the biggest problems in digital pathology using deep learning is the small number of labeled images. Although label information at pixel-level (segmentation tasks) or patch-level (classification tasks) is required to train the deep network, most labels of whole-slide images are at case-level (e.g., global diagnosis) at most. However, only expert pathologists can label the image accurately, and labeling a huge image, such as a whole-slide image, requires a lot of labor [9]. One possible solution to overcome this problem is to reuse publicly ready-to-use data such as the popular ImageNet dataset [10]. In the last few years, several challenges have been proposed in the field of digital pathology [26,27]. However, these public datasets are focused on a specific disease or task and they cannot be employed for other applications [9]. For this reason, an “open-data” mentality should be adopted to increase the number of public datasets and images. Moreover, there is an inherent challenge due to big data: compared to other imaging modalities, a digital histological image of an entire slide can reach enormously large dimensions. Typically, a whole-slide histological image can present  $100,000 \times 100,000$  RGB pixels which contain complex cellular patterns and contextual biology that carry notably more information. Despite the attractive qualities of deep neural networks, it is still prohibitive to apply them in huge high-resolution images as the size of the network is limited mainly by the amount of memory available on the workstation. In the field of digital pathology, computer-aided diagnosis (CAD) systems have a very high expected performance, both in terms of accuracy and computational time. However, a CAD system must also overcome important computational challenges, in particular for real-time applications, such as the run-time and memory hold-ups due to the processing of high-resolution images (e.g.,  $40\times$  magnification).

### 3. Pre-processing algorithms

The need for a standardization of both procedures and reagents in histological practice is highlighted in the study by Lyon et al. [28]. However, complete standardization cannot be achieved with the current technology, due to manual sectioning variability and stains fading over time. To minimize visible variability in staining and its impact on diagnostic quality, the current practice is limited to procedural and physical quality-control methods, such as subjective visual evaluation of stain quality and interlaboratory staining comparisons. However, the color appearance of histological samples can still vary significantly across laboratories and even across staining batches within the same lab. These variations in tissue and stain appearance complicate quantitative tissue analysis [29].

Considering the current challenges in digital pathology (big images, tissue artifacts, stain variability, etc.), specific pre-processing and data curation steps are mandatory to train a stable deep learning model. In recent years, several pre-processing techniques have been proposed to mitigate the artifacts caused by the manual preparation of histological slides. The term “pre-processing” refers to all the strategies applied to the raw data (i.e., the whole image of the histological specimen) in order to optimally prepare the network input so as to obtain a more robust and accurate final model. The pre-processing methods proposed so far can be grouped into three categories: i) tissue & artifact detection, ii) stain color normalization algorithms and iii) patch selection techniques. Table 1 summarizes all the pre-processing strategies described in this section, along with the dataset used for their validation.

#### 3.1. Tissue & artifact detection

In whole slide imaging (WSI), the quality of scanned images is an interplay between the condition of the tissue slide itself and the hardware specifications of the scanning device [30]. Tissue artifacts such as bubbles and folds affect the efficiency of a whole slide scanning system in selecting the focus points. The presence of these artifacts can produce blurred or unfocused images [31,32]. For this reason, information on the location of WSI artifacts should be known to produce the best image quality.

Histopathology slides typically contain a tissue area of approximately  $15 \text{ mm} \times 15 \text{ mm}$ . Whole digital slides are captured at an incredibly high resolution, resulting in images that can have a size of up to several gigapixels. Given that processing very large images requires a high computational cost, it is common practice to first identify the slide regions that are of clinical interest before performing a more detailed image analysis. In fact, there are typically large sections of the whole slide that do not contain histological tissue, which should be removed from the detailed image analysis in order to reduce computational time. For this reason, tissue segmentation is an essential prerequisite for an accurate and efficient diagnosis in digital pathology.

In the last decade, several approaches have been proposed to perform histological tissue segmentation [33,34]. Wang et al. [35] and Vandenbergh et al. [36] applied a global threshold to remove a large portion of non-informative background. Ertosun et al. [37] employed a hysteresis thresholding strategy, while Bug et al. [38] combined prior knowledge with morphological filters for foreground extraction. Arvaniti et al. [39] automatically detected the tissue regions using a three-step pipeline. A Gaussian filter was initially applied to remove noise, followed by global thresholding to separate tissue from background. Then, the detected tissue mask was further refined using morphological operators. Some authors have employed a threshold strategy on the optical density (OD) image and HSV color space [4,40]. For example, Litjens et al. [4] and Ambrosini et al. [41] implemented a fixed threshold on the optical density of the RGB channels to discard the background regions. Salvi et al. [16] employed an RGB high-pass filter to enhance the texture of the histological tissue. Then, Otsu thresholding and morphological operators were applied to obtain the tissue binary mask. Wang et al. [40] adopted a threshold-based segmentation method where the RGB image of the sample was first converted to the HSV color space and the optimal threshold was calculated for each of the three channels. Then, the final tissue mask was obtained by combining the segmentation results of the H and S channels. Recently, more sophisticated strategies were proposed to locate the tissue within the histological slide. Bándi et al. [34] compared traditional techniques (i.e. those that employ a threshold) with deep convolutional neural networks to perform tissue segmentation. Foreground extraction using CNNs outperformed simple thresholding strategies, allowing up to a 6.7% increase of the Jaccard index.

However, a good tissue segmentation is not sufficient to guarantee satisfying results for a computer-aided system. When analyzing a WSI, expert pathologists naturally avoid tissue regions with artifacts, so a computer-based system should also be able to both detect and avoid these artifact regions. Indeed, the performance of deep neural networks deteriorates when applied to images containing artifacts [42]. One of the most common artifacts in WSI is image blur. Image blurring can occur during the acquisition of a WSI, for example when a portion of tissue is not aligned with the focal plane of the scanner (Fig. 2i). Automatic detection of image blur areas can improve the quality of WSI-based diagnostic pipelines [31]. Gao et al. [43] automatically detected out-of-focus regions using an AdaBoost classifier. Several texture features were extracted from each image patch in order to distinguish between in-focus and out-of-focus areas. Wu et al. [31] proposed a classifier trained on local pixel-level metrics. Histogram features of local blur metrics were then used for the classification of blurry and sharp image patches. The usage of local features instead of global features

increased the blurred image region classifier accuracy by 22%. Blurry regions can be prevented during the image acquisition stage by using advanced scanning systems with dynamic focus. However, avoiding tissue folds (i.e., when a thin piece of tissue folds on itself) is difficult to prevent during slide preparation. As a result, the tissue section is thicker at these areas, and a clear difference in saturation values can be observed between normal and folded tissue. The folded parts are remarkably more saturated than the normal tissue regions (Fig. 2e). Palokangas et al. [32] proposed an automated algorithm for tissue fold segmentation. First, the image was converted to HSI color space and the intensity and saturation components were processed to enhance the discrimination of the objective pixels. Then, k-means clustering was performed to detect all the fold pixels. Bautista et al. [44] employed an adaptive shifting of the RGB values based on the difference between the luminance and saturation of each image pixel. This approach allowed to effectively outline the presence of tissue folds while preserving the hue of other tissue structures. Kothari et al. [42] proposed an automated method for detecting tissue folds in WSIs using color and connectivity properties of tissue structures. The strategy consisted of the segmentation of tissue folds in low-resolution WSIs using adaptive thresholding based on the connectivity of tissue structures. The threshold was then combined with a neighborhood criterion to find tissue folds.

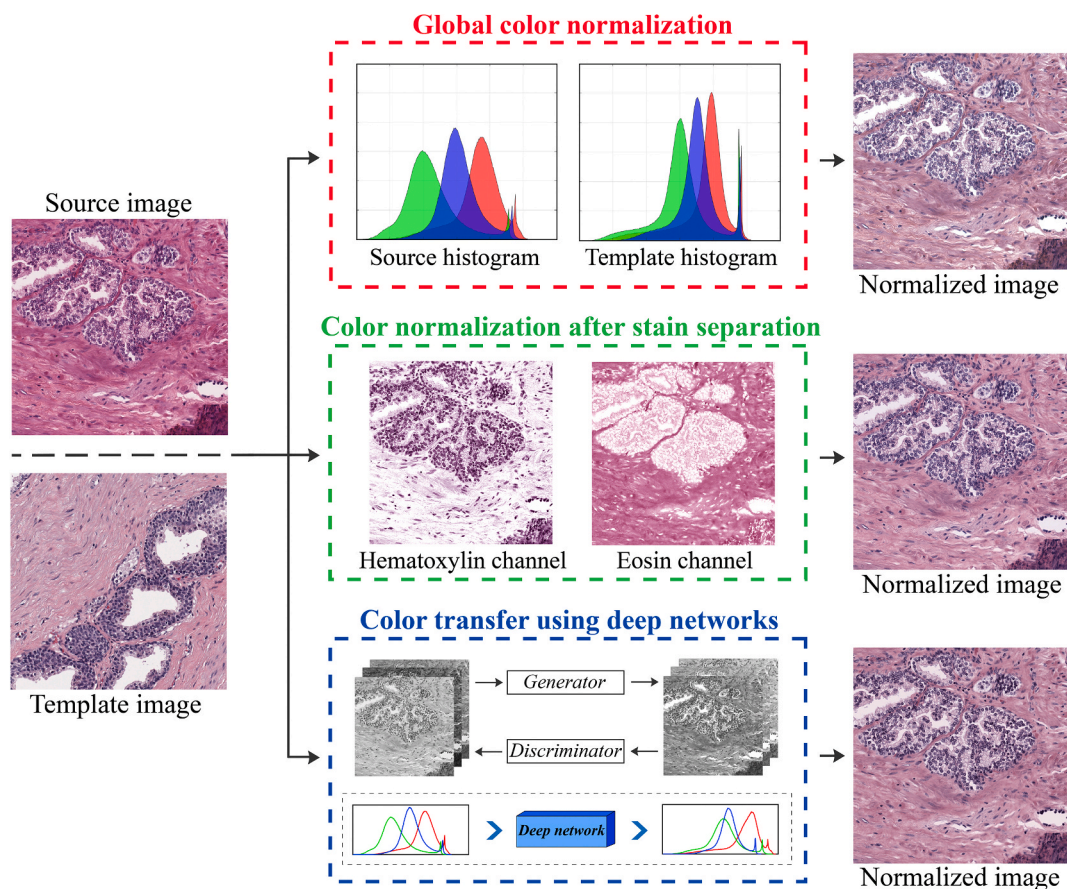
The correct detection of tissue and artifacts within the histological image is essential for the development of automatic pipelines. CADs can analyze a whole slide in a much quicker way by processing image regions containing only histological tissue (Table 1). Similarly, artifact detection is fundamental in that it can allow the exclusion of processing regions containing an altered morphology or intensity (tissue folds, out-

of-focus regions, etc). For example, Kothari et al. [42] have shown that the exclusion of tissue artifacts from processing allows increasing the AUC (Area Under the ROC Curve) of CAD for cancer detection up to 5%.

### 3.2. Stain color normalization

Stain normalization is a common pre-processing step in almost all of the deep learning frameworks in digital pathology [45,46]. The procedure of stain normalization involves transforming an image  $I$  into another image  $I_{NORM}$ , through the operation  $I_{NORM} = f(I, \theta)$ , where  $\theta$  is a set of parameters extracted from a predefined template image and  $f(\cdot)$  is the mapping function that matches the visual appearance of a given image to the template image [46]. The template image is a single image with the most optimal visual appearance and tissue staining. As a result, all stain-normalized images will have their intensity distribution mapped to match the color distribution of the template image. Based on the approach employed to normalize the histological image, the current stain normalization methods can be classified into: (1) Global color normalization, (2) Color normalization after stain separation, and (3) Color transfer using deep networks. Fig. 3 shows the normalization strategies used in current deep learning frameworks.

*Global color normalization* is done after separating intensity and color information using different color spaces. Histopathological images present autocorrelation coefficients or spatial dependency of pixel intensity values that make global color normalization a very suitable technique. Reinhard et al. [47] implemented a global color transformation between the target and source image in the  $\alpha\beta$  color space using Principal Component Analysis (PCA). In order to do so, the target image mean



**Fig. 3.** Stain normalization strategies in deep learning frameworks. Global color normalization: global information (e.g., RGB histogram, brightness) is extracted from the template image and then mapped to the source image. Color normalization after stain separation: the contribution of the individual dye is isolated to alter the original image according to the color distribution of the template image. Color transfer using deep networks: a style transfer approach is employed to change the stain-style of the source image to the template one.

color is transferred to the source image in a way so that the source image intensity variations are preserved, and the obtained contrast of the processed image is roughly the same as the target image contrast. Another global color normalization technique is known as histogram specification [29,48]. In this method, the source image histogram is mapped to the target image histogram to make it, so the color statistics and brightness of the source image resemble those of the target image. Both histogram specification and global histogram enhancement methods employ contrast stretching, which imposingly stretches the source image histogram to mimic the target image histogram, producing an unnaturalness in the obtained image. This unnatural process may occasionally produce artifacts in the processed image. Moreover, if the target and source image present cellular structures that are very different between each other, these global approaches based on the image histogram can ultimately fail in color normalization.

*Color normalization after stain separation* is carried out by isolating the contribution of the individual dyes used during the staining process. The color intensity assumed by the specific cell component depends on the absorption of the amount of stain, according to the Beer-Lambert law [49]. The RGB intensity values cannot be directly used for stain separation since the relationship between the concentration and light intensity of each stain is nonlinear. Thus, before stain separation, the source image should be transformed into the OD (optical density) space so that they act linearly [50]. The intensity of the image in the OD space ( $V$ ) can be defined as the logarithm of the ratio of incident ( $I_0$ ) to transmitted ( $I$ ) light intensity:

$$V = \log_{10} \left( \frac{I_0}{I} \right) = W \cdot H \quad (1)$$

According to equation (1), the corresponding OD value of each pixel can be defined as the product of the stain color appearance matrix ( $W$ ) and the stain density map ( $H$ ). By using  $W$ , an image can be decomposed into each individual stain components via color deconvolution [51]. The obtained components can then be altered and recomposed into an image that appears to contain different stain amounts when compared to the original image. In the last decade, several supervised methods have been proposed to estimate the stain color appearance matrix  $W$  [19,50]. The automated extraction of the matrix  $W$  has been done by using: (i) prior information of the stain vectors [52], (ii) singular value decomposition [51,53] (iii) support vector machines [25], (iv) gaussian mixture model [54] and (v) cellular structures segmentation [55–57]. Recently, unsupervised techniques were also applied to normalize the histological images. Spectral matching [22,58], Non-negative Matrix Factorization (NMF) [59], and Independent Component Analysis (ICA) [60] methods have been employed to estimate both the stain density map  $H$  and the stain color appearance matrix  $W$ . Recently, Gupta et al. [61] proposed a unified framework that corrects stain chemical, illuminant variation and color quantity by exploiting the color vector space's geometry.

*Color transfer using deep networks* is done by using a generative learning and style transfer approach. Style transfer consists of discovering image representations that independently model dissimilarities in the semantic image content and its subsequent presented style. In the last few years, Generative Adversarial Networks (GANs) have been extensively used to perform stain normalization in histological images [62]. GANs are deep networks that take advantage of adversarial training. Adversarial training consists of a generative and a discriminative model trained through an objective function using a minmax game. The goal in GANs is learning a generative distribution  $P_G(x)$  that matches the real data distribution  $P_{DATA}(x)$ . The GAN model includes a generator network  $G$  that generates sample  $G_z$  using a noise variable  $z$ . The generator 'plays' against a second network, the adversarial discriminator network  $D$ , that tries to distinguish between real data ( $x$ ) and generated ones ( $z$ ). The objective function of the minmax game is defined as:

$$\min_G \max_D V(G, D) = E_{x \sim P_{DATA}} [\log(D(x))] + E_{z \sim \text{noise}} [\log(1 - D(G_z))] \quad (2)$$

While the generative network  $G$  tries to minimize the objective function, the adversarial discriminator network  $D$  learns to maximize it until both networks arrive at their optimal state. Through the above procedure, every stained image might be transferred to have the desired stain-style. Cho et al. [63] presented a stain-style transfer method based on GANs to minimize the difference between latent features of the source image and that of the target image. BenTaieb et al. [64] built a discriminative model with an intrinsic stain normalization component while Shaban et al. [65] designed a GAN for stain normalization without the need to pick a template image. Swiderska-Chadaj et al. [66] proposed a cycle-GAN to correct both color and style of prostate histological images. Hence, CNNs can produce powerful deep feature representations that can be exploited to independently manipulate both the style and content of natural images [67].

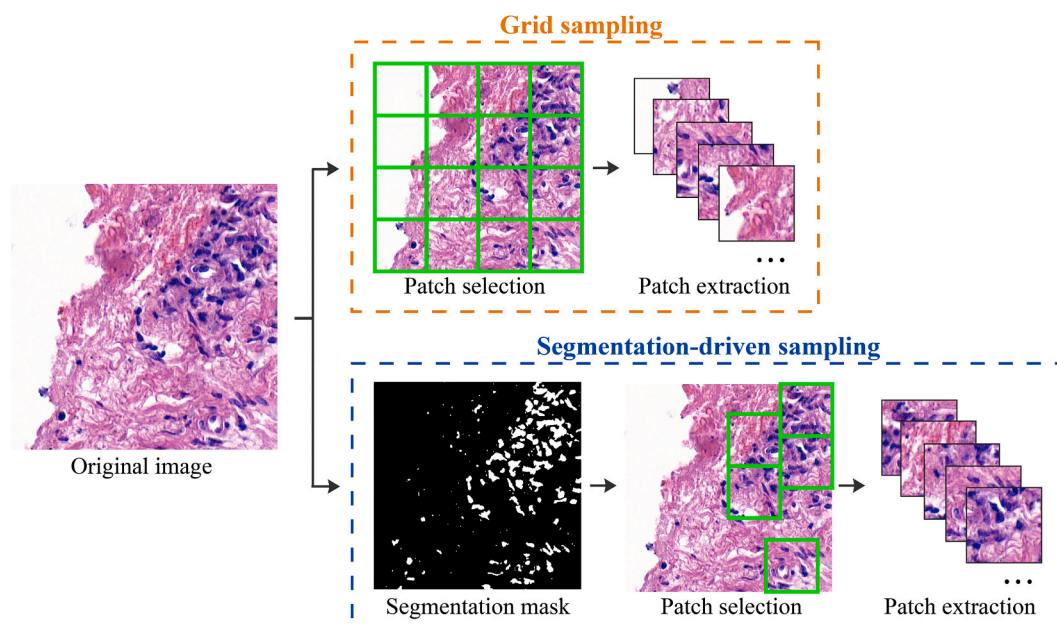
Stain color normalization has been shown to have a great influence on deep learning frameworks (Table 1). Almost all the published deep learning methods for quantitative analysis of histological images integrate a stain normalization process [46,68]. With color normalization and regardless of the task or the dataset, we consistently observed a rise in the performance of the deep network. Stain normalization has increased the accuracy of a CAD for prostate and breast cancer detection [45,53,56,63,69,70], colon glands segmentation and classification [46,58,64], nuclei segmentation [57,71,72], and mitosis detection [64]. This confirms the need for normalization in automatic histopathology image analysis.

### 3.3. Patch selection techniques

Deep learning algorithms are generally applied on entire biopsies or WSIs. However, applying a CNN directly to a WSI has several drawbacks. First of all, discriminative details could be lost due to the necessity of extensive image downsampling. Secondly, a CNN could potentially learn from only one of the many discriminative patterns in the image, which would result in data inefficiency [73]. In histological images, the discriminative information is encrypted in high resolution patches; for this reason, the key is to train the network on the high-resolution patches and subsequently predict the label of the entire WSI based on the patch-level predictions. DL techniques learn the models directly from the provided data, so it is of utmost importance to select representative patches of the image for training. Over the years, several patch selection techniques have been proposed to train deep networks, ranging from random sampling to segmentation-guided tiles extraction (Fig. 4).

The most common strategy is to split the target WSI into a grid and adopt a sliding window approach to extract all the patches for the CNN training [41,74,75]. When the image presents classes that are not equally represented, del Toro et al. [76] and Lucas et al. [77] proposed a method based on randomly extracting a fixed number of patches for each class. This allowed the network to be trained using the same number of patches for each class, resulting in the network not being polarized towards a specific class [78].

For prostate cancer detection, some authors proposed a glands-guided patch extraction strategy. In this way, patches were extracted only within the glandular areas, i.e. where prostate cancer occurs. Zhou et al. [14] performed a glands segmentation using a k-means algorithm in the LAB colorspace. Then, the CNN was trained using only the patches extracted from the segmented glands. Compared to a simple grid approach, this strategy led to an improvement in cancer detection accuracy of 23.9%. This higher performance can be contributed to the fact that the k-means algorithm selected only the useful malignant glands for the network training and testing, while the discarded areas by the k-means algorithm (e.g., stroma, cell cytoplasm, cell nuclei) are not significant for correct cancer grading. Using an analogous approach, Chen et al. [45] employed a k-means algorithm on the HSV colorspace to



**Fig. 4.** Patch selection techniques in deep learning frameworks. Grid sampling: the image is split into a grid and a sliding window approach is adopted to extract all the patches for the CNN training. Smart sampling: the patch selection is guided by an automated segmentation mask to extract images only on regions of interest (in this case areas containing nuclei).

detect the glandular areas. Then the CNN was trained and tested only inside those regions. This patch extraction strategy resulted in a 26.9% improvement in cancer detection accuracy.

Similarly, several authors have proposed a nuclei-guided sampling strategy for breast cancer detection. Since breast cancer distorts the nuclei texture, shape, and spatial organization, some authors have proposed a smart patch extraction for both training and network testing [79,80]. These approaches are based on extracting patches only where there is a high density of nuclei, thus bypassing the adipose and stromal regions [81]. Zheng et al. [80] employed a color deconvolution to identify the position of the nuclei within the image. In particular, non-informative regions (containing few nuclei or large empty regions) were detected by a threshold of the nucleus number, thus avoiding unnecessary calculations. This framework achieved a speed-up of 500% of the test-time inference, while maintaining an accuracy comparable to previously published papers. Xu et al. [82] presented a deep hybrid attention approach to breast cancer classification. They realized a soft-attention network that highlighted only the valuable information (i.e., clusters of nuclei) within the image. Then, the patches were extracted only on the detected regions. Based on this approach, only a fraction of the pixels in the raw image was processed, resulting in a significant saving in computational resources without sacrificing performances.

The selection of representative patches has a great impact on the training of a deep learning model (Table 1). Compared to the classic grid approach, a smart patch selection allows both to increase the performance of the model [14,45,83] and reduce computational times during the test evaluation [80,82].

#### 4. Post-processing algorithms for classification problems

##### 4.1. Deep networks for classification

In the context of deep networks, image classification is a supervised learning problem in which a set of target classes are defined, and a model is trained to recognize them using labeled example images. Convolutional neural networks (CNNs) can be exploited to progressively extract higher and higher-level representations of the image content; in fact, the CNN takes the raw pixel data as input and “learns” how to

extract specific features, such as textures and shapes, and ultimately infer the object they are a part of. This discovery was a breakthrough in the building of models for image classification, as it precludes the necessity to preprocess the data to derive specific features. The input image of a CNN has dimensions  $W \times H \times C$ , where  $W$  and  $H$  are the width and height of the image in pixels, respectively, and  $C$  is the number of image color channels. In general, a CNN consists of a stack of various modules, which each perform three operations: (i) *convolution*, which creates a filter map by applying numerous different filters over the input image; (ii) *ReLU* (Rectified Linear Unit) transformation to the convolved feature, which introduces nonlinearity into the model and (iii) *Pooling*, where the CNN downsamples the convolved feature, which preserves the most important feature information while still reducing the number of dimensions of the feature map. One of the final modules in a CNN is when the obtained feature map is reshaped into a long vector and then one or more fully connected layers are engaged to carry out the final classification task. The final fully connected layer typically includes a softmax activation function, whose output is a probability value between 0 and 1 for each classification label the model is attempting to predict (Fig. 1a). In this review, any image processing method that takes the softmax activation function and either combines it or modifies it compared to baseline methods is considered as a “post-processing” technique.

The performance of deep networks for image classification is generally assessed by calculating the accuracy. The overall accuracy is a common metric used in classification problems and it is defined as the ratio between the correctly classified images and the total number of images. Recently, CNNs have become the reference algorithm for solving the task of patch-based classification in medical imaging [84]. Moreover, recent challenge competitions in digital pathology [84–86] have shown that CNN-based methods can perform just as well as, if not better than, pathologists at the task of analyzing histopathological images. Several “standard” deep network architectures have been employed in the methods described in the following sections: the *AlexNet* [10] has five convolutional and three fully connected layers with approximately 60 million parameters. It has facilitated the rapid adoption of deep learning by introducing the ReLU as activation functions. *VGGNet* [11] is twice as deep as AlexNet and contains 16 or 19 convolutional layers for a



**Table 1**  
Summary of the pre-processing strategies employed in deep learning frameworks.

Task	Reference	Year	Dataset	Method	Performance/Pre-processing impact
Tissue segmentation	Bug et al. [38]	2015	43 WSI	Thresholding with morphological filters and median blurring	95.9% Jaccard index
	Ertosun et al. [37]	2015	Brain (200 WSI)	Hysteresis thresholding on grayscale image	–
	Litjens et al. [4]	2016	Breast (173 WSI); Prostate (225 WSI)	Thresholding on the optical density of the RGB channels	–
	Wang et al. [40]	2016	Breast (400 WSI)	Global thresholding on HSV color space	Computational time reduced by 82% (after excluding the background)
	Bándi et al. [34]	2017	54 WSI from breast, lymph node, rectum and tongue	Semantic segmentation using a deep network (UNET architecture)	92.9% Jaccard index (6.7% increase compared to simple thresholding)
	Vandenberghe et al. [36]	2017	Breast (74 WSI)	Global thresholding	–
	Arvaniti et al. [39]	2018	Prostate (886 images)	Gaussian filtering and Otsu thresholding followed by morphological operators	–
	Wang et al. [35]	2018	Lung (939 WSI)	Otsu thresholding	–
	Ambrosini et al. [41]	2020	Prostate (128 WSI)	Thresholding on the optical density of the RGB channels	–
	Salvi et al. [16]	2020	Liver (385 images)	RGB high-pass filter followed by Otsu thresholding and morphological operators	99.30% accuracy
Artifact detection	Palokangas et al. [32]	2007	Aorta (10 images)	Tissue fold detection using HSI color space and k-means clustering	–
	Bautista et al. [44]	2010	12 images from breast, esophagus and liver	Tissue fold segmentation through adaptive shifting of the RGB values	37.02% Jaccard index
	Kothari et al. [42]	2013	Kidney (1092 WSI); Ovarian tissue (563 WSI)	Tissue fold segmentation using connectivity-based thresholding and color properties	5% improvement of a cancer-grade prediction model (after excluding tissue folds)
	Wu et al. [31]	2015	Endomyocardial tissue (1000 images)	Classification of sharp and blurry images through local histogram features	98% accuracy
Stain normalization	Gao et al. [43]	2010	120 images from different tissues	Detection of out-of-focus regions using AdaBoost classifier	92.77% accuracy
	Reinhard et al. [47]	2001	Natural images	Color transfer using LAB color space and Principal Component Analysis	–
	Ruifrok et al. [52]	2001	Images from breast and lung	Estimation of color appearance matrix using prior information of stain vectors	–
	Coltuc et al. [48]	2006	Natural images	Image normalization through histogram specification	–
	Macenko et al. [51]	2009	12 sample with various stain combinations	Singular Value Decomposition on OD space	–
	Magee et al. [54]	2009	Colon (12 images); Liver (48 slides);	Probabilistic prior and Bayesian Gaussian Mixture Model	50% more consistency of RGB colors compared to the original image
	Khan et al. [25]	2014	Breast (50 images); Esophagus (12 images); Liver (48 images)	Nonlinear mapping of channel statistics and image reconstruction	11% improvement in Dice coefficient during breast tumor segmentation
	Bejnordi et al. [55]	2015	Liver (30 slides); Lymph node (125 images)	HSD color model followed by structures segmentation and template matching	23.78% improvement on AUC during necrosis quantification
	Li et al. [22]	2015	Breast (58 images); Lymphoma (375 images)	Illuminant normalization and spectral matching	–
	Vahadane et al. [59]	2016	295 images from bladder, colon, prostate, and stomach	Sparse Non-negative Matrix Factorization (SNMF)	–
	Alsubaie et al. [60]	2017	Breast (3 WSI); Colon (7 WSI); Lung (2 WSI)	Wavelet decomposition and Independent Component Analysis	–
	BenTaieb et al. [64]	2017	Breast (9600 images); Colon (165 images); Ovarian tissue (135 images)	Style transfer and generative learning using GANs	Accuracy increase in mitosis detection (18.5%), colon glands segmentation (13.5%), and ovarian cancer detection (16.9%)
	Bug et al. [67]	2017	Lung (9 WSI)	Feature extraction and RGB color shifting using a deep neural network	–
	Cho et al. [63]	2017	Breast (400 WSI)	Stain-style transfer learning through GANs	2.85% AUC increase for breast cancer detection
	Janowczyk et al. [57]	2017	Breast (25 images); Gastro-intestinal tissue (182 WSI)	Sparse AutoEncoders to standardize the color distribution of the image	10% improvement in Dice coefficient for a nuclei segmentation task
	Zanjani et al. [62]	2018	Lymph node (625 images)	End-to-end model based on CNNs to learn image-content and color attributes	–
Anghel et al. [53]	2019		Singular Value Decomposition on OD space	5% improvement on the F1-score during prostate cancer detection	

(continued on next page)

Table 1 (continued)

Task	Reference	Year	Dataset	Method	Performance/Pre-processing impact
Patch selection			Breast (500 WSI); Lymph node (400 WSI); Prostate (126 WSI)		
	Shaban et al. [65]	2019	Breast (400 WSI)	Stain-style transfer learning through GANs	80% improvement on AUC for breast cancer classification
	Tosta et al. [58]	2019	Breast (58 images); Colon (165 images); Lymph node (374 images)	Estimation of sparsity parameters and spectral matching	9.5% improvement on AUC during colon cancer detection
	Zheng et al. [70]	2019	Cervical tissue (47 WSI); Lymph node (1400 WSI); Lung (39 WSI)	Adaptive color deconvolution in OD space	7.2% improvement on AUC for breast cancer classification
	Swiderska-Chadaj et al. [66]	2020	Prostate (717 WSI)	Cycle-GAN normalization	10% improvement on AUC for prostate cancer detection
	Gupta et al. [61]	2020	Bone (60 images); Breast (58 images)	Multi-stain approach based on color vector space's geometry	22.5% improvement on AUC for breast cancer detection
	Salvi et al. [56]	2020	270 images from adrenal, breast, colon, liver and prostate	Stain color adaptive normalization through cellular structures detection	11.28% accuracy improvement for breast cancer detection
	Cruz-Roa et al. [74]	2017	Breast (349 WSI)	Grid sampling	–
	Del Toro et al. [76]	2017	Prostate (235 WSI)	Extraction of a fixed number of patches based on Blue Ratio image (BR)	–
	Zheng et al. [80]	2017	Breast (715 WSI)	Nucleus-guided patch extraction using color deconvolution and Gaussian filter	WSI classification for breast cancer detection up to 19x faster
	Zhou et al. [14]	2017	Prostate (234 WSI)	K-means algorithm to extract relevant patches	24.26% accuracy improvement for prostate cancer classification
	Golatkari et al. [79]	2018	Breast (400 images)	Patch extraction based on local nuclear density	10% accuracy improvement for breast cancer classification compared to grid sampling
	Janowczyk et al. [83]	2018	Breast (141 WSI)	A CNN (AlexNet) is employed to identify all the nuclear regions	Computational time reduced by around 85% for a nuclei segmentation task
	Ryu et al. [75]	2019	Prostate (1833 WSI)	Grid sampling	–
	Lucas et al. [77]	2019	Prostate (96 WSI)	Random patch extraction followed by a balanced partition for each class	–
	Xu et al. [82]	2019	Breast (7909 images)	Soft-attention network that select only the relevant patches	Computational time reduced by 85% for breast cancer classification
Chen et al. [45]	2020	Prostate (32 WSI)	Patch extraction using HSV color space and k-means to detect glandular areas	26.9% accuracy improvement in prostate cancer detection	
George et al. [81]	2020	Breast (2027 images)	Color deconvolution to detect cell nuclei and perform patch extraction	–	

**Table 2**  
Summary of the post-processing strategies for classification task.

Task	Reference	Year	Dataset	Method	Performance/Post-processing impact	
Prostate cancer	Kallén et al. [108]	2016	Prostate (213 images)	CNN + majority voting	89.2% accuracy in Gleason grading	
	Litjens et al. [4]	2016	Prostate (225 WSI)	CNN + cumulative patch histogram to perform cancer detection in WSI	0.99 AUC (median analysis) in cancer detection	
	Zhou et al. [14]	2017	Prostate (234 WSI)	CNN + majority voting	70.41% accuracy in intermediate prostate cancer classification	
	Arvaniti et al. [39]	2018	Prostate (886 images)	CNN + majority voting	57.75% accuracy in Gleason grading	
	Bulten et al. [99]	2019	Prostate (1243 WSI)	UNET + thresholding on tumoral patches	61.33% accuracy in Gleason grading	
	Campanella et al. [110]	2019	Prostate (24,859 WSI)	Slide-level aggregation using recurrent neural network (RNN)	0.99 AUC in cancer detection	
	Duong et al. [106]	2019	Prostate (1005 images)	ResNet with scale embedding	95.3% accuracy in cancer detection. 1.9% accuracy improvement compared to the single network	
	Karimi et al. [15]	2019	Prostate (333 images)	Multiscale decision aggregation using a logistic regression model	86% accuracy in Gleason grading. 10% accuracy improvement compared to the single network	
	Lucas et al. [77]	2019	Prostate (96 WSI)	GoogleNet + majority voting	81% accuracy in Gleason grading	
	Nagpal et al. [107]	2019	Prostate (1557 WSI)	Nearest-neighbor classifier to label the WSI starting from the CNN heatmap	69% accuracy in Gleason grading	
	Ryu et al. [75]	2019	Prostate (1833 WSI)	CNN + majority voting	59% accuracy in Gleason grading	
	Breast cancer	Cruz-Roa et al. [117]	2014	Breast (162 WSI)	CNN + thresholding	84.23% balanced accuracy in cancer detection
		Kovalev et al. [118]	2016	Breast (340 WSI)	GoogleNet + thresholding	0.99 AUC in cancer detection
Litjens et al. [4]		2016	Breast (173 WSI)	CNN + connected component analysis	0.90 AUC in cancer detection	
Wang et al. [40]		2016	Breast (400 WSI)	Feature extraction from CNN heatmap to perform slide-based classification	0.93 AUC in cancer detection	
Araújo et al. [116]		2017	Breast (269 images)	VGGNet + majority voting	77.8% accuracy in cancer grading	
Couture et al. [127]		2017	Breast (571 images)	CNN + SVM classifier	82% accuracy in cancer grading	
Liu et al. [78]		2017	Breast (270 WSI)	CNN + ensemble model for WSI classification	0.97 AUC in cancer detection	
Vandenbergh et al. [36]		2017	Breast (74 WSI)	CNN + thresholding	78% accuracy in cancer detection	
Awan et al. [123]		2018	Breast (400 images)	ResNet + majority voting	90% accuracy in cancer detection	
Gecer et al. [121]		2018	Breast (240 WSI)	CNN + majority voting	0.92 AUC in cancer detection	
Iesmantas et al. [119]		2018	Breast (400 images)	CNN + majority voting	87% accuracy in cancer grading	
Kohl et al. [115]		2018	Breast (400 images)	VGG + connected component analysis	94% accuracy in cancer detection. 1.5% accuracy improvement compared to the baseline	
Nawaz et al. [120]		2018	Breast (400 images)	AlexNet + majority voting	81.25% accuracy in cancer grading	
Nazeri et al. [128]		2018	Breast (400 images)	CNN + ensemble model for image-level classification	95% accuracy in cancer detection. 5% accuracy improvement respect to majority voting	
Rakhlin et al. [94]		2018	Breast (400 images)	CNN + maximum probability score	87.2% accuracy in cancer grading	
Vang et al. [129]		2018	Breast (400 images)	GoogleNet + logistic regression model	87.5% accuracy in cancer grading. 6% accuracy improvement respect to the baseline score	
Vesal et al. [69]		2018	Breast (400 images)	ResNet + majority voting	89.58% accuracy in cancer grading	
Wang et al. [122]		2018	Breast (400 images)	VGGNet + majority voting	93% accuracy in cancer grading	
Ahmad et al. [89]		2019	Breast (260 images)	ResNet + majority voting	85% accuracy in cancer grading	
Matos et al. [92]		2019	Breast (7797 images)	GoogleNet + majority voting	88.7% accuracy in cancer detection	
Roy et al. [124]	2019	Breast (400 images)	CNN + majority voting	90% accuracy in cancer grading. 2.5% accuracy improvement compared to simple thresholding		
Le et al. [126]	2020	Breast (1090 WSI)	CNN + patch aggregation using neighborhood information	73% positive predictive value (PPV) in cancer detection. 4% improvement in PPV compared to the baseline		
Yan et al. [98]	2020	Breast (4020 images)	GoogleNet + deep network for image-wise classification	91.3% accuracy in cancer grading.		
Lung cancer	Coudray et al. [130]	2018	Lung (1634 WSI)	GoogleNet + patch aggregation	0.97 AUC in cancer detection	
	Graham et al. [135]	2018	Lung (64 WSI)	ResNet + random forest classifier	81% accuracy in cancer grading. 3% accuracy improvement respect to majority voting.	
	Wang et al. [35]	2018	Lung (939 WSI)	CNN + random forest classifier	97.1% accuracy in cancer detection. 27.4% accuracy improvement respect to majority voting.	

(continued on next page)

Table 2 (continued)

Task	Reference	Year	Dataset	Method	Performance/Post-processing impact
Colon cancer	Li et al. [136]	2019	Lung (200 WSI)	CNN + conditional random fields	79.7% accuracy in cancer detection. 11% accuracy improvement respect to the single deep network
	Wei et al. [132]	2019	Lung (422 WSI)	ResNet + thresholding	76.7% agreement in cancer grading respect to three pathologists
	Sirinukuwattana et al. [139]	2016	Colon (100 images)	CNN + patch aggregation using neighborhood information	0.92 AUC in colon nuclei classification
Brain cancer	Korbar et al. [131]	2017	Colon (2074 images)	ResNet + majority voting	91.3% accuracy in cancer detection
	Xu et al. [140]	2015	Brain (106 images)	CNN + SVM classifier	97.5% accuracy in cancer detection
	Hou et al. [73]	2016	Brain (1000 images)	CNN + logistic regression model	77.1% accuracy in cancer grading. 3.5% and 6.5% accuracy improvement respect to majority voting and SVM, respectively

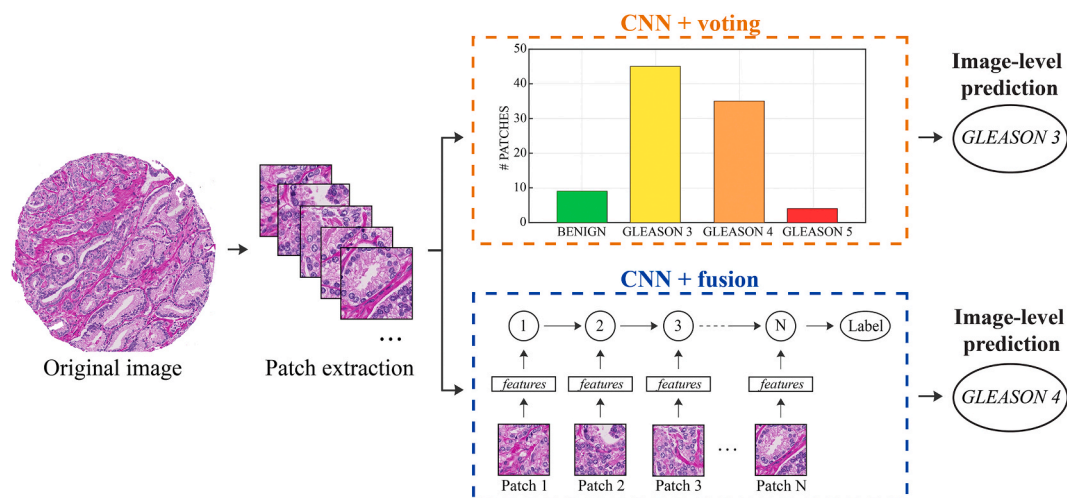


Fig. 5. Post-processing strategies for classification tasks. Prostate cancer grading is used as an explanatory example. CNN + voting: for each classified patch, a ‘voting’ procedure is applied to determine the final class of the entire image. CNN + fusion: this approach integrates a decision model to aggregate the classification of each patch and obtain the label of the entire image.

total of 150 million parameters. This network showed the effect of the network depth on performance: deeper architectures allow to obtain higher accuracy performance, but also lead to optimization challenges (training time, computational power and storage space). *GoogLeNet* [12] presents a 22-layer architecture with around 5 million parameters. This network introduced the ‘Inception’ module that consisted in the concatenation of convolutional layers having different kernel sizes. However, in general, as the depth of a deep network increases, the accuracy gets saturated. *ResNet* [87] addressed the problem of Inception networks by using skip connections while building deeper models. This network was the first to adopt the batch normalization, allowing to design even deeper CNNs (up to 145 layers) without compromising the model’s generalization power.

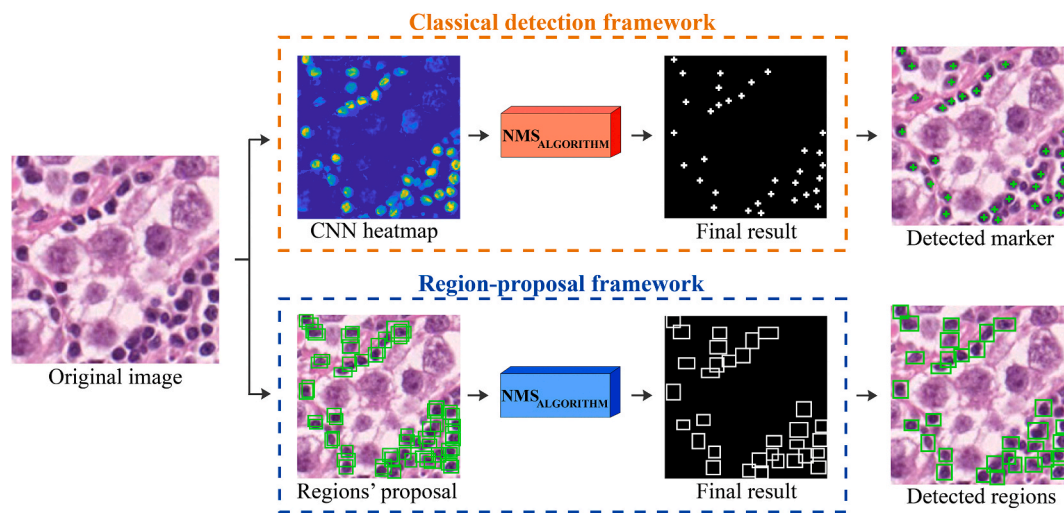
In almost all the works reported in this review, transfer learning strategies are applied to train the network. Transfer learning is a method used to transfer knowledge acquired from one task to resolve another [88,89]. This strategy can overcome the problem of small datasets [90] and, at the same time, it can help reduce the training time [91]. Two main approaches can be adopted to apply transfer learning: (i) take advantage of a pre-trained network as a feature extractor and then use these features to train a new classifier [92–94] or (ii) fine-tuning of the pre-trained network parameters according to the new required tasks [95,96]. Another common technique to improve results and avoid overfitting of a deep network is data augmentation. Since histopathological images do not have a canonical orientation, most of the authors applied data augmentation to increase the robustness of their network [97,98]. The following augmentation procedures were typically used: flipping, scaling, rotating, color alterations (saturation, hue, contrast, and brightness), additive noise, and Gaussian blurring [99]. Recently,

deep networks were also employed to perform data augmentation [100]. Table 2 summarizes all the classification approaches described in this section, along with their post-processing strategy and the database used for their validation.

#### 4.2. Prostate cancer

Prostate cancer (PCa) is the most common cancer in men and the fifth cause of cancer-related death globally [101,102]. The need for an accurate prognostic factor stratification has become mandatory, and the Gleason Score assessment performed on prostate biopsies is considered the gold-standard technique. The Gleason Score is a five-grade based score that evaluates the architecture of neoplastic glands, with 1 representing healthy and well-formed glands and 5 representing the most aggressive gland pattern, showing single cells and necrosis. However, an important issue is the reproducibility of its outcome. Several reports in the literature show how the inter- and even the intra-reproducibility of Gleason Score assessment is very low, and the leading causes of this variance could be identified both in the subjectivity of the evaluation and in its ‘simplicity’ itself [103–105].

In recent years, several deep learning methods have been developed for PCa detection [39,45]. Since pathologists assess a tissue specimen at different resolutions to make a diagnostic decision, Duong et al. [106] proposed a multiscale CNN for prostate cancer grading. In this way, detailed cellular characteristics can be assessed at a higher resolution whereas the overall tissue structure can be observed with a lower resolution. Then, a patch-wise classification approach was applied to each image of the test set. Respect to a single-scale network, the proposed method allowed to obtain a higher accuracy in cancer detection (95.3%



**Fig. 6.** Post-processing strategies for detection tasks. Lymphocyte detection is used as an explanatory example. Classical detection framework: the CNN heatmap is created in a patch-wise manner using a sliding window approach. Then, the non-maxima suppression (NMS) algorithm is employed to locate each object. Region-proposal framework: a selective search generates the regions' proposal and the NMS algorithm deletes overlapping regions to locate the bounding box of each object.

vs 93.4%). Typically, a WSI is classified using a sliding window and a heatmap is generated with the classification results. Starting from the heatmap, the algorithm should aggregate the patch-level classification and then assign a categorical class for the entire image (e.g., benign, malignant, Gleason Score, etc.). In the last few years, several post-processing strategies have been adopted to aggregate the classifications of all patches and give the global class of the WSI [73,107]. These strategies can be grouped into two categories: CNN + voting and CNN + fusion. Fig. 5 illustrates the two most common post-processing techniques adopted in classification tasks.

The CNN + voting approach consists in evaluating the number of patches classified for every considered class. Subsequently, a 'voting' procedure is applied to determine the final class of the entire image. For example, Bulten et al. [99] applied a simple threshold on the percentage of patches classified as 'tumor'. If at least 1% of the patches were classified as tumoral, the entire WSI was labeled as malignant. Litjens et al. [4] constructed the normalized cumulative histogram of the WSI's heatmap, using 100 bins equally spaced between 0 (healthy) and 1 (tumor). A percentile analysis was employed to find the best threshold to divide between benign and malignant cases. Other studies employed majority voting to perform cancer grading [14,75,77,108]. In other words, the final predicted label of a WSI is equal to the predicted label of the patch with maximum probability over all other patches and classes. Using this strategy, Arvaniti et al. [39] obtained a global accuracy comparable with the inter-pathologist agreement.

The CNN + fusion approach integrates a supervised decision fusion to aggregate the classification of each patch. This is based on the fact that the WSI global class information is not based solely on the most represented class but also on the spatial distribution of the various classes within the heatmap [73]. Several studies demonstrated that aggregating patch-level CNN predictions for WSI classification significantly outperforms patch-level CNNs with major voting [73,109]. Karimi et al. [15] employed three different CNNs with different input sizes to classify the prostate tissue. Then, a logistic regression model was used to determine the Gleason Score based on the predictions of the three CNNs. Respect to the single network, the authors achieved up to 10% improvement in accuracy during cancer grading. Campanella et al. [110] implemented a recurrent neural network (RNN) to aggregate the classified patches into the WSI label while Nagpal et al. [107] applied a nearest-neighbor classifier that used a summary of the heatmap to classify the entire whole-slide. In particular, the work of Nagpal et al. showed that the average accuracy of the deep learning system was 8% higher than that of a cohort of 29 pathologists.

As can be seen, the main post-processing method in prostate cancer detection is the patch aggregation to label the entire image (Table 2). Several strategies have been proposed to assign the class of a WSI, obtaining a performance improvement of up to 10% when compared to methods that do not employ any post-processing.

#### 4.3. Breast cancer

The most common cancer presenting a high mortality and morbidity among women worldwide is breast cancer [111]. In the histopathological analysis of breast cancer, pathologists analyze the overall tissue architecture and distribution of cells along with the nuclei density and organization. The nuclear pleomorphism and the spatial arrangement of cellular structures are generally assessed to distinguish between normal tissue, non-malignant (benign), and malignant lesions (in situ or invasive carcinoma) [26]. Manual examination of breast histopathological images calls for an intense workload and, on average, there is only a 75% diagnostic concordance between specialists [112]. This motivated the development of automated systems to increase the level of inter-observer agreement and improve diagnosis efficiency [113].

In the last few years, several deep learning pipelines have been proposed to perform breast cancer grading [27,114]. After the deep network training, all test images are classified using a patch-based approach. Recently, several post-processing strategies have been proposed to aggregate the patch-level classifier and generate the image-level classification [115,116]. Similarly to prostate cancer, we can divide these strategies into CNN + voting and CNN + fusion. In the CNN + voting approach, simple thresholding or major voting is applied to the classified patches to choose the label of the entire image. Cruz-Roa et al. [117] and Vandenberghe et al. [36] applied an empirical threshold on the CNN probability map to detect breast cancer. Kohl et al. [115] implemented a median filter to smooth the probability map and a small dilation of all the tumoral classes to slightly decrease the false-negative rate and slightly increase the size of tumor regions. Using this approach, a 1.5% improvement in accuracy was achieved compared to the baseline network (VGG). Kovalev et al. [118] and Litjens et al. [4] applied a fixed threshold followed by a connected component analysis of the heatmap. All components with a diameter smaller than a predefined value were removed to get rid of spurious detection caused by artifacts (tissue deformation and dust). One of the most common strategies in CNN + voting approaches is majority voting [69,119–121]. In this case, the image level label is decided by a majority voting on the labels of classified patches. Different network architectures, such as VGG [122],

**Table 3**  
Summary of the post-processing strategies for detection task.

Task	Reference	Year	Dataset	Method	Performance/Post-processing impact
Lymphocyte detection	Chen et al. [156]	2014	Breast (42 images)	CNN + non-maxima suppression algorithm	94.3% accuracy
	Janowczyk et al. [1]	2016	Breast (100 images)	CNN + iterative cleaning	96% true positive rate (TPR) and 87% positive predicted value (PPV). 6% TPR incrementation and 23% PPV increase respect to single network
	Garcia et al. [157]	2017	Gastro-intestinal tissue (45 images)	CNN + non-maxima suppression algorithm	96.8% accuracy
	Bidart et al. [153]	2018	Breast (92 WSI)	CNN + non-maxima suppression algorithm	94.6% accuracy. 7% accuracy improvement respect to SVM classifier
	Li et al. [154]	2018	Liver (5040 images)	CNN + morphological operators	93.7% accuracy
	Swiderska-Chadaj et al. [158]	2018	58 WSI from breast, colon and prostate	YOLO + non-maxima suppression algorithm	78% F1-score
	Rijthoven et al. [13]	2018	58 WSI from breast, colon and prostate	YOLO network + non-maxima suppression algorithm	74.7% F1-score. 3% F1-score incrementation with a speed-up of 4.3X respect to YOLO network
	Saltz et al. [151]	2018	5202 images from breast, colon, lung, prostate, skin and ovarian tissue	CNN + thresholding	0.95 AUC
Mitosis detection	Swiderska-Chadaj et al. [147]	2019	Breast (33 WSI); Colon (28 WSI); Prostate (22 WSI)	YOLO + non-maxima suppression algorithm	79% F1-score. 8% F1-score increase compared to the baseline
	Wang et al. [170]	2014	Breast (73 images)	Cascade ensemble handcrafted + CNN-derived features	73.5% F1-score. 4.81% F1-score improvement compared to single network
	Albarqouni et al. [163]	2016	Breast (23 WSI)	Multiscale decision aggregation using three different CNNs	74.2% F1-score. 22% F1-score improvement compared to the single-scale CNN
	Albayrak et al. [166]	2016	Breast (10 WSI)	CNN + blob analysis	96.8% accuracy
	Chen et al. [162]	2016	Breast (50 images)	CNN + local maximum analysis	79% F1-score. 13.1% F1-score increase compared to the baseline
	Chen et al. [169]	2016	Breast (50 images)	CNN + second deep network for mitosis detection	78.8% F1-score
	Janowczyk et al. [1]	2016	Breast (311 images)	CNN + thresholding	54% F1-score
	Zerhouni et al. [160]	2017	Breast (107 images)	CNN + majority voting	64.8% F1-score
	Akram et al. [165]	2018	Breast (656 images)	CNN + non-maxima suppression algorithm	64.2% F1-score. 16% F1-score increase compared to the baseline
	Li et al. [172]	2018	Breast (1746 images)	FASTER-RCNN + non-maxima suppression algorithm	76.8% F1-score
	Rao et al. [171]	2018	Breast (4638 images)	FASTER-RCNN + non-maxima suppression algorithm	95.5% F1-score
	Saha et al. [164]	2018	Breast (2762 images)	CNN + handcrafted features	90% F1-score. 14% F1-score improvement respect to the single CNN
	Beevi et al. [168]	2019	Breast (92 images)	CNN + active contour model	88.6% F1-score
	Li et al. [173]	2019	Breast (1634 images)	CNN + morphological filtering	80.21% F1-score. 9.27% F1-score improvement compared to the baseline
	Wahab et al. [167]	2019	Breast (500 WSI)	CNN + thresholding	71.3% F1-score.
Mahmood et al. [174]	2020	Breast (1746 images)	FASTER-RCNN + textural features	73.1% F1-score. 7% F1-score improvement compared to the baseline	

ResNet [123] and Inception [92], were combined with major voting to perform breast cancer classification. Roy et al. [124] considered all image patches as one unit and assigned to the image the class that the maximum number of patches presented, ignoring many misclassified patches. In this way, compared to the fixed threshold proposed in Ref. [117], the accuracy in image-wise classification was improved by 2.5%. Araújo et al. [116] employed a VGG-like network followed by major voting to obtain the final image label from the individual patch classifications. Ahmad et al. [89] adopted the same strategy but with a deeper network (ResNet), thus achieving a 7.2% increase in accuracy. Finally, all the patch predictions were averaged by Rakhlin et al. [94] and the image-level class was defined using the maximum probability score.

The CNN + fusion approaches employ more sophisticated strategies to assign a label to the entire image. In fact, the majority of patch-based classification algorithms (i.e., major voting) predict the final image label without taking into consideration the spatial distribution of patches

within the WSI. However, the probability of a patch to be cancer-positive is correlated with its surrounding patches [125]. To incorporate this information into a deep learning pipeline, several aggregation approaches have been proposed as post-processing steps [126,127]. Le et al. [126] took into consideration the labels and characteristics of neighboring patches by computing an aggregation operation on nearby patches within a specific distance of the considered patch, which was then used to compute the final classification probability value. By integrating this post-processing method within their pipeline, they achieved a 4% improvement in the positive predictive value (PPV). Wang et al. [40] extracted 28 morphological and geometrical features for each heatmap of the training set. Then, a random forest classifier was built using these features to discriminate between malignant and benign WSI. Liu et al. [78] and Nazeri et al. [128] employed an ensemble model to obtain the image-level classification. Since histopathological images do not have a canonical orientation, the authors applied for each patch the left-right rotation and rotations to obtain predictions of 8 different

orientations. All the 8 predictions were averaged to obtain a robust rotation-invariant classifier. This approach allowed to improve the image-wise accuracy by 5% compared to major voting [128]. Similarly, Vang et al. [129] first evaluated the class-map for the 8 different orientations separately and then computed a classes histogram across all 8 orientations. The histogram data was then used to train a logistic regression classifier to detect different cancer subtypes. This post-processing strategy achieved a 6% improvement in accuracy compared to the baseline score [116]. Couture et al. [127] aggregated the probability of each patch into a quantile function and used an SVM classifier to predict the class of the whole image. This post-processing allowed to obtain an 80% accuracy in breast cancer grading. Finally, Yan et al. [98] employed two different deep networks to perform image-wise classification. First, the entire image is divided into 12 small patches and then feature representations are extracted from each patch with an Inception network. Then, the 12-feature vector (one for each patch) is used to train a second deep network to obtain the final image classification.

The main post-processing method employed in deep learning frameworks for breast cancer classification is patch aggregation (Table 2). The most common strategy is the CNN + fusion approach, where a classifier is employed to assign the image label starting from the CNN heatmap. Compared to baseline methods, this approach achieved between a 2.5% and 6% improvement in classification accuracy.

#### 4.4. Other tumors

Different post-processing strategies have also been integrated within deep learning frameworks for the analysis of other histological tumors [130,131]. For example, one of the currently most studied histological tumors is lung carcinoma [132], which is the leading cause of cancer death in the western world [133]. Treatment for lung cancer is based on the grade and stage of the tumor. Tumor size, single nodules classification, and the presence of metastasis have shown a very important prognostic relevance during the assessment of lung cancer [134]. Coudray et al. [130] proposed an Inception network followed by a patch aggregation approach. Using this strategy, the per-tile results were aggregated on a per-slide basis by averaging the probabilities obtained on each patch. Wei et al. [132] combined a ResNet with an ad-hoc heuristic pipeline for the classification of lung carcinoma patterns. Once generating the heatmap for a generic WSI, all the patch predictions with low confidence (e.g., lower than a predefined threshold) were discarded. Subsequently, the predominant label was assigned as the most frequent class, and minor labels were assigned as the remaining cancerous patterns. The proposed model was robust to tissue staining artifacts and single-patch misclassifications, thanks to discarding low confidence predictions and aggregating over a large number of patches. Graham et al. [135] employed a ResNet with two different post-processings: majority voting and random forest. Majority voting simply assigned the class of the WSI as the one with the largest number of positive patches in its corresponding probability map. For the random forest model, several morphological and statistical features were extracted from the heatmap to train the classifier and assign the overall WSI classification. The ResNet with random forest obtained the highest performance compared to majority voting, with a 3% accuracy increase. Wang et al. [35] proposed different patch aggregation methods to classify lung WSI, ranging from simple voting approaches to random forest models. The best performing method combined the deep features extracted from each patch into a global descriptor vector. Then, the feature vector was fed into a random forest classifier for WSI-level prediction. This approach provided an effective holistic representation of the entire WSI, allowing for a 27.4% improvement in accuracy over majority voting. Finally, Li et al. [136] implemented a deep model followed by conditional random fields (CRFs) for lung cancer detection. CRFs were adopted for noise elimination and boundary smoothing of the tumor contour. Using this strategy, the authors obtained an 11%

accuracy improvement compared to the single deep network.

Another commonly studied tumor with histological analysis is colon cancer. Colon cancer is the fourth most common cause of cancer death overall (after lung, stomach, and liver) and it represents the second most common cause of cancer in women and the third most common cause in men [137,138]. Korbar et al. [131] proposed a colorectal polyp classification on WSI using deep learning. Using majority voting, the most common colorectal polyp class among the associated patches was used to label the entire WSI. Sirinukunwattana et al. [139] proposed a post-processing based on neighbors to classify colon cancer histology images. The authors proposed a neighboring ensemble predictor to be used in conjunction with a standard CNN. Based on spatial ensembling, this predictor leveraged all relevant patch-based predictions in the local neighborhood of an extracted tile, which in turn produces more accurate classification results than its single-patch based counterpart.

Xu et al. [140] and Hou et al. [73] proposed different strategies for brain tumor detection in histopathological images. In particular, Xu et al. [140] combined a deep convolutional network with an SVM classifier to obtain the final WSI prediction. Hou et al. [73] employed different post-processing strategies to aggregate the patch-level classification into the WSI label. Their CNN was combined with majority voting, SVM and a logistic regression model for brain glioma classification. The authors achieved the best results with the logistic regression model, obtaining an accuracy increase of 3.5% compared to a simple majority voting and an increase of 5.6% compared to the SVM.

Also in cases of other cancers such as lung or colon cancer, the main post-processing strategy is the aggregation of the patches using the CNN softmax followed by a classifier (Table 2). This strategy allows to increase the performance of the DL model by up to 27% compared to other techniques.

## 5. Post-processing algorithms for detection problems

### 5.1. Deep networks for detection

Image classification networks classify entire images or patches into only a single category, that typically corresponds to the most salient object. However, assigning a single label per image presents numerous drawbacks, such as not allowing a spatial localization of the object and precluding important information about the number of objects within the image. Hence, object detection network models are more appropriate to locate and identify multiple relevant objects within the same image. A generic object detection performs an instance segmentation, where each existing object is located and labeled with rectangular bounding boxes (Fig. 1b). In this review, any method that takes the potential objects as determined by the network and further elaborates them to provide the final result is considered as a “post-processing” technique.

The performance of deep networks for detection tasks is generally assessed by calculating the F1-score. The F1-score is a measure of accuracy and it is calculated as the harmonic mean between precision and recall [141]. A deep network for object detection can follow two different approaches: region proposal and regression/classification (Fig. 6). *Region proposal* frameworks follow a two-step process, in which a traditional object detection pipeline initially generates all region proposals and then classifies each proposal into different categories of objects using a deep network. *Regression/classification* frameworks regards object detection as a regression or classification problem, adopting a unified framework to achieve final results (categories and locations) directly.

In the last few years, several CNNs have been released for object detection challenges [3]. The *R-CNN* model [142] combines the selective search method [143] to detect region proposals and a deep network to find the object in these regions. The selective search generates around 2000 region proposals using bottom-up grouping to reduce the searching space in object detection. Then, high-level features are extracted

**Table 4**  
Summary of the post-processing strategies for segmentation task.

Task	Reference	Year	Dataset	Method	Performance/Post-processing impact	
Nuclei segmentation	Song et al. [186]	2015	Breast (53 images)	CNN + graph partitioning model	95% Dice score	
	Pan et al. [183]	2017	Breast (58 images)	CNN + morphological operators	83.9% F1-score	
	Xing et al. [187]	2015	Brain (31 images); Breast (35 images); Pancreas (22 images)	CNN + deformable model	81% F1-score	
	Chen et al. [188]	2017	Brain (33 images)	CNN + morphological operators	87.6% Dice score	
	Kumar et al. [180]	2017	30 images from bladder, breast, colon, kidney, liver, stomach and prostate	three-class CNN + region growing	76.2% Dice score. 6.89% Dice score increase compared to traditional two-class CNN	
	Naylor et al. [185]	2018	Breast (55 images); 25 images from bladder, colon, kidney, liver, stomach and prostate	UNET + distance transform	82.3% Dice score	
	Cui et al. [190]	2019	Breast (77 images); 25 images from bladder, colon, kidney, liver, stomach and prostate	three-class CNN + connected component analysis	81% Dice score. 5% Dice score increase compared to the baseline	
	Sornapudi et al. [184]	2018	Cervical tissue (133 images)	CNN + superpixel approach	98.2% Dice score	
	Xie et al. [71]	2018	Brain (15 WSI); 30 images from bladder, breast, colon, kidney, liver, stomach and prostate	MASK R-CNN + watershed transform	90.4% Dice score. 2.06% Dice score increase compared to the baseline	
	Jung et al. [72]	2019	Breast (38 images); 25 images from bladder, colon, kidney, liver, stomach and prostate	MASK R-CNN + multiple inference	81.2% Dice score. 11% Dice score improvement respect to the baseline	
	Koyun et al. [182]	2019	670 images from breast, colon, kidney and prostate	CNN + thresholding	81% F1-score. 10% F1-score increase respect to the single network	
	Mandloi et al. [181]	2019	670 images from breast, colon, kidney and prostate	CNN + thresholding	86% F1-score	
	Zeng et al. [191]	2019	30 images from bladder, breast, colon, kidney, liver, stomach and prostate	three-class UNET + thresholding	82.7% F1-score. 2% F1-score increase respect to the standard UNET model	
	Wan et al. [189]	2020	243 images from breast and lung	CNN + concave point detection algorithm	80.75% Dice score. 8.7% Dice score improvement respect to the baseline	
	Xie et al. [192]	2020	95 images from bladder, breast, colon, kidney, liver, stomach and prostate	three-class CNN + watershed transform	88.6% F1-score. 17% F1-score increase respect to the baseline	
	Tubule/Glands segmentation	Bentaieb et al. [198]	2016	Colon (165 images)	CNN + thresholding	80% Dice score
		Chen et al. [193]	2016	Colon (80 images)	Two CNNs + morphological operators	81.3% Dice score
		Ren et al. [201]	2017	Prostate (22 images)	CNN + borders mirroring	84.6% F1-score. 1% precision improvement compared to the single network
Xu et al. [196]		2017	Colon (165 images)	Ensemble of three CNNs	86.8% Dice score	
De Bel et al. [199]		2018	Kidney (24 WSI)	UNET + connected component analysis	73.6% Dice score	
Graham et al. [203]		2018	Colon (378 images)	Two CNNs + thresholding	87% F1-score	
Tang et al. [197]		2018	Colon (165 images)	CNN + thresholding	87.2% Dice score	
Van Eyck et al. [195]		2018	Colon (165 images)	CNN + connected component analysis	84.1% Dice score	
Binder et al. [204]		2019	Breast (25 images); Colon (165 images)	CNN + morphological operators	84.5% Dice score. 3% Dice score improvement compared to the baseline	
Qu et al. [202]		2019	Colon (165 images)	Three-class CNN + connected component analysis	92.4% F1-score	
Rezaei et al. [200]		2019	Colon (165 images)	CNN + morphological operators	84.4% Dice score	
Ding et al. [205]		2020	Colon (165 images)	Three-class CNN + morphological operators	88.2% F1-score	

from each region using a pre-trained CNN and they are fed into a simple classifier (e.g., SVM) employed to recognize each known class. However, the whole detection framework could not be optimized in an end-to-end manner, making it difficult to obtain a global optimal solution. *FAST R-CNN* [144] and *FASTER R-CNN* [145] are the evolution and the extension of *R-CNN* networks. They were designed to speed-up both training and testing time of traditional *R-CNNs* by introducing a CNN in the region proposal pipeline. For regression/classification approaches, a widely used network is the YOLO model [146]. This architecture employs a single end-to-end trained neural network that takes the image as input and directly provides the location and class labels for each bounding box. The following sections go into specific detail on how these networks have been integrated within deep learning frameworks for histological structure detection. Table 3 summarizes all the detection approaches described in this section, along with their post-processing strategy and the database used for their validation.

## 5.2. Lymphocyte detection

Lymphocytes are a subtype of white blood cells and play an important role in the immune system, where an immune response is characterized by lymphocytic infiltration, in which the lymphocyte density greatly increases at sites of disease or foreign bodies [1]. Accurate detection and assessment of lymphocyte presence in cancer can potentially allow for the design of new biomarkers to help monitor the rapid progression of a tumor [147], as recent studies have shown how the lymphocytes can serve as a fundamental biomarker to predict clinical outcomes and treatment response [148,149]. Lymphocytes present a similar appearance to cell nuclei in terms of hue, often making them difficult to be differentiated. Typically, however, lymphocytes tend to be circular, more chromatically dense, and smaller [1]. In recent years, several automated tools have been developed to localize and quantify the density of these immune cells with the aim to predict the presence



and development of metastases and the overall survival of the patient [147].

The most common strategy to detect lymphocytes in histological images is the segmentation through object detection which output bounding boxes around objects of interest (cells). These methods implement a CNN as a two-class (e.g., lymphocytes, background) classifier to detect these cells in a patch-wise manner using a sliding window [150]. By applying the CNN across an entire image, a heatmap is generated that indicates the probability of each pixel being a lymphocyte (probability map). Still, the obtained heatmap does not allow for a precise localization of the lymphocytes and the results are often unsatisfactory when there is a cluster of cells very close to each other. Hence, it is necessary to implement post-processing techniques to correctly identify the center of each lymphocyte.

Janowczyk et al. [1] applied a convolution with a disk kernel to the probability map in order to highlight the center of each lymphocyte. In this method, the highest point in the probability map is taken as the center of a lymphocyte and the radius is cleared. This is done in an iterative manner, and the considered radius is the same dimension as a typical lymphocyte. The dimension of the radius is of critical importance, as the radius needs to be large enough to ensure that there is only one prediction per nucleus, but at the same time also small enough so as to not suppress a neighboring cell's nucleus. This technique prevents multiple centers from being recognized for the same lymphocyte and allowed for a 6% TPR (true positive rate) incrementation and a 23% PPV (positive predictive value) increase. On the other hand, Saltz et al. [151] applied a simple threshold to the heatmap to locate each lymphocyte. Then, local spatial features [152] were extracted from the detected cells to associate the lymphocytes' infiltration with molecular readouts and clinical outcomes. Bidart et al. [153] chose to employ a non-maxima suppression (NMS) algorithm to locate the center of each lymphocyte. The NMS algorithm, similarly to the method by Janowczyk et al. [1], is based on the assumption that the points in the heatmap with the highest probability correspond to the center of the cells. Similarly to Ref. [1], the pixels with the highest cell probability are found in an iterative manner and then the pixels found within a determined radius  $r$  are set to zero. In this application, the optimal value of  $r$  was found empirically by looking at the distribution of the distance from a nucleus to its closest neighbor. Compared to a SVM classifier, this method was able to achieve an improvement of 7% when considering accuracy, and a 13% sensitivity and 6% specificity improvement. Finally, Li et al. [154] adopted a region proposal framework to detect the lymphocytes by combining a CNN with a dual morphological operation and distance transform. Dual morphological grayscale reconstruction [155] was adopted to highlight the cell from the background and a H-maxima transform after distance transform was employed to locate the center of each candidate cell. Then, a CNN was used to classify each region proposal in two classes: lymphocyte and non-lymphocyte.

Several strategies have also been proposed for the detection of lymphocytes in immunohistochemical (IHC) images. Chen et al. [156] proposed a CNN combined with color deconvolution to locate these cells. Sparse color unmixing was performed to separate the image into DAB and hematoxylin channels using Ruifrok color deconvolution [52]. Then, the heatmap of the lymphocytes' location was obtained using a patch-wise classification on the DAB channel and a non-maxima suppression algorithm was used to yield the final detection. A similar approach was adopted by Garcia et al. [157], where an NMS algorithm was employed after a CNN to locate the immune cells in gastric cancer IHC images. Swiderska-Chadaj et al. [158] employed a YOLO architecture, where during inference, predicted bounding boxes with an overlap are considered as detecting the same lymphocyte using a non-maxima suppression algorithm. Rijthoven et al. [13] proposed a modified version of the YOLO model for the detection of the lymphocytes in WSI's of colon, breast and prostate cancer. In this architecture, named YOLLO (You Only Look on Lymphocytes Once), the grid cell used for prediction was forced to be  $32 \times 32$  pixels and the number of convolutional layers

was reduced from 23 to 8 in order to simplify the entire model. The proposed modifications, namely guided sampling strategy and simplified architecture, allowed to increase the detection performance of 3% and gain in speed of up to 4.3 times quicker during inference time compared to the traditional YOLO model. In another work, Swiderska-Chadaj et al. [147] adopted a YOLLO network followed by non-maxima suppression. The proposed method allowed to distinguish well the cells in clusters, achieving an improvement in F1-score of 8% compared to a simple CNN.

As can be seen, the most common post-processing method in lymphocyte detection is the non-maxima suppression algorithm (Table 3). Starting from the CNN heatmap, this technique allows to accurately detect the location of each lymphocyte and has shown an improvement in detection performance up to 8% compared to methods that do not employ any post-processing.

### 5.3. Mitosis detection

Mitosis is a process of duplication where a single cell divides into two genetically identical daughter cells. Precise quantification of mitotic figures is one of the most important prognostic factors in cancer grading as it gives an assessment of the tumor proliferation [159]. However, mitotic count is time-consuming and difficult due to the variations in morphological appearance of mitotic cells. These variations are caused by various factors including the non-uniform stain variation, mitotic phase, irregular illumination and tissue damage during the slide preparation (Section 2).

In recent years, there has been a growing interest in the development of fully automatic solutions for an accurate and precise quantification of the mitotic activity [1,160]. Ciresan et al. [161] proposed a CNN as a pixel-wise classifier to detect mitosis with a sliding window. However, this approach was very computationally expensive, making the algorithm not practical in clinical settings. To overcome this limitation, several methods proposed a patch-wise approach [1,162–164]. These methods employ a CNN on the entire image that outputs a heatmap indicating each pixel's probability of belonging to a cell in mitosis. In order to correctly locate all the mitotic events, different post-processing methods have been applied to the probability map. Janowczyk et al. [1] convolved the heatmap with a kernel disk and identified a mitotic event as those image locations that were above a certain probability threshold. Chen et al. [162] computed the local maximum of the heatmap while Akram et al. [165] employed the NMS algorithm, with a distance of 50 pixels, to remove duplicates. These strategies achieved an improvement of the F1-score equal to 13.1% and 16%, respectively, when compared to a CNN without any post-processing. Saha et al. [164] proposed a deep architecture model reinforced with handcrafted features. The combination of handcrafted (HC) features with the high-level ones extracted by the CNN allowed to increase the overall accuracy of the classifier up to 14%. Zerhouni et al. [160] adopted a majority vote to perform pixel clustering inside the heatmap. All the clusters containing fewer than 30 pixels were discarded, and the final prediction was taken as the centroid of each remaining cluster. Finally, Albarqouni et al. [163] implemented a multi-scale CNN for mitosis detection, combining the output probability of three different CNNs. The input image was downsampled to different scales (i.e. 0.33, 0.66 and 1) and three different networks were trained for each downsampled scale. Then, during inference time, the final positive response was obtained as the average of the three output probabilities from each single CNN. Using this multiscale approach, an overall improvement in the F1-score of 22% was observed compared to the single-scale CNN.

Given the fact that a patch-wise approach can be computationally demanding and time-consuming, some authors proposed a two-step strategy, by first detecting all nuclei (*detector step*), and then classifying each nucleus separately as mitotic or non-mitotic (*discriminator step*). Different techniques have been applied to detect candidate nuclei, ranging from k-means and blob analysis [166], blue-ratio binary

thresholding [167], active contour models [168] to deep neural networks [169]. Wang et al. [170] applied a Laplacian of Gaussian filter followed by a fixed threshold to identify all the candidate nuclei. Then, a cascade ensemble of CNNs and handcrafted features were adopted for mitosis detection. This nuclei sampling strategy, along with the combination of CNN and HC features, allowed to both reduce the computational time at test time and to obtain a 4.81% improvement of mitotic form detection. Rao et al. [171] and Li et al. [172] proposed a modified version of the FASTER-RCNN tuned for mitosis detection, with comparable speed of previous CNN models and more accurate localization. Small bounding boxes were discarded, and region proposals were further refined using an NMS algorithm with fixed threshold as post-processing. In another work, Li et al. [173] proposed a more refined post-processing to their strategy. Starting from the heatmap, the mitotic cells were found using a heuristic method. Firstly, a smoothing and a binary processing were applied to the probability map, aiming to yield the detected blobs. Then, a morphological filtering step based on the confidence score and area of the segmented objects was employed to delete false-positive shapes. This strategy was able to obtain up to a 9.27% improvement of the F1-score with respect to previously published methods. In a recent work, Mahmood et al. [174] employed a FASTER-RCNN followed by a post-processing based on textural features. First-order statistical features, local binary pattern (LBP) and histograms of oriented gradients (HOG) were adopted to reject false positive shapes detected by the deep network. This strategy led to an improvement in the F1-score of 7% compared to the single CNN.

Similarly to lymphocytes detection, the clustering strategies applied to the CNN heatmap (NMS, majority voting, local maxima) are the main post-processing methods employed in mitosis detection (Table 3). Compared to the single CNN, these techniques have shown an accuracy improvement of up to 17%.

## 6. Post-processing algorithms for segmentation problems

### 6.1. Deep networks for segmentation

The task of segmentation is the requirement of delimiting a precise boundary of the desired objects (e.g., histologic primitives such as nuclei, tubules, epithelium, etc). This is done so that accurate morphological features can subsequently be extracted from the segmented object. Detection tasks (i.e., mitosis and lymphocyte detection) differ from segmentation tasks in that in order to detect an object, precise boundary or contour determination is not necessary, as the goal is typically to only identify the center of the region of interest. For this reason, segmentation typically tends to be more challenging than detection since each pixel is classified into instances, each instance (or category) corresponding to an object of the image. A general segmentation pipeline performs a semantic segmentation, where each pixel of the image is classified into meaningful classes of objects (Fig. 1c). In this review, any method that further processes the semantic segmentation determined by the network to provide the final result is considered as a “post-processing” technique.

The performance of deep networks for segmentation problems is generally assessed by calculating the F1-score and the Dice score. The Dice score measures the spatial overlap between two binary shapes [175]. A semantic segmentation architecture can be thought of as an encoder network followed by a decoder network. The encoder network is typically a pre-trained CNN such as ResNet/VGG designed to extract high-level features from the input image. The decoder network aims to semantically project the discriminating characteristics (lower resolution) learned by the encoder on the pixel space (higher resolution) to obtain a dense classification [176]. Basically, the idea is to scale up, the scale down effect made by all the encoder layers. Based on the decoding mechanism, a semantic segmentation network can follow two approaches: region-based segmentation or fully convolutional network-based segmentation. *Region-based* semantic segmentation

follows the “segmentation using recognition” pipeline, which first extracts the regions containing the object. Then, region-based predictions are transformed into pixel predictions, generally by labeling each pixel within the ROI. On the other hand, *fully convolutional network* (FCN) semantic segmentation learns a mapping from pixels to pixels, without extracting the region proposals [177].

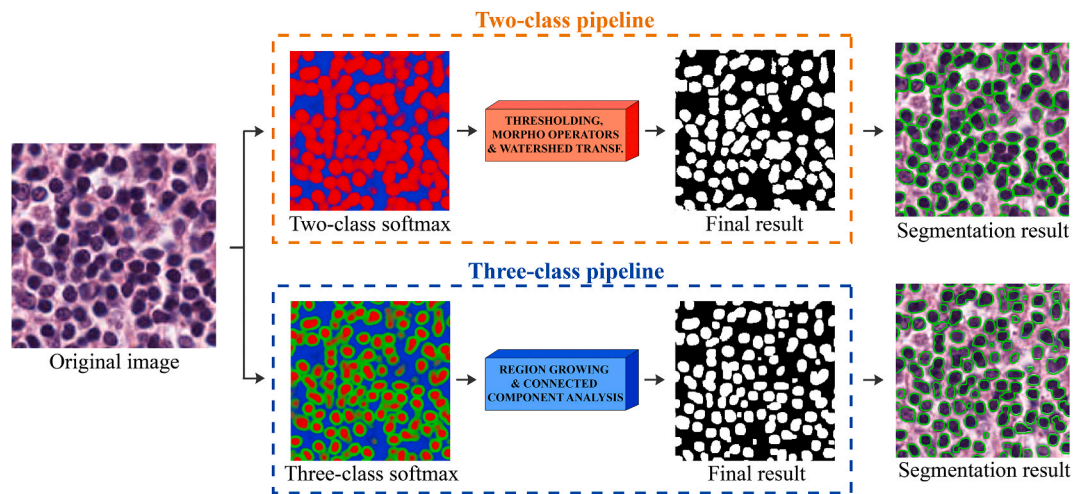
In the last few years, two deep architectures have become popular for medical image segmentation: MASK R-CNN and UNET. MASK R-CNN is an evolution of the FASTER R-CNN architecture (Section 5.1) specifically designed for pixel-level segmentation [178]. The MASK R-CNN is a R-CNN with three output branches: the first one computes the bounding box coordinates, the second one computes the associated class and the last one computes the binary mask to segment the object. The particularity of the MASK R-CNN model is its multi-task loss combining the losses of the bounding box coordinates, the predicted class and the segmentation mask. The model tries to solve complementary tasks leading to better performances on each individual task. The UNET model was proposed by Ronneberger et al. [5] and it was specifically designed for biological microscopy images. This architecture is composed in two parts: a contracting part to compute features (downsampling) and an expanding part to spatially locate patterns within the image (upsampling). The downsampling subnet has an FCN-like architecture that extracts features with convolutional layers while the upsampling part uses up-convolution to reduce the number of feature maps while increasing their height and width. Cropped feature maps from the downsampling part of the network are copied within the upsampling part to avoid losing pattern information. The following sections will describe in depth how these segmentation networks have been employed to comprehend the spatial relationships between histological structures within the image. Table 4 summarizes all the segmentation approaches described in this section, along with their post-processing strategy and the database used for their validation.

### 6.2. Nuclei segmentation

Accurate nuclei segmentation is a crucial step in cancer analysis and grading [68]. During cancer diagnosis, pathologists analyze biopsies to make prognostic and diagnostic assessments, mainly based on the nuclei morphology and their spatial arrangement. In this context, an automated algorithm could assist the pathologist to obtain reliable and quantitative statistics about cell morphology. However, the automatic segmentation of cell nuclei is a challenging task due to the extremely variable shapes and sizes of overlapping nuclei, as well as weakly defined boundaries and different staining methods. Nowadays, the current challenge is to precisely define cell boundaries or/and divide overlapping nuclei [179].

To solve this problem, several algorithms based on deep learning have been proposed to obtain an accurate segmentation of nuclei in histopathological images [1,180]. These strategies can be grouped into two categories: *two-class pipeline* and *three-class pipeline* (Fig. 7). In the two-class pipeline, the CNN is employed for binary segmentation (nuclei vs background) while the deep network also estimates the cell boundaries in the three-class pipeline.

In two-class pipelines, authors employed a simple fixed threshold on the CNN softmax to detect the nuclei boundaries [181,182]. Pan et al. [183] implemented a series of morphological operations as post-processing to improve the segmentation performance. After the thresholding, morphological cleaning and hole filling were applied to reduce errors due to image artifacts and background clutters. Hence, all regions detected with an area less than a predefined value were eliminated as they were considered too small to be cell nuclei. Sornapudi et al. [184] combined a superpixel approach with CNN binary segmentation (nuclei vs background) to perform cell segmentation. Their approach required a reduced memory when compared to pixel-wise approaches and also reduced the number of parameters to be tuned, thanks to the superpixel (i.e., a group of similar pixels) classification.



**Fig. 7.** Post-processing strategies for segmentation tasks. Cell nuclei segmentation is used as an explanatory example. Two-class pipeline: a CNN is employed for binary segmentation (object vs background). Then, traditional techniques (morphological operators, watershed transform) are employed to refine the segmentation. Three-class pipeline: a deep network is implemented to segment both objects and their contours. Then, more sophisticated strategies such as connected component analysis are employed to perform individual object detection.

However, these approaches do not completely solve the problem of clustered and overlapping nuclei.

Recently, more sophisticated post-processing techniques were proposed to perform individual nuclei detection [71,185]. The framework proposed by Xie et al. [71] adopted a stain normalization followed by MASK R-CNN and watershed as post-processing. The watershed transform was used to separate touching cells. Using this post-processing, the Dice score was improved by 2.06% with respect to the single network. Song et al. [186] proposed a CNN followed by a graph partitioning model to refine the nuclei segmentation. The method consisted of three parts: 1) CNN to obtain feature representation and a preliminary pixel-level segmentation; 2) superpixel and graph cut to accurately segment the nuclei boundaries; 3) marker-based watershed to separate clustered nuclei. Xing et al. [187] applied an optimized post-processing on the CNN probability map to detect individual nuclei. In particular, a selection-based sparse shape model and local repulsive deformable model combination was used as a segmentation algorithm. This specific deformation model effectively segmented nuclei with either weak or missing boundaries. Jung et al. [72] employed a MASK R-CNN followed by multiple inference as post-processing to boost segmentation performance. A total of seven augmented images including the original image were generated and used as the input for multiple inference. Then, majority voting at the pixel-level was performed on the seven images and all the pixels with a score higher than 50% were selected as the final segmentation result. This strategy allowed to increase the F1-score by 3.3% and the average Dice score by more than 11%. Naylor et al. [185] proposed a post-processing method based on distance map to handle the issue of touching objects. The authors employed a UNET model to predict the distance transform of the cells instead of directly predicting the nuclei binary mask. Then, local maxima were founded, and a simple thresholding operation was employed to obtain object pixels. Chen et al. [188] proposed a novel deep contour-aware network to cope with the issue of merged nuclei. This network combined nuclei foreground and edge information to obtain instance segmentation results. Post-processing steps including smoothing, disk filtering and hole filling were performed to remove small spurious objects. Recently, Wan et al. [189] proposed a CNN followed by a concave point detection algorithm to accurately segment highly overlapping nuclei. This post-processing improved the performance in separating clustered and touching nuclei by more than 8% (Dice score) compared to simple thresholding.

Some authors have tried to solve the problem of overlapping nuclei by designing CNNs that predict both objects and their contours (i.e.

three-class output: inside, contours, background) [180,190]. Kumar et al. [180] employed a three-class CNN to segment the cell nuclei. A region growing on the inside probability map was initialized as post-processing. Seeds were found by thresholding the inside class map at 0.5. While the seeded region grows, the average boundary class probability of its contour pixels increases, while their average inside class probability decreases. The nuclei are stopped from growing when the average boundary class probability of the pixels on the border of an area reaches a local maximum. This approach achieved a 6.9% Dice score increase when compared to a simple two-class CNN. Cui et al. [190] proposed a nuclei-boundary model to was applied to each connected component to recover the shape. Using this post-processing, an improvement of 5% of the Dice score was observed. Zeng et al. [191] employed a three-class UNET model to segment cell nuclei. Since the predicted nuclei had many overlapping cells, a post-processing technique was implemented to refine the segmentation mask. Firstly, the inside and contour probability maps were thresholded at 0.5. The contour mask was subtracted from the inside mask and each cell was dilated using a disk template of 3 pixels-radius. This processing resulted in an increase of 2% of the F1-score compared to the standard UNET model. Finally, Xie et al. [192] designed a network with multiple segmentation tasks for learning the foreground, marker, and interval of nuclei, simultaneously. The foreground result is then refined using logical operators thanks to the learned interval between overlapping nuclei. Then, the touching nuclei were split thanks to a marker-controlled watershed algorithm using the learned marker result and nuclei segmentation. Compared to other deep networks and different post-processing methods, this strategy achieved an increase in the F1-score of up to 17%.

As can be seen, the main issue when segmenting nuclei is cell separation (Table 4). The main post-processing methods involve the use of a third CNN class (boundary mask) to carefully separate the touching nuclei. These strategies allow to improve performance up to 11% compared to methods that do not employ any post-processing.

### 6.3. Tubules and glands segmentation

A typical tubule/gland is composed of a lumen area surrounded by a ring of epithelial cells. In later stages of cancer, the tubule regions become massively disorganized [193]. Tubule and gland morphology is routinely used by expert pathologists to assess the cancer malignancy degree in several epithelial tissues such as prostate, breast and colon [194]. In order to obtain reliable morphological statistics for a

quantitative diagnosis, the tubules/glands must be accurately segmented in histology images [1]. However, this task is non-trivial due to the large variability in glandular morphology as well as the existence of touching and poorly defined structures in pathological conditions.

In the last few years, several deep learning methods have been proposed for the gland segmentation task from pathology images [195, 196]. Tang et al. [197] and Bentaieb et al. [198] proposed an encoder-decoder CNN to segment colon glands in histopathological images. Starting from the softmax, the authors applied a global thresholding to segment all the glandular regions. De Bel et al. [199] employed a similar approach to segment the renal tubules. The CNN softmax was first thresholded at 0.9 likelihood and then a connected component analysis was employed to remove objects smaller than 300 pixels. Rezaei et al. [200] applied more robust post-processing methods to segment the glands contours. Firstly, the Otsu method was used to estimate the best threshold and to produce the segmentation binary mask from the probability map. Then, the final solid segmentation was obtained by using morphological operations on the binary mask for denoising and filling holes. Ren et al. [201] implemented an encoder-decoder network to segment the prostate glands. Since the semantic segmentation can be less accurate near image borders, the authors proposed a post-processing step to retain the boundary information during inference time. Specifically, a mirror border of 320 pixels was synthesized in each direction and the CNN was applied in a sliding window fashion. Only the center of each output image was used to form the seamless segmentation mask. Using this strategy, the global precision was improved by 1% with respect to the single segmentation network. In order to separate touching glands, Qu et al. [202] proposed a three-class CNN that predicted both the contour and the glandular inner regions simultaneously. Starting from the three-class segmentation map, the final segmentation was obtained by connected component labeling, removing small area objects, and dilating with a disk filter.

Xu et al. [196] tried to solve the issue of touching glands by combining three different CNNs. The first network was designed for foreground segmentation, the second was optimized for edge detection while the third was employed for the detection of individual glands. The features generated by the three deep networks were concatenated by a CNN that produced the segmented instances. The combination of regional, boundary, and location information allowed to accurately split and segment the colon glands. Chen et al. [193] proposed a novel deep contour-aware network that both depicted the gland object contours and output segmentation probability maps. To separate touching glands and output the final segmented object and contour masks, features maps from hierarchical layers were upsampled with two different branches. Hole filling, smoothing, and small area removal were applied to the segmentation result and each connected component was labeled with a unique value for representing one segmented gland. A similar approach was followed by Graham et al. [203] for colon glands segmentation. The authors applied a threshold of 0.5 to all predicted probability maps. Then, a morphological opening operation is used with a disk filter radius 5 to obtain the final result. Binder et al. [204] employed a deep network for multi-organ gland segmentation. After stain normalization, the histological image was fed into the CNN, obtaining two probability maps: contour and inner regions. Both of these maps were then thresholded with two present threshold values, resulting in two binarized masks relative to the contour and glands. Afterward, the contour mask was subtracted from the binarized gland mask to separate overlapping glands, so that individual glands were accurately identified. Finally, the gland mask was dilated using a disk element with a radius equal to the thickness of the contour mask to retrieve the boundary information. After this post-processing, an improvement of the Dice score of 3% and a reduction of the Hausdorff distance equal to 20% was observed. Recently, Ding et al. [205] proposed a three-class CNN (background, gland object, gland boundary) followed by an ad-hoc post-processing. In particular, the probability map of the gland boundary was subtracted from the gland's interior probability map. Then, segmented glands were

obtained by thresholding the resulting probability map with a fixed value of 0.8. Finally, a morphological dilation using a disk template of 5 pixels-radius was used to obtain an accurate gland contour.

As can be seen, the main post-processing methods in tubules and glands segmentation are the morphological operators applied on a three-class CNN to precisely define tubules and glands boundaries (Table 4). These approaches have shown a performance improvement of up to 20% when compared to methods that do not employ any post-processing.

## 7. Discussion

This paper aims to provide an overview of the main pre and post processing techniques adopted in deep learning frameworks in digital pathology. Deep learning algorithms, in particular convolutional networks, have rapidly become the main methodology for analyzing medical images. However, it is not trivial to manage the network prediction errors as they can occur randomly or due to the chosen network model. In the last few years, several authors have started to integrate traditional pre- and post-processing methods with deep networks as a tool to increase the performance and robustness of their approaches [15,61,64, 162,182]. Different from other reviews which typically focus on specific applications, this review focuses instead on the impact of different pre and post processing methods that are implemented within deep learning frameworks to deal with the very complex patterns of histological images. Many of the techniques presented here, especially the post-processing methods, are not limited only to histological image analysis but can be applied to almost any image analysis field.

Regarding pre-processing strategies, the most common algorithm adopted in digital pathology is stain normalization. The stain normalization process standardizes the stain color appearance of a source image with respect to a reference image. The current stain normalization methods can be based on different approaches, ranging from global color normalization to color transfer using generative adversarial networks (GANs). The standardization of histological images lets the deep network learn not only the certain color distribution but also the histopathological patterns. Moreover, including the stain normalization pre-processing gives forth more stable performances both on the train and test sets, especially if the data come from different centers and therefore with a variability of stains, scanners and sample preparation [63]. Numerous studies have also shown how including this pre-processing steps gives forth higher performances when using deep networks [56,57]. Another crucial pre-processing step during CNN training is patch selection. Different traditional algorithms based on thresholding, color deconvolution, and active contour models have been employed to identify the regions-of-interest in which to extract the CNN patches. Selecting patches only within specific regions of interest and not over the entire image increases the overall accuracy of a CNN as patches only containing significant information for the particular problem are analyzed. In particular, these approaches process only a fraction of the pixels in the raw image avoiding unnecessary calculations without sacrificing performances. Various studies have shown how a smart patch selection allows both to reduce computational times during inference and increase the model performance [79,82]. Finally, when processing whole-slide images (WSIs), it is fundamental to correctly detect the histological tissue and the various artifacts that can occur when preparing the histological slide (Section 2). Over the years, several strategies have been proposed to perform histological tissue segmentation and artifacts detection, mainly based on changing the color space and adaptive thresholding [32]. An accurate detection of tissue and artifacts lets a CAD program process a whole slide quicker, by excluding the background regions and avoiding regions that contains an altered morphology or intensity. Moreover, integrating these pre-processing strategies within a CAD pipeline also increases the performances of a deep-learning method [61,66].

As for post-processing strategies, we focused on all the three main tasks in computer vision tasks: classification, detection, and

segmentation. During classification tasks, a CNN is generally employed to predict the class label. In this review that is specifically focused on digital pathology, the main classification tasks considered were regarding prostate, breast, liver, and colon cancer. The main post-processing method applied during image classification is patch aggregation. A patch aggregation approach takes into consideration the characteristics and the labels of all the patches extracted by the CNN to predict the final image label. These strategies can adopt a simple voting procedure (i.e. max voting) as well as more sophisticated models, such as random forest or nearest-neighbor classifiers. Aggregating a large number of patches to predict the entire image class makes the deep learning model more robust to low-confidence predictions and single-patch misclassifications. Several studies demonstrated that the use of a patch aggregation strategy as a post-processing method improves the performance of a deep learning framework for cancer detection and grading [15,128,136].

Regarding detection tasks, a deep network is generally adopted to locate the centroid or the bounding box of the objects of interest within the image. Here, the main focus was put on the two most common detection tasks in histopathology: lymphocyte and mitosis detection. The main post-processing in deep learning-based detection frameworks is the non-maxima suppression (NMS) algorithm, which is an iterative method that takes as input the regions' proposal provided by the network and provides a list of filtered proposals. This specific post-processing allows the removal of overlapping bounding boxes while maintaining a high level of sensitivity. The integration of the NMS algorithm within deep learning networks improves the detection performance compared to CNNs that do not employ any post-processing [13, 153,165].

During object segmentation tasks, a CNN is employed to perform a pixel-level segmentation. In this review, the most common segmentation tasks in histological image analysis were analyzed, which are nuclei segmentation and tubules/glands segmentation. Two main post-processing strategies have been proposed for this task: a two-class pipeline and three-class pipeline strategy. In two-class pipelines, a CNN is adopted to perform binary segmentation (foreground vs background) and traditional techniques, such as morphological operators and the watershed transform, are employed to refine the segmentation. More recently, three-class pipelines have become commonly used, as they are able to simultaneously estimate the background, the inside, and the border of the object of interest. It is then possible to use more sophisticated post-processing techniques (e.g., connected component analysis) to both accurately and efficiently segment touching or high overlapping objects. Various studies have shown how these post-processing techniques allow to further reduce the network prediction errors and at the same time accurately define the borders of the objects of interest [70,192,204].

Over the last few years, there has been an ever-growing trend to use increasingly "deep" networks together with more and more sophisticated pre- and post-processing techniques in order to obtain progressively higher-performing CAD methods [66,110,128]. Combining more traditional techniques with deep learning networks has made it possible to improve the performance of single networks, bypassing some of their current limitations (random misclassification, pixel prediction errors). We strongly believe that in the future, increasingly refined pre- and post-processing strategies will be integrated into deep learning networks to create robust and reliable frameworks in the field of medical image analysis. This is confirmed by the recent exponential increase of publications that have integrated at least one pre- and/or post-processing stage to their pipeline.

Finally, it is encouraging to see how an "open data" mentality is becoming the norm especially in the field of deep learning, with researchers sharing both their dataset and codes, stimulating the development of deep learning frameworks that are progressively more robust and reliable. In addition, cloud-based systems are gradually spreading to overcome some of the current limitations of digital pathology like the

huge dimension of WSIs and the hardware resources needed to train deep models.

## 8. Conclusion

Due to its powerful learning ability and advantages in dealing with complex patterns, deep learning methods have been a research hotspot recently. Most machine learning methods have either a pre or a post-processing stage or both that are employed to make the subsequent classification, detection, or segmentation problem easier to solve. The integration of pre- and post-processing methods within deep learning frameworks has attracted much interest, and to date, the combination between these two techniques has become the standard method for image analysis in almost all research fields, in particular digital pathology.

## Conflicts of interest

The authors declare no conflict of interest.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

- [1] A. Janowczyk, A. Madabhushi, Deep learning for digital pathology image analysis: a comprehensive tutorial with selected use cases, *J. Pathol. Inf.* 7 (2016) 29, <https://doi.org/10.4103/2153-3539.186902>.
- [2] M.Z. Alom, C. Yakopcic, M.S. Nasrin, T.M. Taha, V.K. Asari, Breast cancer classification from histopathological images with inception recurrent residual convolutional neural network, *J. Digit. Imag.* 32 (2019) 605–617.
- [3] Z.-Q. Zhao, P. Zheng, S. Xu, X. Wu, Object detection with deep learning: a review, *IEEE Trans. Neural Networks Learn. Syst.* 30 (2019) 3212–3232.
- [4] G. Litjens, C.I. Sánchez, N. Timofeeva, M. Hermsen, I. Nagtegaal, I. Kovacs, C. Hulsbergen-Van De Kaa, P. Bult, B. Van Ginneken, J. Van Der Laak, Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis, *Sci. Rep.* 6 (2016) 26286.
- [5] O. Ronneberger, P. Fischer, T. Brox, U-net, Convolutional networks for biomedical image segmentation, in: *Int. Conf. Med. Image Comput. Comput. Interv.*, Springer, 2015, pp. 234–241.
- [6] W. Samek, T. Wiegand, K.-R. Müller, Explainable artificial intelligence: understanding, visualizing and interpreting deep learning models, *ArXiv Prepr.* (2017), [ArXiv1708.08296](https://arxiv.org/abs/1708.08296).
- [7] A. Janowczyk, A. Madabhushi, Deep learning for digital pathology image analysis: a comprehensive tutorial with selected use cases, *J. Pathol. Inf.* 7 (2016).
- [8] A. Zietman, J. Smith, E. Klein, M. Droller, P. Dasgupta, J. Catto, Describing the grade of prostate cancer: consistent use of contemporary terminology is now required, *Eur. Urol.* 70 (2016) 1.
- [9] D. Komura, S. Ishikawa, Machine learning methods for histopathological image analysis, *Comput. Struct. Biotechnol. J.* 16 (2018) 34–42.
- [10] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, *Adv. Neural Inf. Process. Syst.* (2012) 1097–1105.
- [11] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, *ArXiv Prepr.* (2014), [ArXiv1409.1556](https://arxiv.org/abs/1409.1556).
- [12] S. Ioffe, C. Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift, *ArXiv Prepr.* (2015), [ArXiv1502.03167](https://arxiv.org/abs/1502.03167).
- [13] M. van Rijthoven, Z. Swiderska-Chadaj, K. Seeliger, J. van der Laak, F. Ciompi, You Only Look on Lymphocytes once, 2018.
- [14] N. Zhou, Y. Gao, Optimized color decomposition of localized whole slide images and convolutional neural network for intermediate prostate cancer classification, in: *Med. Imaging 2017 Digit. Pathol.*, International Society for Optics and Photonics, 2017, p. 101400W.
- [15] D. Karimi, G. Nir, L. Fazli, P.C. Black, L. Goldenberg, S.E. Salcudean, Deep learning-based gleason grading of prostate cancer from histopathology images—role of multiscale decision aggregation and data augmentation, *IEEE J. Biomed. Heal. Informatics.* 24 (2019) 1413–1426.
- [16] M. Salvi, L. Molinaro, J. Metovic, D. Patrono, R. Romagnoli, M. Papotti, F. Molinari, Fully automated quantitative assessment of hepatic steatosis in liver transplants, *Comput. Biol. Med.* (2020) 103836.
- [17] A. Monte-Alto-Costa, L.C. Porto, Special stains for extracellular matrix, in: *Histol. Protoc.*, Springer, 2010, pp. 131–140.
- [18] J.D. Bancroft, M. Gamble, *Theory and Practice of Histological Techniques*, Elsevier health sciences, 2008.









- [181] A. Mandloi, U. Daripa, M. Sharma, M. Bhattacharya, An automatic cell nuclei segmentation based on deep learning strategies, in: 2019 IEEE Conf. Inf. Commun. Technol, IEEE, 2019, pp. 1–6.
- [182] O.C. Koyun, T. Yildirim, Adversarial nuclei segmentation on H&E stained histopathology images, in: 2019 IEEE Int. Symp. Innov. Intell. Syst. Appl, IEEE, 2019, pp. 1–5.
- [183] X. Pan, L. Li, H. Yang, Z. Liu, J. Yang, L. Zhao, Y. Fan, Accurate segmentation of nuclei in pathological images via sparse reconstruction and deep convolutional networks, *Neurocomputing* 229 (2017) 88–99.
- [184] S. Sornapudi, R.J. Stanley, W.V. Stoecker, H. Almubarak, R. Long, S. Antani, G. Thoma, R. Zuna, S.R. Frazier, Deep learning nuclei detection in digitized histology images by superpixels, *J. Pathol. Inf.* 9 (2018).
- [185] P. Naylor, M. Laé, F. Reyal, T. Walter, Segmentation of nuclei in histopathology images by deep regression of the distance map, *IEEE Trans. Med. Imag.* 38 (2018) 448–459.
- [186] Y. Song, L. Zhang, S. Chen, D. Ni, B. Lei, T. Wang, Accurate segmentation of cervical cytoplasm and nuclei based on multiscale convolutional network and graph partitioning, *IEEE Trans. Biomed. Eng.* 62 (2015) 2421–2433.
- [187] F. Xing, Y. Xie, L. Yang, An automatic learning-based framework for robust nucleus segmentation, *IEEE Trans. Med. Imag.* 35 (2015) 550–566.
- [188] H. Chen, X. Qi, L. Yu, Q. Dou, J. Qin, P.-A. Heng, DCAN: deep contour-aware networks for object instance segmentation from histology images, *Med. Image Anal.* 36 (2017) 135–146.
- [189] T. Wan, L. Zhao, H. Feng, D. Li, C. Tong, Z. Qin, Robust Nuclei Segmentation in Histopathology Using ASPPU-Net and Boundary Refinement, *Neurocomputing*, 2020.
- [190] Y. Cui, G. Zhang, Z. Liu, Z. Xiong, J. Hu, A deep learning algorithm for one-step contour aware nuclei segmentation of histopathology images, *Med. Biol. Eng. Comput.* 57 (2019) 2027–2043.
- [191] Z. Zeng, W. Xie, Y. Zhang, Y. Lu, RIC-Unet: an improved neural network based on Unet for nuclei segmentation in histology images, *Ieee Access* 7 (2019) 21420–21428.
- [192] L. Xie, J. Qi, L. Pan, S. Wali, Integrating deep convolutional neural networks with marker-controlled watershed for overlapping nuclei segmentation in histopathology images, *Neurocomputing* 376 (2020) 166–179.
- [193] H. Chen, X. Qi, L. Yu, P.-A. Heng, DCAN: deep contour-aware networks for accurate gland segmentation, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit, 2016, pp. 2487–2496.
- [194] A. BenTaieb, J. Kawahara, G. Hamarneh, Multi-loss convolutional networks for gland analysis in microscopy, in: 2016 IEEE 13th Int. Symp. Biomed. Imaging, IEEE, 2016, pp. 642–645.
- [195] Y.-R. Van Eycke, C. Balsat, L. Verset, O. Debeir, I. Salmon, C. Decaestecker, Segmentation of glandular epithelium in colorectal tumours to automatically compartmentalise IHC biomarker quantification: a deep learning approach, *Med. Image Anal.* 49 (2018) 35–45.
- [196] Y. Xu, Y. Li, Y. Wang, M. Liu, Y. Fan, M. Lai, I. Eric, C. Chang, Gland instance segmentation using deep multichannel neural networks, *IEEE Trans. Biomed. Eng.* 64 (2017) 2901–2912.
- [197] J. Tang, J. Li, X. Xu, Segnet-based gland segmentation from colon cancer histology images, in: 2018 33rd Youth Acad. Annu. Conf. Chinese Assoc. Autom, IEEE, 2018, pp. 1078–1082.
- [198] A. BenTaieb, G. Hamarneh, Topology aware fully convolutional networks for histology gland segmentation, in: Int. Conf. Med. Image Comput. Comput. Interv, Springer, 2016, pp. 460–468.
- [199] T. de Bel, M. Hermsen, G. Litjens, J. van der Laak, Structure instance segmentation in renal tissue: a case study on tubular immune cell detection, in: *Comput. Pathol. Ophthalmic Med. Image Anal*, Springer, 2018, pp. 112–119.
- [200] S. Rezaei, A. Emami, H. Zarrabi, S. Rafiei, K. Najarian, N. Karimi, S. Samavi, S.M. R. Soroushmehr, Gland segmentation in histopathology images using deep networks and handcrafted features, in: 2019 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc, IEEE, 2019, pp. 1031–1034.
- [201] J. Ren, E. Sadimin, D.J. Foran, X. Qi, Computer aided analysis of prostate histopathology images to support a refined Gleason grading system, in: *Med. Imaging 2017 Image Process*, International Society for Optics and Photonics, 2017, p. 101331V.
- [202] H. Qu, Z. Yan, G.M. Riedlinger, S. De, D.N. Metaxas, Improving nuclei/gland instance segmentation in histopathology images by full resolution neural network and spatial constrained loss, in: Int. Conf. Med. Image Comput. Comput. Interv, Springer, 2019, pp. 378–386.
- [203] S. Graham, H. Chen, J. Gamper, Q. Dou, P.-A. Heng, D. Snead, Y.W. Tsang, N. Rajpoot, MILD-net: minimal information loss dilated network for gland instance segmentation in colon histology images, *Med. Image Anal.* 52 (2019) 199–211.
- [204] T. Binder, E.M. Tantaoui, P. Pati, R. Catena, A. Set-Aghayan, M. Gabrani, Multi-organ gland segmentation using deep learning, *Front. Med.* 6 (2019) 173.
- [205] H. Ding, Z. Pan, Q. Cen, Y. Li, S. Chen, Multi-scale fully convolutional network for gland segmentation using three-class classification, *Neurocomputing* 380 (2020) 150–161.