Check for updates

Received: 24 December 2021 | Revised: 14 June 2022 | Accepted: 22 June 2022

DOI: 10.1002/widm.1474

ADVANCED REVIEW

WIREs DATA MINING AND KNOWLEDGE DISCOVERY

WILEY

A survey on artificial intelligence in histopathology image analysis

Mohammed M. Abdelsamea[1,2] | **Usama Zidan**[1] | **Zakaria Senousy**[1] | **Mohamed Medhat Gaber**[1,3] | **Emad Rakha**[4] | **Mohammad Ilyas**[4]

[1]School of Computing and Digital Technology, Birmingham City University, Birmingham, UK

[2]Department of Computer Science, Faculty of Computers and Information, University of Assiut, Assiut, Egypt

[3]Faculty of Computer Science and Engineering, Galala University, Galala, Egypt

[4]Department of Histopathology, Queens Medical Center, Nottingham University, Nottingham, UK

Correspondence
Mohammed M. Abdelsamea, School of Computing and Digital Technology, Birmingham City University, Birmingham, UK.
Email: mohammed.abdelsamea@bcu.ac.uk

Edited by: Mehmed Kantardzic, Associate Editor and Witold Pedrycz, Editor-in-Chief

Abstract

The increasing adoption of the whole slide image (WSI) technology in histopathology has dramatically transformed pathologists' workflow and allowed the use of computer systems in histopathology analysis. Extensive research in Artificial Intelligence (AI) with a huge progress has been conducted resulting in efficient, effective, and robust algorithms for several applications including cancer diagnosis, prognosis, and treatment. These algorithms offer highly accurate predictions but lack transparency, understandability, and actionability. Thus, explainable artificial intelligence (XAI) techniques are needed not only to understand the mechanism behind the decisions made by AI methods and increase user trust but also to broaden the use of AI algorithms in the clinical setting. From the survey of over 150 papers, we explore different AI algorithms that have been applied and contributed to the histopathology image analysis workflow. We first address the workflow of the histopathological process. We present an overview of various learning-based, XAI, and actionable techniques relevant to deep learning methods in histopathological imaging. We also address the evaluation of XAI methods and the need to ensure their reliability on the field.

This article is categorized under:
Application Areas > Health Care

KEYWORDS

Actionability, artificial intelligence, deep learning, histopathology, image analysis, machine learning

1 | INTRODUCTION

Artificial Intelligence (AI) has recently shown great and considerable success in the analysis of medical images. The development of whole slide scanners enabled the digitization of histopathological process by generating whole slide images (WSI) and facilitating the pathologist's workflow. This has encouraged the production and application of various AI techniques in histology diagnosis. AI-augmented histopathological diagnosis includes segmentation, quantification,

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2022 The Authors. *WIREs Data Mining and Knowledge Discovery* published by Wiley Periodicals LLC.

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstine), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

and classification. A typical histopathology image analysis pipeline is composed of a training and a testing phase. The training phase usually builds statistical models of the region of interest (ROIs) on the basis of previously labeled images while the testing phase uses those trained models to detect, segment, and/or classify the ROIs in previously unseen images. The training phase will then typically consist of (a) image pre-processing to normalize and clean the input images in order to deal with noise, artifacts, and intensity inhomogeneities; (b) region detection, segmentation, and/or classification to semantically extract the ROIs; (c) feature extraction; and (d) model building as a means to learn the morphological structure (e.g., appearance and/or shape) of the ROIs. The testing phase starts with the same (a-c) steps and then compares the extracted features to the training-built models to produce a classification map. Finally, a quantitative analysis component is used to assist the extracted knowledge based on the same criteria that are usually used by pathologists to make the diagnosis. Histopathology images have very high visual appearance variability based on the morphological and architectural characteristics of the biological structures that typically reflect the output of a complex and multidimensional molecular processes in the component cells. Several other technical factors such as staining differences, orientation of the objects, and the scanning magnifications of the biological samples, just to mention a few, are contributing to the visual heterogeneity of the histopathology image, see Figure 1. Moreover, computer vision challenges such as multi-class intensity/morphological overlaps, fully/partially occlusion, illumination variations, and objects with weak (or ill-defined) boundaries are usually presented in histopathology images, making it hard to distinguish between the different classes in the images, see Figure 2.

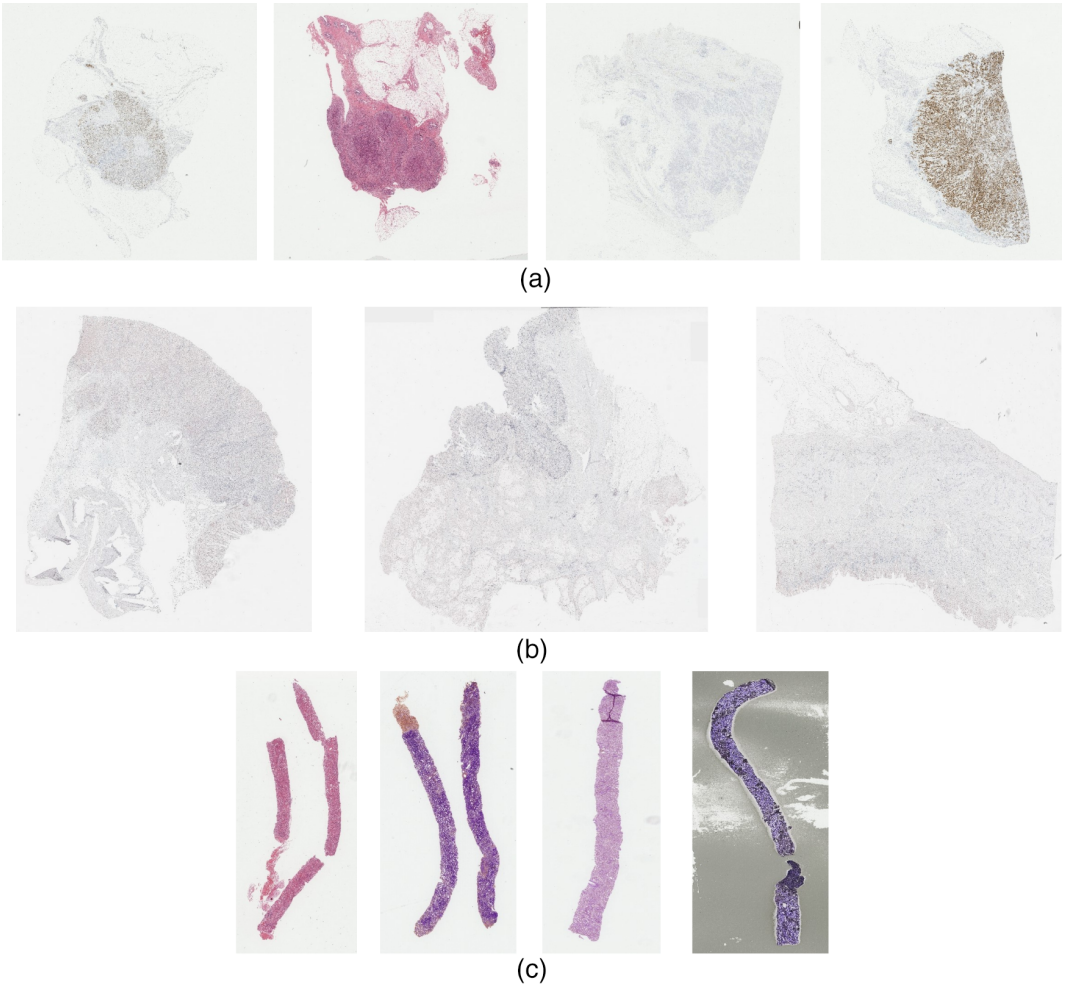

FIGURE 1 Examples of different stains with different histopathological sections, from https://anhir.grand-challenge.org/. (a) Breast tissues stained with, from left to right, estrogen receptor (ER), hematoxylin and eosin (H&E), Her2-neu, and progesterone receptor (PR), and Her2-neu; (b) kidney tissues, stained with, from left to right, H&E, Masson, PAS, and Methenamine; and (c) gastric mucosa and gastric adenocarcinoma tissues staining with, from left to right, CD4 (clone 4B12), CD8 (clone C8/144B), and CD68 (clone PG-M1).

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstinet), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

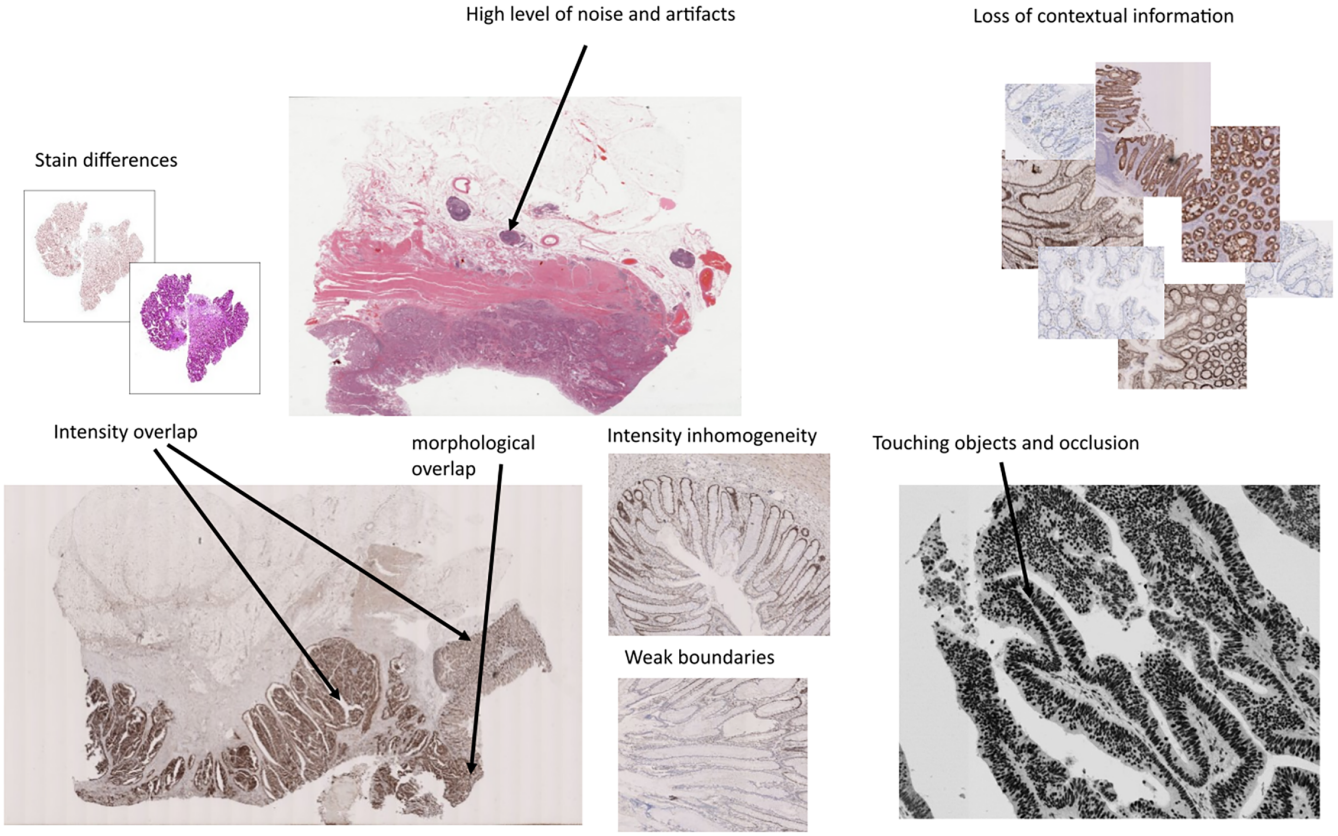

FIGURE 2 Some computer vision challenging demonstrated on some of our WSIs (Abdelsamea, Grineviciute, et al., 2019a; Abdelsamea, Pitiot, et al., 2019b).

AI techniques have the capability to tackle such computer vision tasks through Hand-designed methods (Feng et al., 2016; Fu et al., 2014; Gunduz-Demir et al., 2010; Liu et al., 2016; Phoulady et al., 2016; Shahul Hameed et al., 2017; Wu et al., 2005) and trainable methods (Chen et al., 2017; Kainz et al., 2015; Li et al., 2017; Sirinukunwattana et al., 2015). Hand-designed methods divide an image into multiple regions based on the similarities of pixels in different feature spaces (e.g., color, texture, gradient, regional statistics, shape, etc.) within each sub-region and the dissimilarities across different sub-regions. With the recent trend of deep learning, more publications are using learning-based methods due to their superior performance (Ibrahim et al., 2020). This class of methods involves a training process to learn prior knowledge about the particular type of images to be segmented. The prior knowledge can be a probabilistic approach that represents the probability distribution of each sub-region. This is then used to infer the probability of each pixel in unseen images belonging to each of the classes. However, in practice, the probability distribution is difficult to be estimated accurately due to the limited number of training images. Alternatively, a feature descriptor can be used to describe the local characteristics in the training images, followed by training a classifier to best discriminate different classes based on the feature descriptor. Subsequently, in the segmentation stage, the classifier is applied to estimate the best possible class for each pixel based on the extracted feature descriptor. Many methods have been proposed which have supreme performance than Hand-designed methods (Ballerini et al., 2004; Belhassen & Zaidi, 2010; Katouzian et al., 2012; Naik et al., 2008; Nguyen et al., 2010; Sirinukunwattana et al., 2015). With a deep learning architecture and a large number of training images, the feature descriptors can be automatically learnt (Chen et al., 2017; Chen & Chefdhotel, 2014; Kainz et al., 2015; Sirinukunwattana et al., 2016). However, the learning-based method requires a set of training images with detailed annotations, which is difficult and time consuming to obtain in histological imaging. Nevertheless, learning-based methods offer more promising results with continuing advancements in medical images using deep learning methods.

Despite the advances in the application of AI in a medical workflow, there remain several challenges. These algorithms deliver highly accurate predictions. Yet transparency and understandability are critical factors that should be addressed in order for these technologies to be implemented in real-world workflows. Due to the black-box nature of

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstinet), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

deep learning (DL) techniques, that are utilized in learning-based methods, interpretability methods are needed to understand the underlying mechanism behind these decisions not only to increase users' trust in this novel technology application but also to allow the wide use of algorithms trained for specific tasks in a way akin to human approaches. Explainable artificial intelligence (XAI) seeks to improve the trustworthiness of AI solutions by providing an insight into the predictions of machine learning (ML) models. This can help healthcare workers to provide more qualitative proof of what prompted the algorithm to make these decisions. These techniques can range from intuitive designs such as a presentation of contributing factors projected on the input (e.g., heatmaps) to patterns portraying the inner workings of a model. The demand for explanations is the result of many sectors trying to integrate AI into their workflow but failing to fully rely on them because of a lack of trust, understandability, or transparency.

To identify the position of our survey paper, we have highlighted and summarized the main aspects reviewed by the key survey articles in digital pathology as follow: The emergence of feature engineering and learning approaches in digital pathology, for object detection, image segmentation, and tissue classification has been discussed in Gurcan et al. (2009) and Veta et al. (2014). Some challenging problems such as technical and computational challenges in different applications such as diagnosis and prognosis and opportunities for the quantitation and predictive modeling of digital pathology images have been discussed in (Bueno et al., 2016; Madabhushi & Lee, 2016; Xing et al., 2017). Other challenging problems related to the image storage and management due to the large size (>10 GB) of whole slide images and their impact on the diagnosis-making process have been discussed in (Al-Janabi et al., 2012). Applications of machine learning models in digital pathology (such as computer-assisted diagnosis, content-based image retrieval, and clinicopathological relationships) have been discussed in this mini-review paper Komura and Ishikawa (2018). Focusing on the automation of pathology workflows, Echle et al. (2021) provides a review of deep learning models in histopathology analysis tasks such as detection and grading the context of important clinical applications such as the inference of molecular features, prediction of survival and therapy response. In Srinidhi et al. (2021), different classes of machine learning methods (such as supervised, weakly supervised, unsupervised, and transfer learning) have been discussed from the methodological perspective. The use of deep learning models, their limitations and future perspectives in diagnostic breast pathology have been discussed in Robertson et al. (2018). Transparency and reliability challenging problems in digital pathology have also been discussed in (Pocevičiūtė et al., 2020) through XAI techniques for deep learning methods. To the best of our knowledge, this paper is the first survey to cover the developments in histopathology workflow from annotations to explainability and actionability challenges. We also review the state-of-the-art learning-based and XAI techniques and briefly discuss the evaluation of such techniques and the need to ensure their reliability on the field. This survey is meant for technical researchers and newcomers to the field and aims to provide the requisite information for further development in the field (see Appendix S1 for brief descriptions of the main terminologies used in the survey in the Supplementary Materials). Figure 3 shows the main topics that have been reviewed and covered in this survey.

The adopted methodology in conducting the survey is based on three foundations: (1) the separation of historical work before and after the era of deep learning is used; (2) the categorization of the reviewed methods is based on the application areas in histopathology; and (3) the selection of articles to be included is based on recency and impact of the work. The selection of the papers that have been reviewed in this survey was first carried out by search process using a combination of keywords in histopathology image analysis (such as histopathology, deep learning, computational pathology, digital pathology, artificial intelligence, and machine learning). We then used exclusion criteria to (a) only capture studies that considered AI algorithms applied and contributed to the histopathology image analysis workflow, and (b) focus on the application areas in histopathology. As a consequence, a total of 155 papers were selected and reviewed in this survey paper.

The paper is organized as follows. Section 2 goes over the process of annotating histopathological data. Section 3 reviews the development of computational pathology before and after deep learning era with a focus on deep learning applications. Section 4 demonstrates the current/future research directions in computational pathology, including Scarce annotations problem, the importance of uncertainty quantification and explainability in future efforts and outlines some factors regarding actionability and its value in the medical workflow. Section 5 discusses the most common validation metrics that have been used for explainable techniques in computational pathology. Finally, Section 6 provides a brief summary of the paper.

2 | TRAINING DATA ANNOTATION

In order to create new ways to scan medical (histopathological slide) images to whole slides images (WSI), the quality characteristics of WSI have to be identified. Preparation of glass slides used to generate WSI requires several steps to

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstinet), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

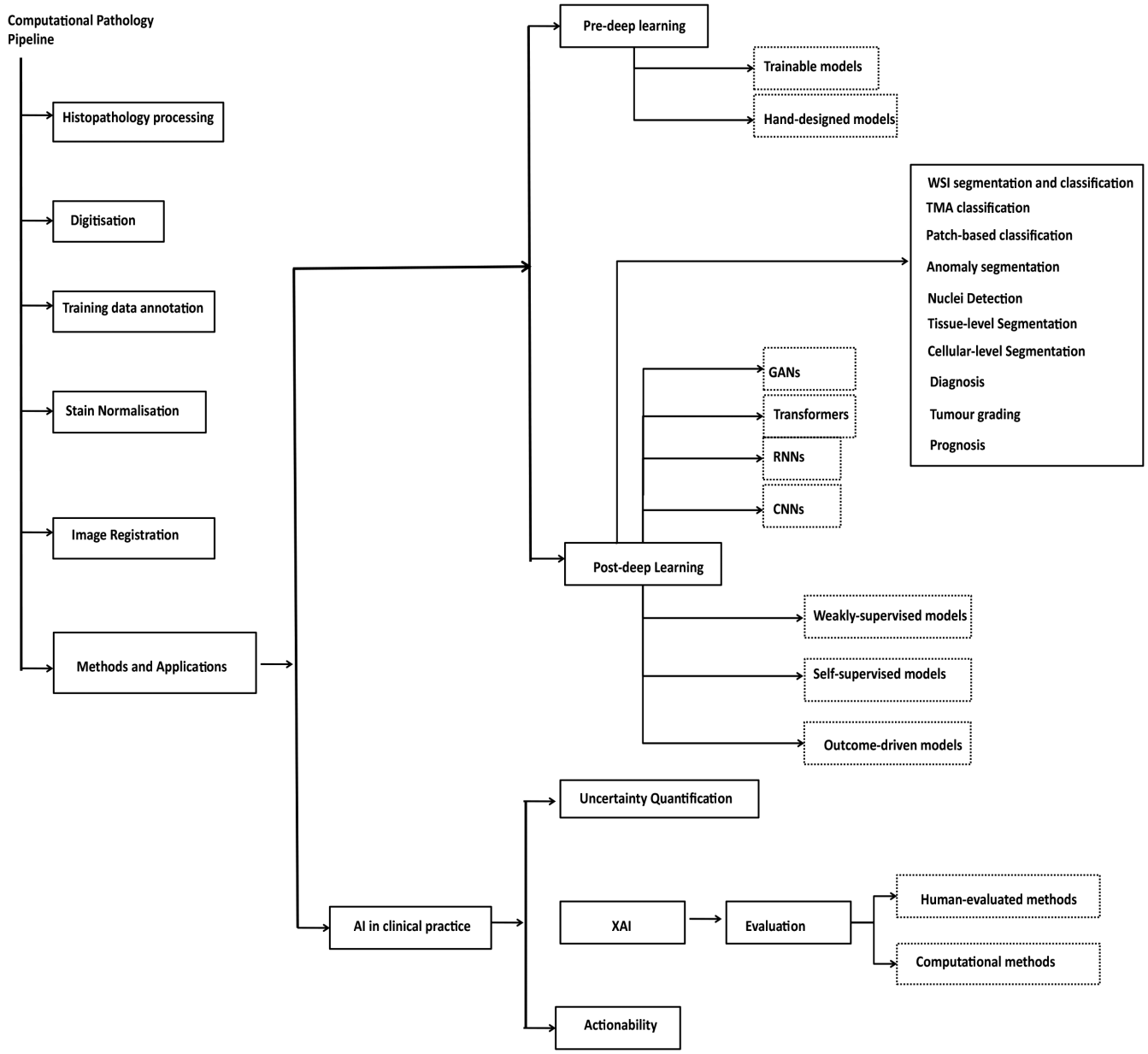

FIGURE 3 The main topics covered in this survey paper

ensure good quality slides suitable for WSI examination. These include tissue procurement, fixation, processing, and staining. The next move is to digitize versions of these slides to entire images (e.g., WSI) using high quality scanners. These images are then examined by a pathologist. Since several clinical reports are written in free text type, standardized labels have to be developed to ensure all articles are the same in appearance. Annotations made by medical practitioners, such as pathologists, are image labels. These annotations are called ground truth for imagery when the imagery is the known reference standard. Choosing an appropriate classification for any imaging tool requires a proper balance between finding the best discriminating categories (e.g., normal vs. abnormal) and clinically relevant granularity (e.g., sub-type of liver lesion) (Figure 4).

To establish consistently structured labels, some labeling strategies have been developed. The commercially available HALO image analysis software (www.indicalab.com) is commonly used for producing manual annotations. For example, manually annotating images is possible but very time consuming and impractical at times when imaging examinations are not satisfactory alone and might require assessing follow-up pathological diagnosis, or clinical outcomes to arrive at the ground truth, see Figure 5.

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstinet), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

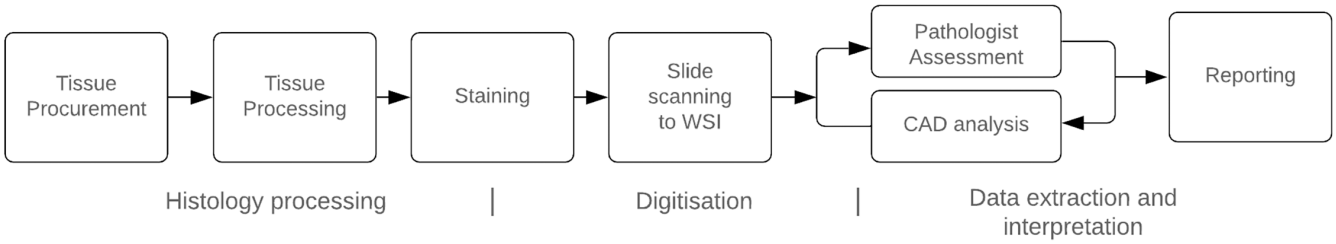

FIGURE 4 An illustration of the histopathological workflow in pathology, adopted from (Barisoni et al., 2020).

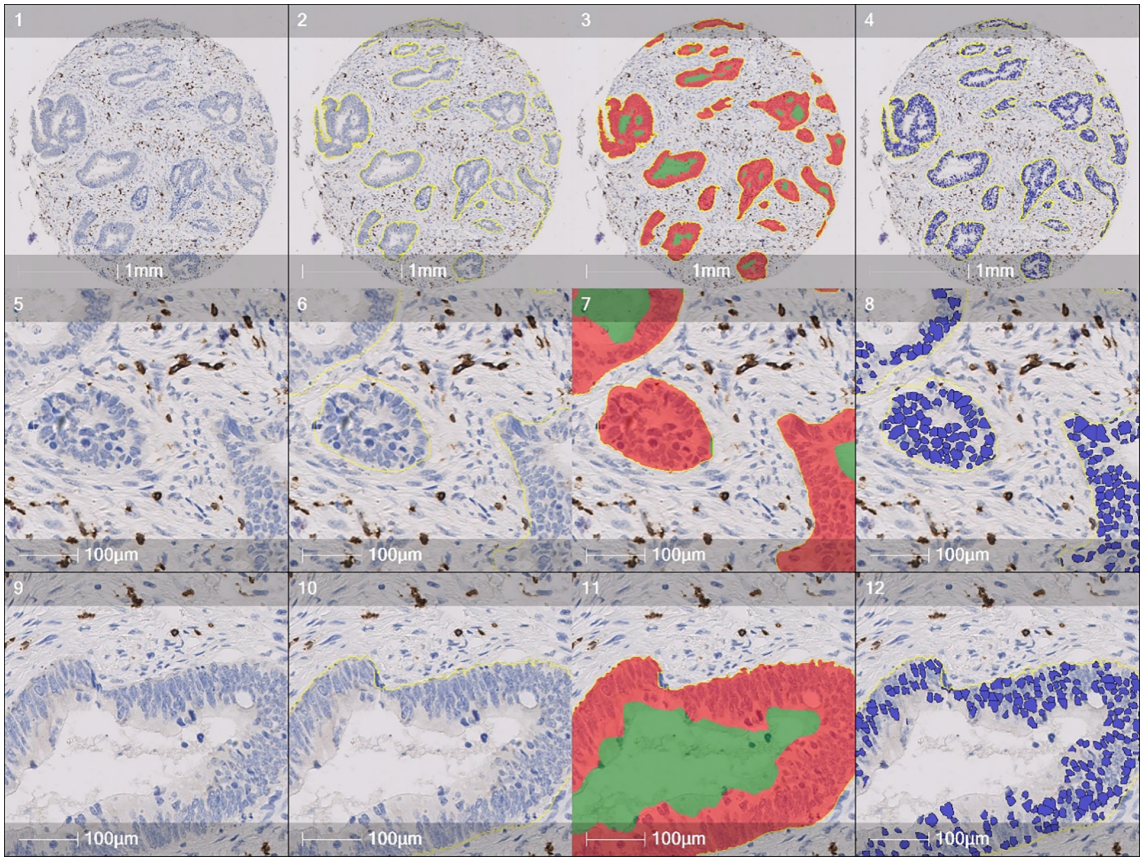

FIGURE 5 Ground truth data annotated and digital image analysis by HALO software (https://indicalab.com/halo/). Images 1, 5, and 9 represent scanned TMA images without analysis. The precise manual annotations (yellow line) were drawn to delineate invasive tumor complexes (2, 6, and 10). Images 3, 7, and 11 illustrate HALO image analysis results to classify tumor epithelium (red) inside manual annotations avoiding artifacts (green) such as necrotic debris, stroma and glandular lumens. HALO Cytonuclear algorithm was calibrated to segment tumor cell nuclei (blue) inside previously classified tumor areas (4, 8, and 12).

Imaging data may typically be classified in a number of ways, including structured labels, image annotations, image segmentation, and/or electronic phenotypes. The implementation of an imaging diagnosis tool based on expert interpretation (on the basis of a reinterpretation of the images or a free-text reports) is the most frequently used method. Automatic techniques, on the other hand, are looking more like promising solutions, although they come with their shortcomings. Natural language processing (NLP) techniques offer more scalable solutions. A useful NLP method is topic modeling, which utilizes free text reports to obtain gross insight over the data set. This approach characterizes report content on the basis of key terms and estimates of the topics covered in the reports. Another class of methods in NLP are recurrent neural networks, which can be trained on small sample reports to achieve state-of-the-art performance in radiology (Banerjee et al., 2019; Lakhani et al., 2012). These methods have shown great promise in generating structured labels in large quantities, allowing for the production of larger imaging data sets.

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstinet), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Recently, interactive reporting has seen an uplift in attention. The premise is that by using hyperlinked text during the process of reporting on a study, the clinical staff would link their descriptions to certain parts of the image. This allows for far more accessible reports with immediate evaluation of the findings. (Folio et al., 2018) used such a technique in a radiological workflow and found that this interconnection between the hyperlinked text and image annotation instrumental in enhancing the value of the examination reports. They found that it did not take longer for the radiologist to apply these hyperlinks. (Yan et al., 2018) used bookmarks made during the report writing to generate bounding boxes for an open-source dataset. Preprocessors could also be implemented to pre-segment regions in the image before clinical interpretation (Do et al., 2020). These annotations could then be approved by the examiner resulting in a more improved target measurement, earlier investigation, and faster notification of incidental actionable findings to referring technicians. These methods offer a potential for a radiological workflow that can bridge the gap between local labeling and expert labeled data sets that can be used for training, validation, and testing of future AI techniques. However, the question whether these methods are applicable or can be adapted to cope with the challenging problems in digital pathology is still not clear.

3 | HISTOPATHOLOGY IMAGE ANALYSIS

Unlike radiology images, histopathology images are far more complex due to the amount of information contained in the images where several regions are present and each region has a special characterization in terms of shape, color, texture, gradient, and/or other features that might be used as similarity or difference criteria, see Figures 6 and 7.

In the context of histopathology image analysis, one of the important initial tasks is to visually compare tissue sections of the same subject but of different biomarkers in a single frame. This is achieved by determining a transformation that maximizes the similarity between a reference image and its associated moving image (Borovec et al., 2018; Borovec et al., 2020; Fernandez-Gonzalez et al., 2002; Gupta et al., 2018). With the aim of improving precision and reducing

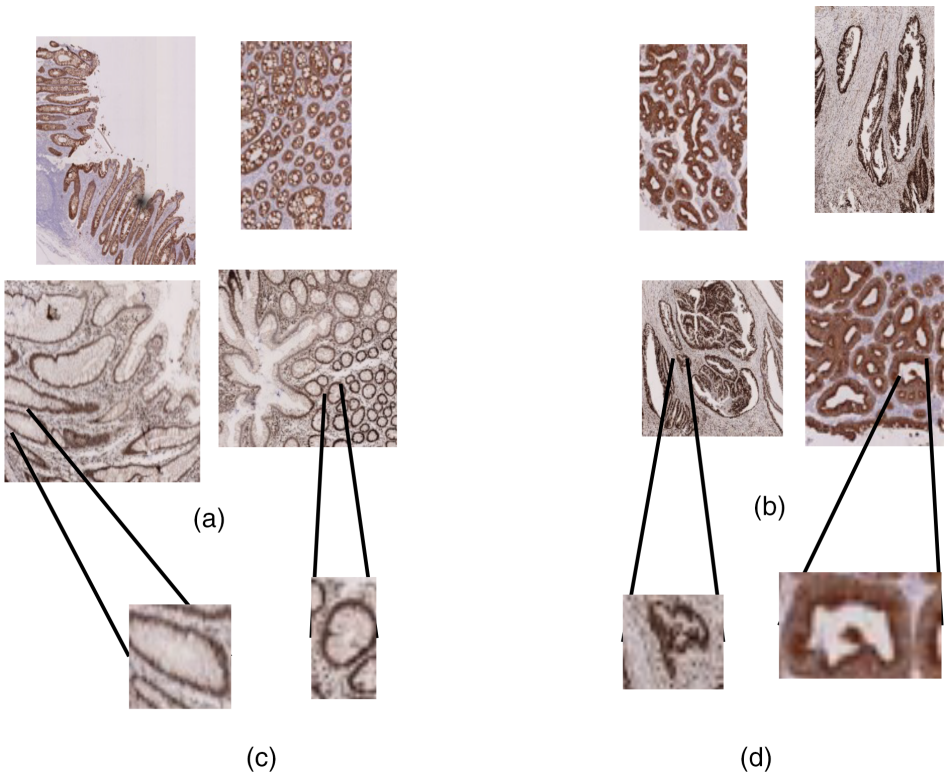

FIGURE 6 Some challenging examples of (a) normal and (b) tumor epithelial regions that have been randomly extracted from WSIs stained by AE1AE3 (bright brown) and BerEP4 (dark brown). These sections share the same intensity profiles with different morphological structures, for example, the nuclei of (c) the normal regions are regularly distributed while (d) they are randomly distributed in the tumor regions.

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstinet), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

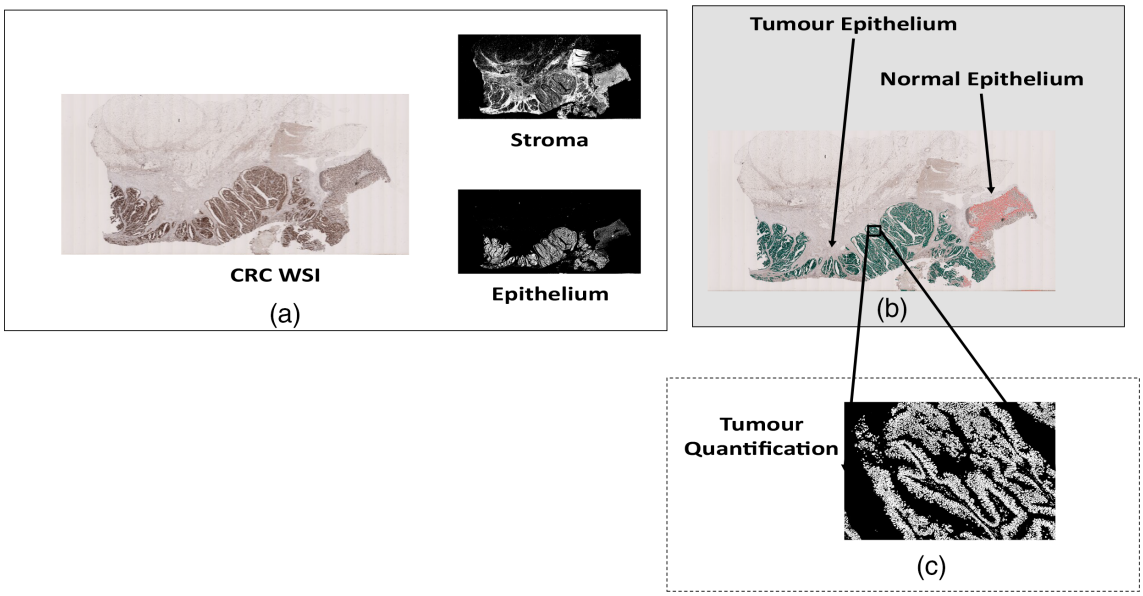

FIGURE 7 The different components of tumor Parcellation and quantification (TuPaQ) tool in analyzing immunohistochemistry stained image (Abdelsamea, Grineviciute, et al., 2019a). First the input WSI is uploaded into TuPaQ for automatic image processing, where the output is two binary masks for the stroma and epithelium regions extracted using an active contour model (a). Then the epithelium binary mask is further divided into normal and tumor regions using the TuPaQ's tumor epithelium identification (TEI) component, with the tumor epithelium highlighted in green and the rest of the image in red (b). Finally, a Tumor Quantification (TQ) component is used to divide the nuclei clusters into individual ones and give predicted nuclei counts in the tumor epithelium region (c).

subjectivity, several different image analysis approaches have been used for analyzing digitized histopathological images. Object mage Segmentation is the most complicated challenging problem in many histopathology image analysis pipeline. Likewise, image/tissue classification is essential and fundamental components in any analysis pipeline, see Figure 8. In the following subsections, we categorize recent computational pathology models based on the different application areas and the development before and after the era of deep learning.

3.1 | Pre-deep learning

Here, we review traditional computational pathology methods before the era of deep learning. We categorize pre-deep learning methods into: (1) Hand-designed methods, and (2) trainable methods. Unlike trainable methods, hand-designed methods are computationally efficient and works effectively to process images with homogeneous regions or with distinctive boundaries. They are also very easy to be adapted from application to another. Although some methods are more robust than others, they all typically have difficulties to achieve consistent results when large regional variations and noises are presented in the images.

3.1.1 | Hand-designed methods

Energy-based methods (Chang et al., 2012) are powerful hand-designed methods that have been developed/used as a part of analyzing different histopathological tissue regions by maximization or minimization of an energy function. *Contour-based* models or "Active contours" are effective energy-based tools that have been extensively used in histopathology image analysis piplines. *Contour-based* models can be divided mainly into parametric and geometric active contours. Parametric active contours, also called "Snakes", rely on a kind of parametrization (Sekhar et al., 2008) to approach the contour and control its evolution. Gradient Vector Flow (*GVF*) (Plissiti et al., 2010) is a parametric active contours that can be considered as an extension version of snakes, which is proposed with the aim of improving its robustness to contour initialization and to boundary concavities. Gradient vector flow snake was applied for the segmentation of nuclei and cytoplasm in blood cells stained images (Ko et al., 2011). However, the segmentation scheme is

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstinet), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

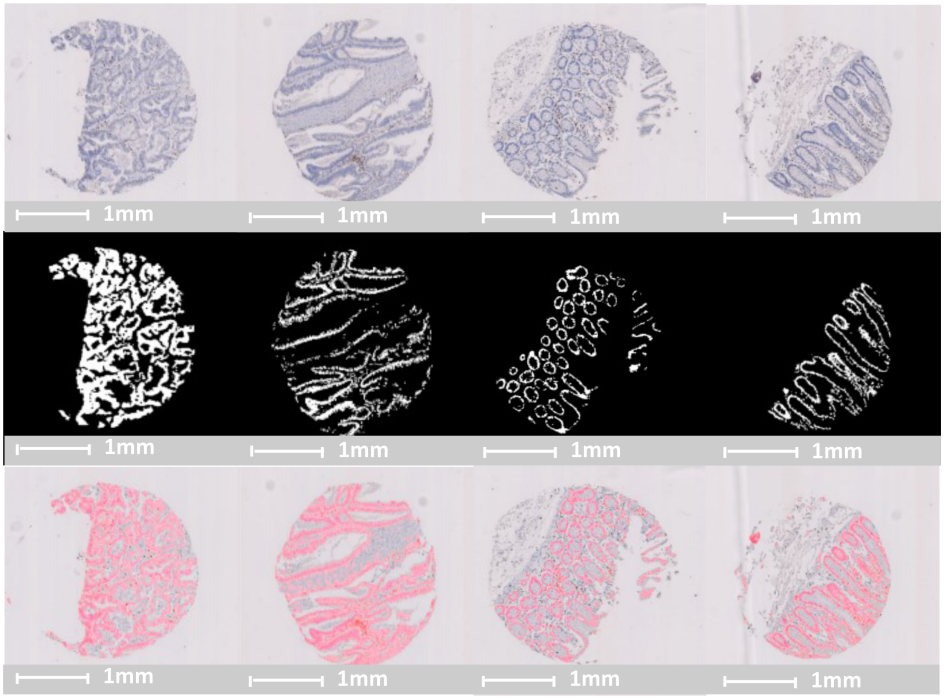

FIGURE 8 A few examples of segmentation and classification of tissue micro arrays (TMAs) cores of tumor (first two columns) and normal (last two columns) colon tissues stained by CD3 (first row), with ground truth images (second row), and results from a previously developed method of (Abdelsamea, Grineviciute, et al., 2019a) to differentiate between epithelial regions (highlighted in pink) stromal regions (third row). The method is working by first segmenting the epithelium using an approach called fuzzy signed pressure force approach (FSPF), based on fuzzy c-means method and level sets (or active contours) where fuzzy c-means was used to control the evolution of the contour in a way to accuratly extracting the epithelium based on color information. Then, the method calculates a set of novel morphometric features based on the Axis of least inertia (ALI) and appearance features based on the intensity profile of the epithelium, which are invariant to staining. Finally, the extracted features were used to distinguish between tumor and normal epithelium using two Self Organizing Maps (SOMs) that have been previously trained on manually labeled images.

computationally expensive and still needs further adjustment to improve the segmentation performance, because of the sensitivity to the contour initialization. Moreover, Geodesic Active Contours (*GAC*) approach (Hafiane et al., 2008) is one of the most popular edge-based active contour approaches that has been used in histopathological object segmentation. As a way of improving the robustness to contour initialization of active contour segmentation frameworks, in (Fatakdawala et al., 2010), an expectation-maximization (*EM*) algorithm to detect the centers of lymphocytes on breast cancer histopathology images has been proposed for automatically initializing the *GAC*. As a consequence, the initial contours were defined using the detected centers. However, only small image patches of size 200 × 200 pixels were considered in this study. Chan and Vese (*C–V*) approach (Ali & Madabhushi, 2012) is the most used region-based approach, which can perform better than edge-based active contour approaches for objects with blurred or weak boundaries in histopathological segmentation. In Hafiane et al. (2008), Chan and Vese has been used for gland structure segmentation as a multiphase vector-based level set method in prostate cancer images. In (Abdelsamea, Pitiot, et al., 2019b), an active contour method based on a fuzzy sign pressure force function, has been proposed to segment out epithelial regions from colorectal cancer cases in a way to be robust and less sensitive to noise, see Figure 9. Our practical recommendation for hand-designed methods is to make use of morphometric features (e.g., features extracted from segmented objects within the images) as part of the histology image analysis pipeline. Morphometric features can describe clearly the morphological structure of ROIs. We also noticed that shape features can accommodate the high visual variability of the ROIs and hence can provide stain-independent solutions.

As a part of grading prostate cancer (*CaP*) in histopathology images, a hybrid segmentation framework that combines color gradient geodesic active contours (*CGAC*) and hierarchical normalized cuts (*HNCut*) has been proposed (Xu et al., 2011). It claimed to improve the robustness to the contour initialization which gave an initial guess of the location of the region of interests (e.g., a gland, which is composed of a central lumen area surrounding epithelial cytoplasm with a ring of epithelial nuclei defines its outer boundary). The *HNCut*, which is mainly inspired by both mean shift

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstinet), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

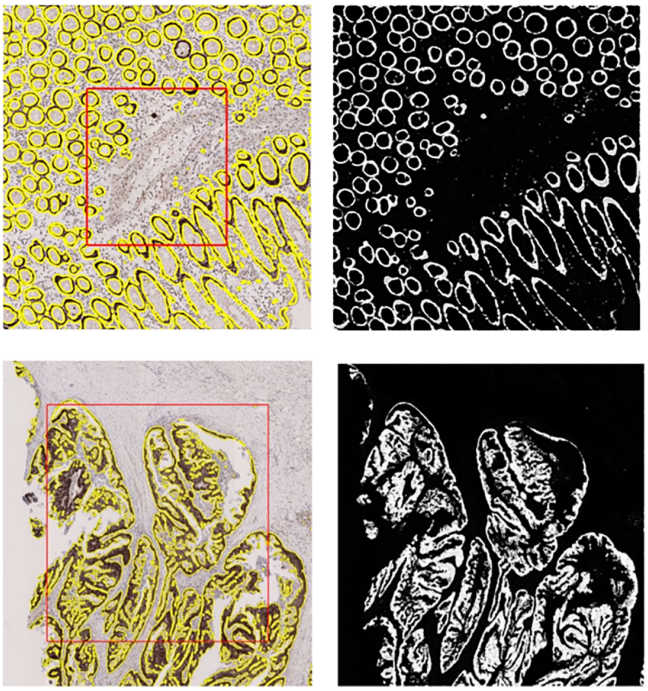

FIGURE 9 Segmentation results obtained by fuzzy sign pressure force function (Abdelsamea, Pitiot, et al., 2019b) on colorectal cancer sections: First row shows original images with the initial contours (in red) and segmented epithelial regions (in yellow) while second row shows the associated binary masks.

clustering and normalized cuts, acts as an initial contour to the *CGAC* segmentation framework. The main aim to develop such a hybrid approach is to improve: the ability to deal with multiple regions; improve the computational time; and decrease the user-intervention.

In spite of the fact that energy-based models, especially the non-parametric active contours, show superiority in performance when dealing with complex distribution, they may fail due to inaccurate initialization. Similarly, boundary-based active contour models have the ability to find accurately the region of interest by incorporating local information around the object boundaries into its edge-detection energy functional. However, they are not robust to the contour initialization and fail to extract multiple objects due to the parametric representation of the contour, which does not allow for the topological changes (e.g., merging and splitting) to be handled implicitly during the evolution.

3.1.2 | Trainable methods

Trainable methods or classical/statistical machine learning methods that have been previously proposed for histopathology image analysis can be classified into two categories, namely, generative and discriminative methods, which mainly consist of an off-line training phase and an on-line deployment phase. The off-line training phase can be summarized in the following three steps: (a) image pre-processing, which aims to attenuate the visual appearance variability in the image including noise, artifacts, and intensity in-homogeneity. Additionally, *ROI* extraction is also often required; (b) feature extraction/representation to project the original data space into feature space (i.e., feature descriptors), which better captures the local characteristic of the data; and (c) based on a training dataset, training a classifier (or building a probabilistic model) to best discriminate (or describe) the features extracted from different segmentation classes. In the on-line phase, the same image pre-processing and feature extraction methods are applied to the unseen image, followed by pixel/region-level prediction using the trained classifier (or learnt probabilistic model).

Generative approaches (Sirinukunwattana et al., 2015) can be used as a segmentation method by building a probabilistic model that learns the probability of each pixel to belong to a specific class (e.g., foreground/background). This is by finding the best estimate of its parameters for some pre-specified parametric form of a probability distribution. Examples of generative approaches include fuzzy *c*-means (Belhassen & Zaidi, 2010), Bayesian inference (Sirinukunwattana et al., 2015), and Gaussian Mixture approaches (Khan et al., 2012). In (Naik et al., 2008), nuclear

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstine), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

and glandular segmentation was achieved by integrating a Bayesian classifier driven by image color, texture, and morphological features for automated grading of both prostate cancer and breast cancer, and distinguishing between cancerous and benign breast histology specimens. Also, a Bayesian voting-based approach has been presented in Hiary et al. (2013), for the removal of stromal cells in histology images by utilizing cell texture and color. In Hafiane et al. (2008), the results from fuzzy *c*-means clustering were employed to initialize the active contour approach for segmenting the nuclei on prostate histopathology. However, such kinds of solutions are not efficient enough especially when dealing with large images.

Generative-based approaches are usually computationally efficient. However, with a small training dataset, the estimated parametric approach may not represent the true probability distribution, which results in an inaccurate prediction.

Discriminative approaches (Katouzian et al., 2012; Naik et al., 2008; Nguyen et al., 2010) have been proposed mainly with the aim of ignoring probability, and the main focus was to construct an accurate decision boundary directly. This approach is often extremely successful because of its ability to discover the underline distribution of the data, especially when no parametric probabilistic approach of the data exists. By using the extracted image features (mentioned above) as input, a model can be trained to learn the decision boundaries that best separate different classes. Such techniques include *K*-nearest neighbor (Lee et al., 2010), Self Organizing Maps (SOMs) (Katouzian et al., 2012; Sertel et al., 2009) and random forest (Sommer et al., 2012) Support Vector Machines (Naik et al., 2008). A three-phase framework has been proposed in (Nguyen et al., 2010) to segment the whole structure of a gland for Gleason grading of prostate tissue histopathology images. The framework is composed of three phases, they are: (1) pixel-based classification, (2) gland exterior boundary extraction, and (3) the whole gland construction. In the pixel-based classification phase, each pixel has been classified into five different classes (e.g., stroma, lumen, and nontissue area, epithelial nuclei, epithelial cytoplasm and mucin). First, the $La * b *$ color components have been extracted, where the L channel for illumination and both $a *$ and $b *$ components for the color-opponent dimensions. Then, around 200 samples from each class was used as training samples to feed a *k*-nearest Neighbor classifier in order to classify each pixel accordingly. A developed segmentation algorithm based on Hierarchical Self-organizing maps (*HSOM*) (Datar et al., 2008) has been proposed with the aim of segmenting tissue images into four classes: glands, epithelium, stroma, and nuclei. Each pixel has been represented by a feature vector describing the texture and color representation of the pixel to feed the trained *HSOM* as a classification model. Similarly, aiming to segment H&E stained cervical tissue images into regions of squamous epithelium, a patch-based hybrid method has been proposed in (Wang et al., 2007). The method relied on support vector machine (*SVM*) to classify patches according to their texture information at a low resolution and then refine the obtained segmentation results, at a higher resolution and in a similar way, as a post-processing phase. It has been shown that the method is computationally not efficient and an over-segmentation problem has been reported.

Overall, the above mentioned approaches share the same limitations as they result in objects with discontinuities and ill-defined boundaries as pixel/patch-based methods based on some statistical assumptions about the appearance distribution of the object(s) to be processed. Hence a post-processing phase is usually required.

3.2 | Postdeep learning

In this section, we review the recent deep learning methods that have been previously proposed in histopathology image analysis. A summary of some common deep learning algorithms, along with information about the image datasets and validation, is reported in Tables 1 and 2. Deep learning approaches (e.g., convolutional neural networks [*CNN*s], see Figure S1 in the Supplementary Materials) have been successfully applied to several tasks, in digital pathology, including image/object segmentation and classification (Chen et al., 2017; Kainz et al., 2015; Malon et al., 2008; Sirinukunwattana et al., 2016). One of the advantages is that a hierarchical feature descriptor, which captures different levels of details, can be automatically learnt in the training process. With a deep learning architecture and a large number of training images, the feature descriptors can be automatically learnt. In this section, we review the different histopathology applications with deep learning as the main computational pathology models.

3.2.1 | Tissue classification

Recent techniques have been proposed to improve the classification performance of histopathological image sections. For instance, Nazeri et al. (2018) introduced a two-stage CNN, where the first stage is a patch-wise network that is

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstine), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 A summary of some classification approaches in digital pathology

Method(s)	Tissue type	References	Application/ image type	Validation	Dataset info
Two-stage CNN	Nuclei	Nazeri et al. (2018)	Breast cancer	Partition data set into training set and test set for classification accuracies and the area under the ROC curve	BACH dataset of 400 breast histopathology images of size 2048 × 1536 pixels with a pixel resolution of 0.42 μ × 0.42 μ
Context-aware learning using ResNet-50 CNN	Nuclei	Awan et al. (2018)	Breast cancer	Training/test sets partition for classification accuracies	BACH dataset
Context-aware model using Representation Aggregation (RA-CNN)	Nuclei	Shaban, et al. (2020)	Colorectal cancer	3-fold cross validation for classification accuracy	139 images (normal, low grade, and high grade) with an average size of 4548 × 7520 pixels obtained at ×20 magnification
Cell-graph approach	Nuclei	Zhou, Graham, et al. (2019a)	Colorectal cancer	3-fold cross validation for patch and image classification accuracy	139 images with an average size of 4548 × 7520 pixels obtained at ×20 magnification
Deep spatial fusion CNN (DSF-CNN)	Nuclei	Huang & Chung (2018)	Breast cancer	10-fold cross validation for classification accuracy	(1) BACH dataset, (2) BIC dataset of 286 high resolution images of size 2048 × 1536 pixels, split into 249 for training and 36 for testing
Weakly supervised feature selection and aggregation	Carcinoma	Wang et al. (2018)	Lung cancer	Classification accuracy	871 histopathology WSIs with lung carcinomas and 68 WSIs of healthy subjects (collected at Sun Yat-Sen University Cancer Centre)
Adaptive Sampling method	Nuclei	Cruz-Roa et al. (2018)	Breast cancer	Detection and classification accuracy—AUC	The Cancer Genome Atlas—National Cancer Institute
Multi-resolution CNN	Adenocarcinoma	Muhammad et al. (2018)	Lung cancer	Accuracy	Lung adenocarcinoma images as a part of the MICCAI 2017 CPM Challenge, with a total number of 10 images of H&E lung adenocarcinoma sample images
EMS-Net	Nuclei	Yang et al. (2019)	Breast cancer	Precision, recall, accuracy, and AUC using 5-fold cross-validation	BACH and BreakHis datasets
Hybrid CNN + LSTM	Nuclei	Yan et al. (2020)	Breast cancer	Sensitivity and accuracy	Breast cancer dataset of 3771 high-resolution histopathology images
Ensemble of DL networks for classification of Biobsy images	Nuclei	Kassani et al. (2019)	Breast cancer	Accuracy, precision, recall, and F1-score	BreakHis, bio-imaging 2015 challenge dataset, and ICIAR 2018 dataset

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstinet), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 (Continued)

Method(s)	Tissue type	References	Application/ image type	Validation	Dataset info
Feature concatenation and ensemble of deep CNNs	Nuclei	Nguyen et al. (2019)	Cervical cancer & multi-organ cancer	Accuracy	(1) 2D Hela dataset, (2) PAP smear dataset, and (3) Hep 2 cell image dataset
DenseNet121-AnoGAN	Nuclei	Man et al. (2020)	Breast cancer	5-fold cross validation for classification accuracy, precision, recall, F1-score, and confusion matrix	BreaKHis dataset of 7909 biopsy images with a dimension of 700 × 460 using four magnification factors
DL-based classification using global labels	Carcinoma	Sun, Xu, et al. (2020a)	Liver cancer	Classification accuracy, precision, recall, F1-score, and ROC curve	Liver cancer WSIs dataset downloaded from The Cancer Genome Atlas (TCGA). A total of 462 liver tissue slide images, containing 79 normal liver histopathological images and 383 liver hepatocellular carcinoma images
HACT-Net	Nuclei	Pati et al. (2020)	Breast cancer	Weighted F1-scores across four test folds.	BReAst Carcinoma Subtyping (BRACS) which consists of 2080 ROIs acquired from 106 H&E stained breast carcinoma whole-slide-images (WSI)
DSNet	Nuclei	Xiang et al. (2022)	Breast cancer & Lung cancer	Classification accuracy, precision, recall, F1- score, ROC curve, and Area Under ROC (AUC)	(1) Camelyon16, (2) The Cancer Genome Atlas Lung Squamous Cell Carcinoma project (TCGA-LUSC), and (3) Early Breast Cancer Core-Needle Biopsy WSI (BCNB)
HistoGAN	Nuclei	Xue et al. (2021)	Cervical cancer & metastatic cancer	Classification accuracy, AUC, sensitivity, specificity	(1) Cervical histopathology dataset, and (2) lymph node histopathology dataset from breast metastatic cancer.

responsible for extracting the most salient features from image patches while the second stage is an image-wise network that can produce image-level class labels. The work introduced by Awan et al. (2018) proposed a context-aware two-stage approach, which consists of two main steps: (1) patch-based deep CNN ResNet-50 which extracts prominent features from image patches, (2) perform image-based classification using the features extracted from overlapped patches with a separate classifier (SVM). They designed their model to capture contextual information between different image patches. Ehteshami Bejnordi et al. (2017) presented a context-aware stacked CNN architecture to classify breast whole slide images. Their model comprises of two stages: first, they trained a CNN to memorize the cellular level features from image patches; and second, they stacked a fully convolutional network on the top of this to permit for joining of global inter-dependency of structures to encourage predictions in nearby regions. A model called Decompose, Transfer, and Compose (DeTraC) (Abbas et al., 2020) has been designed to cope with data irregularity in medical image datasets and was used to classify colorectal cancer sections (see Figure 10). Here, we realized that working on the decomposed dataset or the decomposed classes of a dataset (instead of the original dataset/classes) can provide the deep learning

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstine), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 A summary of some segmentation approaches in digital pathology

Method(s)	Tissue type	References	Application/ image type	Validation	Dataset info
Contour-aware Informative Aggregation method	Nuclei	Zhou, Onder, et al. (2019b)	Multi-organ cancer	Average Jaccard Index and F1-score	MoNuSeg dataset of 2018 MICCAI challenge of 30 images (with 1000 × 1000 pixels) by The Cancer Genomic Atlas (TCGA). The dataset consists of breast, prostate, liver, kidney, colon, bladder, and stomach containing both benign and malignant cases
Two-stage learning method with DLA	Nuclei	Kang et al. (2019)	Multi-organ cancer	Aggregated Jaccard Index (AJI), precision, recall and F1 score	(1) TCGA dataset of 30 Images with 1000 × 1000 pixels and more than 21,000 nuclei, and (2) TNBC dataset of 50 H&E stained images with 512 × 512 pixels and 4022 nuclei
Self-supervised approach using tile magnification level	Nuclei	Sahasrabudhe et al. (2020)	Multi-organ cancer	AJI, Average Hausdorff Distance (AHD), and Average Dice Coefficient (ADC)	MoNuSeg dataset.
Mask-RCNN as backbone with SPCN and watershed method	Glioma	Xie et al. (2019)	Brain cancer	Dice score	Dataset of 15,668 × 583 pixels annotated H&E stain images extracted from a set of Glioblastoma and Lower Grade Glioma images. The dataset set was presented as a part of MICCAI 2017 segmentation challenge (MSC) and 2018 Data Science Bowl (DSB)
Residual block + attention decoder	Glioma	Lal et al. (2021)	Liver cancer	F1-score and Jaccard index (JI)	KMC liver dataset of 80 H&E stained liver histopathology images collected at KMC Mangalore, MAHE, Manipal, India, and Kumar dataset (multi-organ nuclei segmentation) collected by the Indian Institute of Technology, Guwahati
Deep Learning model + concave point detection	Nuclei	Wan et al. (2020)	Multi-organ cancer	Dice similarity coefficient (DSC), mean accuracy (ACC mean), pixel accuracy (ACC pixel), and mean intersection over union (IoU)	(1) TNBC dataset, (2) cancer genome atlas (TCGA) archive, and (3) two other datasets acquired from the China Japanese Friendship Hospital, Beijing
Deep Learning model + Contour-aware	Gland & Nuclei	Chen et al. (2017)	colorectal and breast cancer	F1 score, object-level Dice index, and Hausdorff distance	(1) Dataset for Gland Segmentation (GlaS) Challenge dataset and (2) Nuclei segmentation challenge
Hover-Net	Nuclei			Ensemble Dice and AJI	

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstinet), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 (Continued)

Method(s)	Tissue type	References	Application/ image type	Validation	Dataset info
		Graham et al. (2019)	colorectal cancer		Colorectal nuclear segmentation and phenotypes (CoN-SeP) dataset, 6 consisting of 41 H&E stained images
Multi-scale Adaptive Nuclei Analysis (MANA)	Nuclei	Salvi & Molinari (2018)	Multi-organ cancer	Validation using recall, precision and F1-score	H&E stained images from six different organ tissues, where images have been collected/digitized at the Molinette Città della Salute University hospital (Torino, Italy)
RIC-Unet (residual-inception-channel attention-Unet)	Nuclei	Zeng et al. (2019)	Multi-organ cancer	Dice, F1-score, and AJI	TCGA (The Cancer Genomic Atlas) dataset
SAMS-Net	Nuclei	Graham & Rajpoot (2018)	Multi-organ cancer	Traditional an ensemble Dice coefficients	The dataset was supplied as part of the computational precision medicine (CPM) nuclei segmentation challenge (at the MICCAI 2017)
Nucleus-boundary model using FCN	Nuclei	Cui et al. (2018)	Multi-organ cancer & Breast Cancer	Precision, recall, F1-score, false detection rate (FDR), missing detection rate (MDR), over segmentation rate (OSR), and under-segmentation rate (USR)	(1) Multiple organ H&E-stained images. (2) Breast Cancer histopathology image
super-pixels using linear iterative clustering algorithm	Cervical	Sornapudi et al. (2018)	Cervical Cancer	Validation using precision, recall, accuracy, dice similarity coefficient (DSC), F1-score, and JI	133 digitized images of cervical cancer histopathology images
SC-CNN	Nuclei	Sirinukunwattana et al. (2016)	Colorectal cancer	Precision, Recall, F1-score, and Median Distance with 2-fold cross-validation	100 *H&E* stained histopathology images of 500 × 500 pixels
Multi-layer pseudo-supervision	Nuclei	Han et al. (2022)	Lung cancer and Breast cancer	FwIoU, MIoU, and ACC	54 lung cancer WSIs with four tissue categories labeled and 151 H&E stained whole slide images of breast cancer with 5 annotated classes
A dense dual-task network for tumor-infiltrating lymphocyte detection and segmentation	Nuclei	Zhang et al. (2022)	Breast Cancer	Validation using precision, recall, accuracy, dice similarity coefficient (DSC), F1-score, and AJI	840 images 100 × 100 pixels and 151600 × 1600 images
Dual Segmentation	Nuclei	Razavi et al. (2022)	Breast Cancer	Dice similarity coefficient (DSC) and F1-score	

(Continues)

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstinet), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 (Continued)

Method(s)	Tissue type	References	Application/ image type	Validation	Dataset info
of Mitoses and Nuclei Using Conditional GANs					TUPAC16, ICPR12, and ICPR14 with 256 × 256 patches

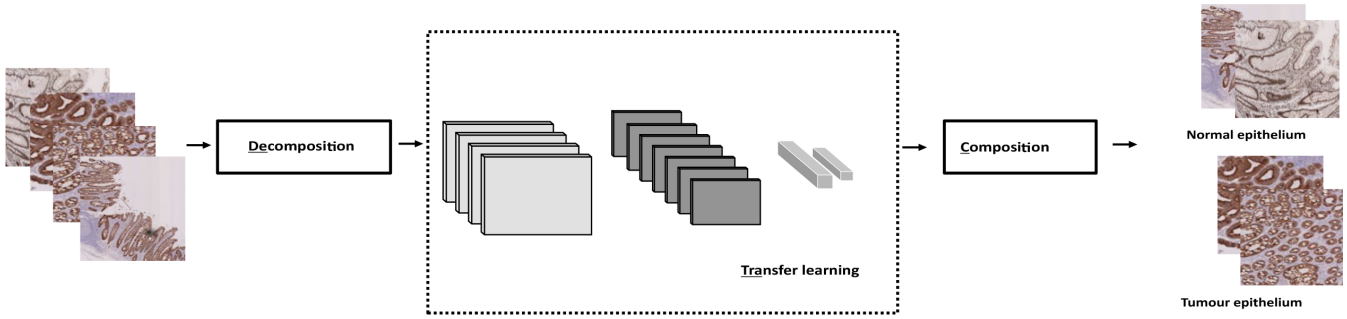

FIGURE 10 The architecture of decompose, transfer, and compose (DeTraC) model. DeTraC has three stages: In the first stage, a class-decomposition method is trained to divide the classes of the original dataset into subclasses, resulting in a new dataset (decomposed dataset). The class decomposition is used to simplify the complexity of the local structure of the image dataset. In the second stage, a pre-trained network is used to classify the decomposed dataset. Finally, the final classification of the original dataset is refined using error-correction criteria applied to a Softmax layer.

models (e.g., ImageNet pre-trained models) with the capability of learning the boundaries between the classes and can help in improving the transferability of features and find more specialized features. Consequently, the authors' reflection on DeTraC is that the method can be adopted in any deep learning architecture (as a generic approach) to cope with complex datasets, especially with the limited availability of training images. A two-stacked CNNs approach was used in (Shaban et al., 2020). The method can learn the local representation of histopathology image and encodes it to a high dimensional features, then the learned features are aggregated considering their spatial pattern using a second CNN.

A deep fusion network to capture the spatial relationship between high resolution histopathology image patches was also proposed in (Huang & Chung, 2018). First, residual network was adapted to learn visual features from cellular-level to large tissue organization. Second, a deep fusion network was developed, which modeled the inconsistent construction of distinctive features over patches and learned to rectify the predictions of residual network. Kohl et al. (2018) studied different approaches for transfer learning and classification of microscopy images using patch-based mechanism. The work conducted in (Yan et al., 2020) proposed a hybrid convolutional and recurrent neural network for the classification of breast cancer histopathological images. Their method introduced the short-term and long-term spatial correlations between image patches using Bidirectional Long Short-Term Memory (LSTM) network. The model starts by capturing feature representations from image patches of histopathology image, then they fed the extracted features to bidirectional LSTM for generating the spatial correlations between feature representations. Koné and Boulmane (2018) proposed a hierarchical system of CNN which automatically classifies histopathological patches into several classes. A cell-graph approach based on CNN was proposed in (Zhou, Graham, et al., 2019a), which transforms each histopathology image to a graph, where nodes are represented by nuclei within the original image and cellular interactions are presented as edges between these nodes based on node likeness. The network utilizes nuclei local features and spatial location of nodes to boost the performance of their work.

DenseNet121-AnoGAN (Man et al., 2020) has been proposed for classification of breast histology images into two classes: benign and malignant. The proposed method is divided into two parts: screening of mislabeled patches with unsupervised anomaly detection using generative adversarial networks (AnoGAN) and extracting multi-layered features from discriminative patches using densely connected convolutional networks (DenseNet). Another work proposed by Hirra et al. (2021) developed a patch-based deep learning method called Pa-DBN-BC to detect and categorize breast

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstinet), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

cancer histopathology images. An unsupervised pre-training and supervised fine-tuning phase is used to extract features. The network extracts features from image patches automatically. Patches from histopathological images are then classified using logistic regression. The model takes the features extracted from the patches as input, and the outcome is presented as a probability distribution of two probability values to differentiate between positive sample (cancer) or negative sample (no cancer). HistoGAN has been proposed by (Xue et al., 2021) to improve classification of histopathology images. It's based on conditional GAN model for generating realistic histopathology image patches depending on class labels. They developed a synthetic augmentation approach which selectively adds HistoGAN-produced synthetic image patches. The framework ensures the quality of synthetic augmentation by picking synthetic images depending on the confidence of their assigned labels and their feature closeness to actual labeled images. They demonstrated that using HistoGAN produced images with selective augmentation improves classification performance significantly.

A transfer learning based approach has been proposed for breast cancer classification by Ahmad et al. (2021). The approach uses patch selection technique to classify histopathology images using transfer learning. Their model works on extracting patches from whole-slide images and fed them into CNN for feature extraction. Discriminative patches are then selected based on the extracted features and then, using pre-trained Efficient-Net, features are extracted which are utilized to train SVM classifier. Similar work has been introduced by Li, Wu, and Wu (2019c) which proposes model for classification of breast histopathology images using (1) patches screening method based on clustering algorithm which aggregates image patches based on their phenotypes and (2) CNN to pick more discriminative patches. Roy et al. (2019) developed a patch-based classifier (PBC) that uses CNN to classify histological breast images. There are two modes of operation for the proposed system: one patch in one decision (OPOD) and all patches in one decision (APOD). The suggested PBC uses OPOD mode to predict the patch's class label. If the class label for all extracted patches is the same as the image's class label, the result is deemed accurate classification. In another option, APOD, the class label of each extracted patch is extracted as in OPOD, and the image's class label is decided using a majority vote technique. Another work (Sun, Xu, et al., 2020a) proposed a deep learning method using only global labels for classification of liver histopathology images. Due to complex features and limited annotated training samples for liver histopathology images, patch-level features are extracted from the images and then used by transfer learning to be combined with multiple-instance learning to generate image-level features for final classification.

Chennamsetty et al. (2018) introduced three CNNs trained on different pre-processing regimes to form an ensemble. The purpose of this work was to prove that there is no single architecture nor a pre-processing regime that can provide a better performance. Yang et al. (2019) proposed a CNN ensemble model called Ensemble of Multi-Scale Network (EMS-Net) utilized to classify hematoxylin–eosin stained breast histopathological images. EMS-Net allows to extract features using multiple pre-trained CNN models at multi-scale and select the optimal subset of the fine-tuned deep models. Likewise, the work conducted in (Dhivya & Vasuki, 2019) presents an approach for identification and classification of tumor in breast histopathology image based on ensemble classification of pre-trained deep CNN architectures (LeNet, AlexNet and VGGNet-16). The construction of classifier model is done by fine-tuning the pre-trained weights of these models separately. Deciding the class label of the full model is done by applying majority voting over the class probabilities obtained from the three pre-trained models. H. Kassani et al. (2019) introduced an ensemble of deep learning models for automatic binary classification of breast histopathology images. The proposed model has been constructed based on three pre-trained CNNs (VGG19, MobileNet, DenseNet). The ensemble model was used as feature extractor and the extracted features are then fed into a multi-layer preceptron classifier to carry out the classification task. Due to the tremendous image sizes and the limited availability of training examples, Makarchuk et al. (2018) adapted a CNN model to classify image patches to increase effective sample size and apply ensembling to build prediction for original images. Both ResNet and DenseNet were used to extract features and XGBoost was used for the final classification. Nguyen et al. (2019) proposed a feature concatenation and ensemble method to combine several CNNs with different depths and structures in order to improve the classification accuracy of biomedical image classification. The proposed model consists of three base models (Inception-v3, ResNet152 and Inception-ResNet-v2) that are pre-trained using transfer learning and a fourth model that works as multi-feature-extractors model. This feature descriptor takes the three feature maps extracted from the three base models and concatenate them into a longer feature vector. An ensemble learning technique is used to ensemble the four feature maps (three from the base models and one from the multi feature descriptor). The work presented in (Vang et al., 2018) provides a deep learning model that performs patch level classification using Inception V3. To achieve the image level prediction, the patch level predictions extracted from the previous stage (Inception V3) are then passed to an ensemble fusion architecture comprising majority voting, gradient boosting machine (GBM), and logistic regression. Then to discover the dominating structural patterns among

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstinet), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

normal image patches, a fully convolutional auto-encoder is utilized. One-class support vector machine and one-layer neural network are used to recognize and evaluate patches that do not share the features of the regular population.

To enhance the structural description of the tissue, hierarchical cell-to-tissue-graph (HACT) representation has been proposed by Pati et al. (2020). Their approach consists of two types of graphs. First, a low-level cell-graph which captures cell morphology and interactions. Second, a high-level tissue-graph, which captures morphological characteristics and spatial distribution of tissue sections, and cells-to-tissue hierarchies, which incorporate the relative spatial patterns of cells in reference to tissue distribution. Additionally, a hierarchical graph neural network (HACT-Net) is introduced to map the HACT presentations into histological breast cancer sub-types. Another work (Shen & Ke, 2020) introduced a Deformable Conditional Random Field (DCRF) model to learn the offsets and weights of nearby patches in a spatially adaptable way from whole-slide images. Also, instead of using overlapped patches, they used adaptive modified offsets in a WSI to locate patches with more robust feature representations.

An interactive whole-slide image diagnostic system has been developed by Chen et al. (2020) for thyroid frozen sections based on the doubtful areas chosen by pathologists. Their system depends on generating feature patterns for doubtful regions through obtaining and fusing patch features using deep neural networks. The feature representations are then used to assess four classifiers and three supervised hashing algorithms for region classification and retrieval. The work conducted by Li, Li, Sisk, et al. (2021c) proposed a multi-resolution multiple instance learning (MIL) model for fine-grained grade prediction that uses important feature map representations to detect doubtful image regions. Their model can be trained end-to-end using only slide-level labels, rather than depending on region- or pixel-level annotations. The model is based on the WSI dataset of large-scale prostate biopsy. Another work developed by Kanavati et al. (2021) proposed a deep learning model for classifying transbronchial lung biopsy (TBLB) WSIs into one of the lung carcinoma sub-types: adenocarcinoma (ADC), squamous cell carcinoma (SCC), small-cell lung cancer (SCLC), and non-neoplastic. Their model is composed of a CNN and an RNN for the task of obtaining patch predictions and aggregating patch predictions into one single WSI classification, respectively.

Because most MIL approaches are based on the independent and identical distribution assumptions, they ignore the connection between distinct instances. The work introduced by Shao et al. (2021) presented a new framework termed correlated MIL and offered a proof for convergence to overcome the issue. They created a Transformer-based MIL (TransMIL) based on this framework, which looked at both morphological and contextual information. With outstanding visualization and interpretability, the suggested TransMIL can cope with unbalanced/balanced and binary/multiple categorization. Another work introduced by Chen, Liang, et al. (2021b) developed a broad framework that uses unit stochastic selection and attention fusion to automatically diagnose various forms of WSIs. A unit on a histopathology slide might be a patch or a cell on a cytopathology slide. Their approach begins with training a unit-level CNN to fulfill two objectives: creating feature extractors for the units and calculating the non-benign probability of each unit. Then, based on CNN's observations, they utilize a stochastic selection technique to select a small segment of units which are considered to be non-benign, termed to as Units Of Interest (UOI). The attention mechanism is then used to merge the UOI representations into a fixed-length description for the WSI's diagnostic.

Li, Chen, Huang, et al. (2021b) argue that further improvement can be achieved by combining pixel-level and image-level annotation. This is problematic in computational pathology because the high resolution of WSIs makes end-to-end classification model training challenging. To deal with this, they developed a hybrid supervised learning system for pathology high resolution images with enough image-level coarse annotations and a few pixel-level fine labels. When used for training patch model, this approach can improve produced pixel-level pseudo labels with the help of coarse image-level labels. Another work (Sharma et al., 2021) introduced Cluster-to-Conquer (C2C), an end-to-end architecture that divides a WSI's patches into k-groups, picks k' patches from each grouping for training, and employs an adaptive attention mechanism to give final slide prediction. They have shown that splitting a WSI into clusters improves model training by exposing it to a variety of discriminative characteristics retrieved from the patches. Due to the difficulty in attaining good classification performance with minimal labeled samples, a deep transferable semi-supervised domain adaptation model (HisNet-SSDA) has been suggested by Wang et al. (2022) for classification of histopathology WSIs. Their approach depends on the knowledge transferred from a highly labeled source domain to a partially labeled target domain via semi-supervised domain adaptation. First, a pre-trained network called HisNet is utilized to extract high-level features from randomly chosen patches in the source and target domains. The characteristics of the two domains are then matched using a multiple weighted loss functions criteria with a new manifold regularization term in semi-supervised domain adaptation. Finally, the estimated probabilities of sampled patches are combined for producing the final image-level classification.

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstinet), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3.2.2 | Tissue-level segmentation

Deep Contour-aware Network (DCAN) (Chen et al., 2017) was introduced to address different issues (such as large appearance variation, existence of strong mimics, and serious degeneration of histopathological structures) during automated detection and segmentation. DCAN is a unified multi-task learning framework that takes advantages of multi-level contextual features and auxiliary supervision to alleviate the problem of vanishing gradient. DCAN has two branches that are working simultaneously where one branch is used for object segmentation while the other is used to achieve contour detection. Each branch is based on different weight parameters during the up-sampling process. Form each branch, a probability map is generated which leads to output of either segmentation of objects or contour information.

As a recent introduction to computer vision, Vision Transformers (Dosovitskiy et al., 2020) have presented a technique that utilizes self-attention to overcome limitations presented by inductive biases inherit in convolutional networks in an efficient way, see Figure S2 in the Supplementary Materials. The spatial limitations imposed by convolutions are not present in the visual representations produced from self-attention components. Instead, they can learn the most appropriate inductive biases based on the task and the stage of the pipeline at which the layer is located. Efforts have been made to use CNNs' localization and transformers' global awareness, resulting in hybrid models. TransUnet (Chen, Lu, et al., 2021a; Lin et al., 2022) is a proposed model that enhances the locality of convolution operations by including a transformer that operates on the feature maps created by an encoding convolutional network. The hybrid CNN-Transformer architecture preserves the advantages of both transformers and CNNs by exploiting the feature rich feature maps of a CNN and the global context encoded by a transformer. This enables better performance in medical image segmentation. State-of-the-art performance was attained by hybrid models, often at the expense of high model complexity and, in most cases, lengthier training times.

Figure 11 shows results of some Hybrid and CNN models on the Gland Segmentation dataset GLAS. DeepLab and Unet are two convolutional networks based on the Unet architecture. DeepLap utilizes atrous convolution with upsampled filters to extract dense feature maps and capture long-range context. These results illustrate the performance of different segmentation approaches in histology and demonstrate the feasibility of deep learning applications on histological imaging. Among the different state-of-the-art deep learning architectures, the authors realize that transformer-based models can help in dealing with the missing of contextual information in histopathology images, due to their ability in encoding global information. Also, it might not be enough to only rely on global information extracted from transformer-based models but there might be a need for other kinds of information (including local information).

Another application field of these models is epithelium tissue segmentation. PESO (Bulten et al., 2019) introduced a novel deep learning algorithm for segmenting epithelial tissue in H&E stained prostatectomy slides using

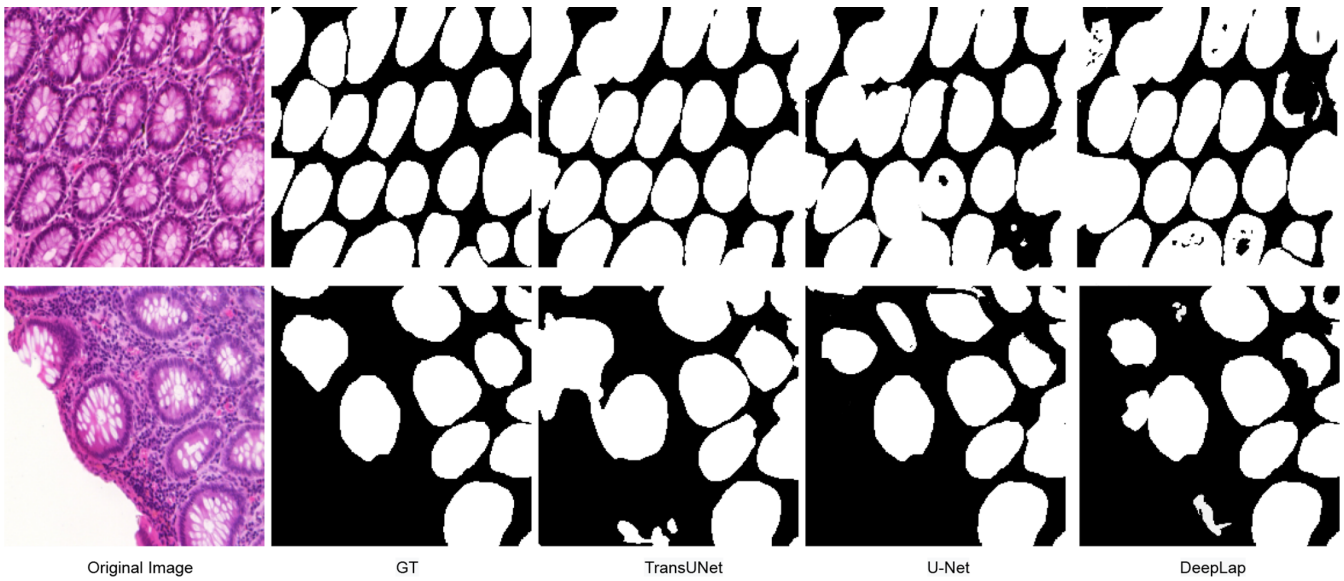

FIGURE 11 Examples of the original input image, GT (ground truth), and the results of different segmentation techniques applied to GLAS dataset

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstinet), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

immunohistochemistry (IHC) as the reference standard. In comparison to human outlining on H&E slides, they employed IHC to generate a precise and objective ground truth, particularly in regions with high-grade PCa. To highlight epithelial features, 102 tissue slices were treated with H&E and then restained with P63 and CK8/18 IHC markers. Following that, each couple was co-registered. A U-Net was trained to segregate epithelial structures in IHC using a subset of preprocessed IHC slides. Second, this network was applied to the remaining slides to produce a reference standard for training a second U-Net on H&E. Their method segmented whole glands as well as individual tumor epithelial cells with high accuracy.

Iizuka et al. (2020) also used convolutional neural networks (CNNs) and recurrent neural networks (RNNs) trained using biopsy histopathology whole-slide images (WSIs) of the stomach and colon. Their models learned to differentiate between adenocarcinoma, adenoma, and non-neoplastic epithelial tumors. Extracted tiles were manually classified into one of three types then features are extracted from the convlutional-based inception-v3 backbone to train an RNN network. Their models attained area under the curves (AUCs) of up to 0.97 and 0.99 for gastric adenocarcinoma and adenoma, respectively, and 0.96 and 0.99 for colonic adenocarcinoma and adenoma, respectively, on three independent test sets. These application results demonstrate the models' generalizability and their great potential for use in a realistic histopathological diagnostic workflow system.

3.2.3 | Cellular-level segmentation

Detection techniques using deep learning models have demonstrated great results in the histopathology image analysis. With the aim of automatic detection of immune cell in immunohistochemistry images, CNN (Chen & Chefdhotel, 2014) has been trained on the immune cell marker image channel in order to produce the probability map of the immune cell locations. First, the RGB image was unmixed in order to produce the immune cell markers components using sparse color unmixing. Second, the immune cell markers images are applied as input to the CNN for learning each pixel into positive (i.e., immune cell) or negative (i.e., the rest). In (Sirinukunwattana et al., 2016), a Spatially Constrained Convolutional Neural Network (SC-CNN) has been proposed with the aim of detecting nuclei in histopathology images. The main contribution of the SC-CNN approach is the integration of the topological domain (i.e., spatial domain) in the CNN architecture. This is by training a spatially constrained regression approach, via a parameter estimation and a spatial constraints layers, to predict the center location of each nucleus in the input patch and provide a probability map of the given locations. The work introduced by Zeng et al. (2019) proposed a U-net network called RIC-U-net (residual-inception-channel attention-U-net) for nuclei segmentation. RIC-U-net was inspired by residual blocks, multi-scale and channel attention to provide more accurate segmentation results. SAMS-Net (Graham & Rajpoot, 2018) is proposed to tackle the trouble in segmenting nuclei with coarse chromatin appearance which leads into a low Hematoxylin intensity. SAMS-Net is a stain-aware multi-scale convolutional neural networks that provides prominent weight to misclassification inside the nucleus with a small Hematoxylin intensity and to misclassification outside the nucleus with a high Hematoxylin intensity. This happens by introducing a number of predefined weight maps that are aware of the intensity of Hematoxylin within each image. Cui et al. (2018) presented an automatic end-to-end deep neural network model for nuclei segmentation. They introduced a nucleus-boundary model to anticipate nuclei and their boundaries at the same time using FCN. The model outputs two maps, one for nuclei and the other for boundary information.

The work introduced in (Wan et al., 2020) developed an image-based method for nuclei segmentation using a deep learning model encoded with a concave point detection algorithm. They used atrous spatial pyramid pooling U-Net (ASPPU-Net) to generate multi-scale nuclei features and contextual information. This is done without decreasing the spatial resolution of feature maps. The use of accelerate concave point detection technique helps in an effective and accurate segmentation even for highly overlapped nuclei. Lal et al. (2021) presented NucleiSegNet for nuclei segmentation of H&E stained liver cancer histopathology images. Their model consists of three blocks. First, a residual block to extract efficient semantic maps. Second, bottleneck block and finally, attention decoder block which utilizes attention mechanism for object localization and improves the performance of their model by minimizing false positives.

Xie et al. (2019) introduced a deep learning framework for automatic nuclei segmentation. Their frameworks is based on Mask-RCNN as backbone with structure-preserving color normalization (SPCN) and watershed for pre- and postprocessing. A two-stage learning framework and Deep Layer Aggregation (DLA) is proposed by Kang et al. (2019) for nuclei segmentation. The two-stage learning framework has been constructed by stacking two U-Nets models. The first stage estimates the morphological information of nuclei while the second stage provides a final fine-grained

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstinet), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

segmentation map. Moreover, an extension is done to U-Nets by adding DLA. This is applied by concatenating features across different levels.

Zhou, Onder, et al. (2019b) proposed a Contour-aware Informative Aggregation Network (CIA-Net) with a multi-level information aggregation module. The network helps in capturing the spatial dependencies information between nuclei and contour. This is done by aggregating task-specific features bidirectionally.

The work conducted in (Sornapudi et al., 2018) introduced a deep learning-based nuclei segmentation approach based on capturing localized information by generating super-pixels using linear iterative clustering algorithm and training with a convolutional neural network. Salvi and Molinari (2018) developed a fully Multi-scale Adaptive Nuclei Analysis (MANA) for nuclei segmentation in different tissues and magnifications. The aim of their work is to automatically detect nuclei in H&E stain images by applying object-based thresholding, area-based correction and nuclei separation (watershed transform). A novel CNN called Hover-Net for simultaneous nuclei segmentation and classification is introduced in (Graham et al., 2019). Their model captures the instance information supported with the vertical and horizontal distances of nuclei pixels. These distances are used to split the clustered nuclei yielding accurate segmentation specially for overlapped instances. Their network predicts the type of nucleus using devoted up-sampling branch.

Mesmer (Greenwald et al., 2022) is a deep learning pipeline that uses PanopticNet and TissueNet to accurately segregate nuclear and whole-cell data. Semantic heads for nuclear and whole-cell segmentation are coupled to a common backbone and FPN in Mesmer's PanopticNet model. For each cell, Mesmer needs a nuclear image (DAPI) and a membrane or cytoplasm image (CD45 or E-cadherin) to describe its form. So the PanopticNet model gets inputs normalized (for robustness) and tiled into fixed-size patches (for processing images of various dimensions). In the last step, the model predicts the centroid and border of each nucleus and cell in the image. The final instance segmentation mask for each nucleus and cell in the picture is created using the centroid and border predictions.

Histological structures offer a different challenge in the computer vision field. Their abnormal, seemingly random, structural boundaries make it harder to transfer weights or learning from solutions that have done on other problems. Recent works have tackled this challenge by introducing methods that are more globally aware, using a higher receptive field to capture these structures at scale. A solution that can be effective but runs into limitations as these structures grow larger and occupy more of an already large image size. Consequently, transformer-based models with their self-attention based aggregation can mitigate the large image size by having attention applied through different overlapping windows in the image. The local objects can identified in each window and cross-window attention can lead to higher recognition of these larger structures. Depending on the type of attention calculation used (Cross, Axial), more efficient models can be utilized to effectively model large histological slides.

3.3 | Deep learning in precision oncology

The histopathological features of pathological tissue reflect a complex interplay between genetic and epigenetic mechanisms that determine the overall protein expression patterns and functional status. Recent data demonstrated strong correlation between the underlying molecular profile of tumors and the morphological features obtained from WSIs of the tumor tissues. In (Cifci et al., 2022), analytic details with a comprehensive review of recent work (e.g., papers published between 2017 and 2021) on deep learning has been provided to show that genetic alterations in tumor tissues are predictable using histopathological images with AI-based methods. There is also direct correlation between morphological features such as tumor grade and types and clinical behavior and response to therapy (Rakha et al., 2021). For example, in (Wulczyn et al., 2020) the direct prediction of clinical outcomes based on morphological features of histopathology images has been explored using deep learning. This study shows promising benefits to provide significant prognostic information using deep learning in multiple cancer types. In (Takamatsu et al., 2022), a deep learning model based on CNN and random forest methods was proposed to extract cancer tiles from WSIs of colorectal cancer cases and then re-label them lymph node metastasis (LNM) status to predict LNM. A multitask deep learning (Bychkov et al., 2022), based on CNN which is supervised by both patient outcome and biomarker status, has been proposed to outcome prediction in breast cancer using H&E images. The study also showed that the performance of the deep learning model could improve with the integration of histologic type and grade of differentiation. Consequently, detailed and comprehensive computational analyses of pathology images assessing myriads of features related to the tumor cytological features, growth pattern and architecture, and tumor microenvironment far beyond what can be achieved using eyeballing assessment by pathologists or conventional image analysis techniques utilizing individual features are likely to provide wealth of invaluable information. Linking such information to clinical behavior and response to therapy can

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstine), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

identify signatures predictive of clinical behavior and response to specific therapy equivalent or more accurate to the current prognostic and predictive tools that rely on molecular alteration or single morphological features and in cost-efficient way (see Figure 12). However, achieving such goals require large and well annotated cohorts of pathology images with detailed treatment and outcome information and robust image annotated protocols.

4 | CURRENT RESEARCH DIRECTIONS

4.1 | Histopathology image registration

In modern digital pathology techniques, sequential imaging enables image acquisition using multiple stained slides from the same histopathological tissue section. This allows an order of magnitude increase in the number of molecular markers to be imaged for the same tissue section. Consequently, automatic image registration (see Figure 13) is an important initial step to align these multi-biomarker tissue slides for subsequent analysis.

Many histological image registration models have been proposed that achieve automated alignment of multi-biomarker images, which we classified here into two categories: intensity-based (Venet et al., 2019; Wodzinski & Skalski, 2019) and feature-based (Awan & Rajpoot, 2018; Schultz et al., 2019) models. Intensity-based models usually require significant effort to minimize the stain variations or to design robust multi-modality image similarity measurements. For instance, Feuerstein et al. (2011) proposed the use of block-face images (e.g., a new set of images that can only be acquired during histological sectioning) as an external reference for registration. This method has shown less sensitivity to the histopathology artifacts, but the availability of such references could be limited. Bagci and Bai (2010) developed an intensity-based registration approach to reconstruct a mouse brain from its associated histological slides. This was achieved by firstly normalizing intensity variations, followed by extracting a set of high fidelity features, and then selecting the best reference slide using an iterative image registration method. A staining-invariant approach (Schwier et al., 2013) was developed for the registration of histological slides. It combined rigid and elastic approaches to firstly align consecutive slides coarsely and followed by a nonlinear registration. However, the method was designed to work on a low magnification level, in which higher details were ignored during the registration process. On the other hand, the feature-based models are more robust where points/landmarks are detected based on the topological structures in the images. A typical point-based registration consists (a) feature points extraction to locate significant structures in the moving (source) and reference (target) tissue images; (b) estimate a point-wise correspondence by

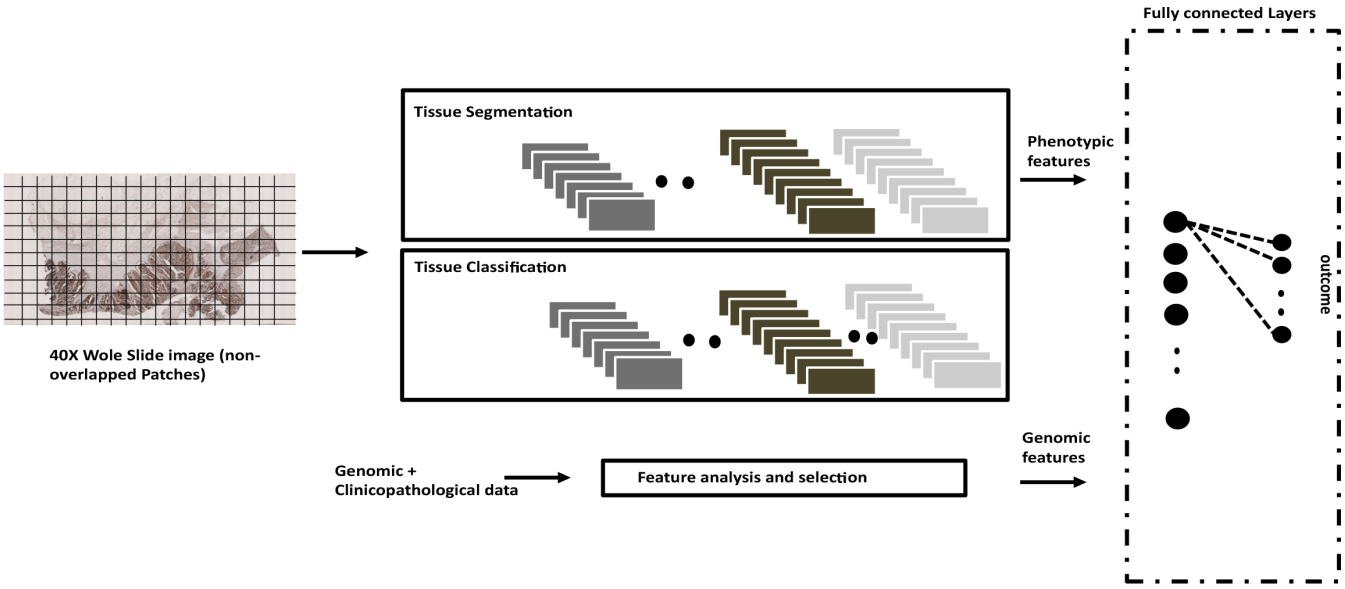

FIGURE 12 The integration of both structured and unstructured data to build a robust diagnostic/prognostic tool. Tissue classification/segmentation components can be used to extract digital biomarkers (or phenotypic features) directly from images while the feature analysis and selection can be used to extract genomic and clinicopathological features. The last fully connected network can be used to learn the relationships between the phenotypic and genomic features to improve the performance of the system.

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstine), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

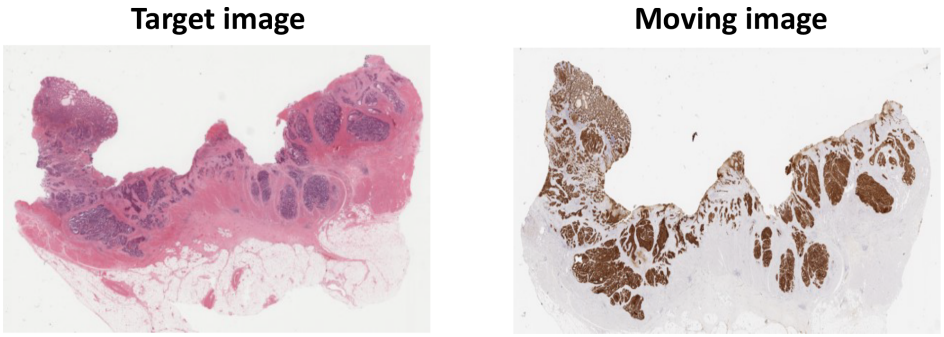

FIGURE 13 Two different stained WSIs of the same tissue. Image registration is an important initial task that can help in visually comparing these two images of the same issue but of different biomarkers. The main task is to determine a transformation that maximizes the similarity between the target image and its associated moving image.

maximizing the similarities in the point space and removing outliers; and (c) point-set registration to find the optimal transformation parameters for transforming the moving image to the reference image space. A number of feature-based image registration methods have been developed for histopathology images, which is considered to be a reasonable choice when compared with the intensity-based methods. High-level features based on a sparse convolutional autoencoder were used as the input to the image registration method for coping with the stain variations in histology images (Awan & Rajpoot, 2018). A segmentation-based image registration method (Borovec et al., 2013), based on a set of features detected by Speeded Up Robust Feature (*SURF*) method (Panchal et al., 2013), was applied to multiple stained histopathology images. The image registration method based on B-Splines was used for multi-stained histopathology sections using common information of the global structure of the images (Obando et al., 2017). Likewise, local feature descriptors based on SURF method (Rossetti et al., 2017) were used in a multi-stage image registration method with multi-resolution mapping to align whole slide images. With the aim of aligning large multi-stained histopathology images, Obando et al. (2017) proposed a semi-automated approach to extract feature points from a down-sampled version of the image, and subsequently applied the estimated transformation (calculated by B-Splines) to the full resolution image. However, the developed approach assumed that no tissue folding and losing was occurring in the moving image which may not be valid in practice.

More recently, the Automatic Non-rigid Histological Image Registration (ANHIR) challenge (Borovec et al., 2020) was organized as a part of the ISBI 2019 to compare the performance of different image registration models on several different histopathology images (see Figure 1). Most of the submitted models to this grand challenge were using similar classical techniques. They have been designed as a multi-step pipeline, which is started with a coarse pre-alignment that is followed by a non-rigid registration step for fine-tuning the registration results. Surprisingly, only one model was designed based on CNNs. The CNN-based model was performing very well in terms of accuracy and speed although not as well as the best-performing methods.

4.2 | Scarce annotations in digital pathology

The main issue that can be considered as a significant obstacle to computational pathology is the scarcity of supervised information that can be provided by professionals in order to build a supervised deep learning approach, this plays an important role in the final diagnosis of a deep learning model. Here in this section we review the current research directions in coping with this problem using the stat-of-the-art deep learning models in several applications.

4.2.1 | Weakly supervised learning methods

Training effective and highly accurate deep learning model as a part of histopathology image analysis workflow requires full supervision to achieve many tasks, including classification, detection, and segmentation of histopathological structures in the images. However, due to the large size of histopathology images, the availability of dense annotations is quite limited and therefore image datasets are usually composed of large coarsely annotated images.

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstine), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Consequently, it is highly beneficial to develop weakly supervised deep learning models that can predict class-level or pixel-level labels without dense annotations.

(Wang et al., 2018) proposed a weakly supervised mechanism for effective classification for whole slide lung cancer images. Their method utilizes patch-based fully convolutional network (FCN) for discriminatory block retrieval. Also, they introduced context-aware feature selection and aggregation to create all inclusive holistic WSI descriptor. A practical and self-interpretable invasive cancer diagnostic approach has been proposed by Li, Radulovic, et al. (2019b) due to the heterogeneity of cancer cell proliferation as well as a range of benign tissue generative lesions in breast cancer histopathology images. The suggested technique exploits contrasting features between normal and malignant images in a weak-supervised way with little annotation information and generates a probability map of anomalies to test its reasoning. A weakly supervised learning technique is used to identify well, intermediate, and poorly differentiated stages of cervical cancer with a multi-layer hidden conditional random fields (MHCRFs)-based cervical histopathology image classification (CHIC) model (Li, Chen, et al., 2019a). Their method starts with extracting deep learning features are extracted from histopathology image patches. The extracted features are then used by neural network, support vector machine, and random forest classifiers to generate patch-level classification probabilities. Then to generate unary and binary potentials, effective classifiers are chosen. Finally, MHCRF model predicts image-level classification outcomes utilizing the produced potentials.

The work introduced by Campanella et al. (2019) demonstrates a deep learning method based on MIL that employs only the reported diagnoses as labels for training, eliminating costly and time-consuming pixel-by-pixel manual annotations. MIL is used to train deep neural networks in the created framework, resulting in tile-level feature representation. These representations are then utilized to integrate the information throughout the entire slide and give the final classification result using a recurrent neural network (RNN). The work conducted by Zhao et al. (2020) suggested a MIL technique for histopathology image classification based on deep graph convolutional networks and feature selection (FS-GCN-MIL). The suggested technique is made up of three parts: feature extraction at the instance level, feature selection at the instance level, and bag-level classification. They created a self-supervised learning method based on a variational auto-encoder and generative adversarial network model to train the feature extractor (VAE-GAN). They also developed a new strategy for selecting discriminative instance characteristics at the instance level. In addition, they used a graph convolutional network (GCN) to learn the bag-level representation and subsequently classify it. Another work presented by Wang et al. (2020) suggested employing graph convolutional networks (GCNs) to describe the spatial organization of cells as a graph to easily capture the growth and structure of tumor cells in a weakly supervised technique for grade classification in tissue micro-arrays (TMA). A contrastive predictive coding (CPC)-based self-supervised technique is used to learn the morphometry of each cell. Courtiol et al. (2019) trained a deep learning network using global slide-level labels based on survival scores. The network was trained to assign survival score to each tile and aggregate a global label on the slide based on the most informative tiles. Tiles were assessed after an encoding process through a CNN backbone that projected the image tiles into an features space that summarizes the important features of the tile.

DeepMIL methods have recently showed great promise in weakly supervised field. In Lu et al. (2020) DeepMIL was used to demonstrate that pre-training a feature extractor using contrastive predictive coding on histopathology tiles increases downstream malignancy classification performance when compared with an ImageNet pre-trained feature extractor, but this has not been tested on entire WSIs. They presented a two-stage semi-supervised strategy in which they first pre-train the feature network using self-supervised feature learning on every single instance in the dataset. Then a margin-based loss function during the second stage to supervise the learning process.

Some work in privacy-preserving federated learning has been done using hundreds of gigapixel entire slide images from several universities. These studies accommodate for the challenges given by the absence of extensive annotations in the majority of real-world whole slide histopathology datasets and illustrate how federated learning can be linked with weakly supervised multiple instance learning to handle binary and multi-class classification tasks using just slide-level labels as supervision.

Lu et al. (2022) proposed an interpretable, weakly supervised framework for survival prediction in computational pathology using full slide images and patient-level prognostic information (demonstrated on renal cell cancer patients). There framework further demonstrates the viability of weakly supervised deep survival models in a federated architecture, paving the path for the development of prognostic models trained on multi-institutional cohorts with diverse demographics.

The work proposed by Kanavati et al. (2020) suggested a weakly supervised deep learning strategy to detect carcinoma in WSIs based on training an EfficientNet-B3, a CNN-based architecture, employing transfer learning and weakly

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstinet), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

supervised learning (WSIs). Their design predicts carcinoma in WSIs of lung cancer using a dataset annotated at a level halfway between two annotation extremes (cell-level annotations and slide-level diagnosis). A similar strategy has been introduced (Kanavati & Tsuneki, 2021) for the categorization of breast invasive ductal carcinoma (IDC) in entire slide images by training deep learning models utilizing transfer learning and weakly supervised learning. The work introduced by Zhou et al. (2021) devised a framework that included characteristics from different magnifications of WSIs to accomplish classification and localization of colorectal carcinoma by using only global labels. Colorectal cancer WSIs from the Cancer Genome Atlas (TCGA) were used to train and test the algorithm.

Most recently, the work introduced by Xiang et al. (2022) suggested that WSI analysis can be carried out efficiently by combining data at both the local and regional levels. They represented the local information by auto-encoding the visual signals in each patch of WSI into a latent embedding vector, and they used a down-sampled WSI with multiple scales to reflect regional information. The WSI label is then predicted using a Dual-Stream Network (DSNet), which accepts the modified local patch embeddings and multi-scale thumbnail images as inputs. These inputs aid in training the framework by using only image-level label. To solve the label constraint problem in glioma subtypes, where each instance is simply labeled with the glioma subtype without detailed annotations of lesion locations. Hsu et al. (2022) devised a hybrid completely CNN-based glioma subtype classification approach that incorporates combination of WSI and multiparametric magnetic resonance imaging (mpMRIs). It consists of two methods: one based on WSI and the other on mpMRIs. They used a 2D CNN on WSIs to classify the glioma subtype for the WSI-based technique. To fix the label limitation, they used a weakly supervised approach to generate representative patches for glioma subtype classification. They devised a 3D CNN-based approach for the mpMRI-based method by examining the mpMRIs. Finally, they combined the WSI and mpMRI data using a confidence measure to improve predictability. HCRF-AM (Hierarchical Conditional Random Field based Attention Mechanism) is developed by Li et al. (2022) for gastric histopathology image classification. The HCRF-AM model is made up of two modules: an Attention Mechanism (AM) and an Image Classification (IC). An HCRF model is created in the AM module to capture attention areas. The IC module trains a CNN with the attention areas specified, and then uses ensemble Learning technique based on probability distribution to generate image-level findings from the CNN's patch-level output. Chen et al. (2022) introduced the IL-MCAM framework for weakly supervised colorectal histopathology image classification, which is based on attention processes and interactive learning. The framework has two phases: automatic learning (AL) and interactivity learning (IL). The AL stage has three separate attention mechanism channels and CNNs to capture multiple channel features for classification. For IL stage, the framework uses an interactive technique to constantly incorporate misclassified images to the training set, improving the model's classification performance.

Weakly supervised techniques offer great solutions to tackle large resolution images, yet a general drawback of these methods has been the greater complexity cost in multi-stage approaches and the computational requirement that might associated with it.

4.2.2 | Self-supervised learning methods

Self-supervised learning (SSL) is another promising mechanism to deal with the scarcity of labeled histopathology datasets. Self-supervised learning aims at learning salient features or digital biomarkers using the original/raw input as the learning signal.

The work conducted by Li, Li, and Eliceiri (2021a) suggested a technique for WSI categorization and tumor identification based on multiple instance learning (MIL) that does not require localized annotations. There are three essential components to their technique. First, they developed a new MIL aggregator that simulates the relationships between instances in a dual-stream design using trainable distance measurement. Second, because WSIs can result in huge or imbalanced bags, which make it difficult to train MIL models, they proposed using self-supervised diverse learning to generate suitable representations for MIL and avoid the problem of high memory consumption for large bags. Third, for multi-scale WSI characteristics, they used a pyramidal fusion technique to increase classification and localization accuracy. The work presented by Wang et al. (2021) suggested a hybrid model (TransPath) that is pre-trained on enormously unlabeled histopathology images using Self Supervised Learning (SSL) to find the underlying image characteristic and record domain-specific feature embedding. The TransPath acts as a collaborative local–global feature extractor that is made up of a CNN and a modified transformer framework. They also suggested a token-aggregating and excitation (TAE) module which is stacked behind the transformer encoder's self-attention for obtaining additional global information.

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstinet), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

In the work introduced by Li, Lin, and Xu (2021d), it thoroughly investigates the inherent properties of WSIs and proposed Spatial Guided Self-supervised Learning on Pathological Images (SSLP). They believed that patch-wise spatial closeness is a key feature of WSIs that, when used appropriately, may offer enough supervision for free. They looked at three types of semantic invariance: (1) self-invariance: the same patch in various augmented perspectives, (2) intra-invariance: patches inside spatial neighbors, and (3) inter-invariance: their feature space equivalent neighbors. Another paper (Koohbanani et al., 2021) suggested Self-Path: a self-supervised CNN approach for learning generalized and domain invariant representations in pathology images utilizing unlabeled data. Self-Path is a multi-task learning architecture in which the primary job is tissue categorization and the pretext tasks are a range of self-supervised tasks with labels embedded in the input images. They developed new self-supervision tasks that take advantage of pathology images' contextual, multi-resolution, and semantic characteristics for semi-supervised learning and domain adaptation. CS-CO has been proposed by Yang et al. (2021) to provide a self-supervised visual representation learning approach for histopathology images that combines the benefits of generative and discriminative models. Their proposed technique is made up of two self-supervised phases: cross-stain prediction (CS) and contrastive learning (CO), which are both built on domain-specific information.

The work conducted by Aryal and Soltani (2022) proposed a self-supervised learning strategy to enhance training with unlabeled data and using a graph convolutional network (GCN) to include context from tumor and adjacent tissues. They visualized the entire slide as a graph, with nodes representing patches from WSIs. The patches in the graph are described as feature vectors derived via self-supervised learning pre-training of WSI patches. The graph is trained with GCN, which takes into account the context of each tissue while grading and classifying cancer. The approach described by Srinidhi et al. (2022) presents task-agnostic unlabeled data based on a self-supervised pretext task that learns a robust supervisory signal for unsupervised representation learning by using the underlying multi-resolution contextual information in histology whole-slide images. Another work introduced by Ciga et al. (2022) addressed the problem of learning domain-specific characteristics without supervision to enhance several task performances that are relevant to the digital histopathology. They collected and pretrained on 57 histopathology datasets without labels using a contrastive self-supervised learning approach. The researchers discovered that mixing numerous multi-organ datasets with various types of staining and resolution qualities improved the quality of the learnt features.

The work presented by Schirris et al. (2022) developed a deep learning-based weak label learning approach for WSIs of H&E stained tumor tissue that does not demand pixel- or patch-level annotations using Self-supervised pre-training and heterogeneity-aware deep Multiple Instance LEarning (DeepSMILE). DeepSMILE is used to predict HRD and MSI (homologous recombination deficiency and microsatellite instability). On histopathological patches of cancer tissue, they use contrastive self-supervised learning to pre-train a feature extractor. In addition, when modeling tumor heterogeneity, they employed a variability-aware deep multiple instance learning to develop the patch feature aggregation function.

Feature extraction is a substantial portion of deep learning modeling, since it is utilized to compress large amounts of information into compact vectors. Recent work in Contrastive predictive encoding has shown significant potential for developing models that do not largely reliant on labeled data to produce an effective feature extractor. Saillard et al. (2021) demonstrated that feature extractors pretrained on TCGA with SSL (MoCo V2) provide state-of-the-art performance for MSI prediction in both colorectal and gastric malignancies. They also discovered that utilizing a feature extractor pretrained on numerous organs using SSL enables both cutting-edge cross-validation and strong generalization from one organ to another.

Abbasi-Sureshjani et al. (2021) also demonstrated that specialized backbones pre-trained on H&E pictures may discover more significant patterns than ImageNet backbones. They utilized embeddings derived from two backbones: one pre-trained on H&E tiles and the other on the ImageNet dataset. The BYOL SSL framework was utilized to train all backbones, and their output embeddings were used as input for training the attention layers in an attention-based MIL pooling. The bag representation of the findings is then processed by a bag-level classifier to deliver the final score for each slide.

In contrast to fully supervised approaches, two-stage weakly supervised methods utilize the MIL method. The signal for the WSI-level label is expected to be present in some of the unlabeled tiles (or instances) in the labeled WSI in a MIL method. To represent the WSI latent characteristics, these methods propose a linear or non-linear combination of the latent features or predicted scores of a subset of the tiles in a WSI. This WSI latent feature is then utilized to classify the WSI, allowing a WSI-level loss to be calculated and transmitted back through the classification network.

A self-supervised approach is proposed by Sahasrabudhe et al. (2020) for segmentation of nuclei in WSIs. Based on the ability to recognize the magnification at which patch is extracted using the size and texture of nuclei, Sahasrabudhe

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstinet), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

et al. (2020) managed to produce a precursory self-supervision signal to determine nuclei location. This occurs depending on the identification of the magnification level for tiles.

Self-supervision prompts a large amount of data, which is not trivial in the medical domain. These methods offer huge potential for a future of modeling in which downstream tasks are highly specialized and otherwise performed using models pre-trained on massive amounts of data from numerous sources. This means that extremely specialized tasks will require far fewer expert annotations. It opens the way for a future in machine learning in which models are not highly specialized algorithms but rather more generalizable techniques used to manage various multimodal datasets and their analysis. Furthermore, the theory of contrastive predictive modeling has been a very insightful breakthrough in this field in which we try to make models that summarize information into very compact vectors that can compress the distinguishing and shared information between different data points so that models can look at a very global scale at the underlying inherit patterns in the data while not ignoring the local intrinsic differences between different categorical d boundaries. Thus, the main direction in self-supervised learning has become what is called Contrastive Learning (van den Oord et al., 2018).

4.2.3 | Stain normalization

The high color variability of histopathology images is another challenging problem that can affect the generalizability of AI models. Such color variability usually results from variations in the thickness of tissue sections, the difference in staining protocols, and the disparity in scanning characteristics, just to mention a few. Since the quality of histopathology images has a direct influence on the performance of AI models, the normalization of the color distribution of images is important for the reliability and accuracy of AI models. This strongly motivates the need for a stain normalization component in the histopathology image analysis pipeline (see Figure 14). According to (Tosta et al., 2019), statistical stain normalization techniques can be classified into three main classes of methods, they are histogram matching, color transfer, and spectral matching. The main idea of these methods is to adjust the color space of the histopathology images in a way to minimize the influence of color variation. Histogram matching methods minimize the color variations using information extracted from the image histogram with the assumption that images have the same histopathological regions. However, unlike histogram matching methods, color transfer and spectral matching perform the normalization using statistical correspondences between histopathological regions (extracted using segmentation methods) and the representation of each stain in the color channels (i.e., the stain concentrations in each image pixel), respectively. For more details about statistical stain normalization techniques, we refer the reader to this recent survey article (Tosta et al., 2019). The article has reviewed the stat-of-the-art computational normalization methods for

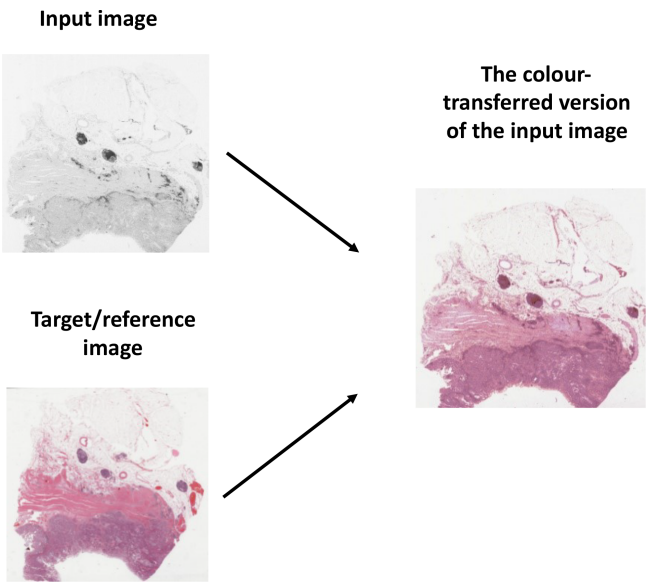

FIGURE 14 Stain normalization method as a process of transforming an input image into a new colored version using a mapping function that resembles the appearance of the input image to a target/reference image.

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstinet), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

histopathology images focusing on H&E stained images with some possible research directions. The authors' reflection on stain normalization methods is that the choice of a reference/target image plays a crucial role in the histology image analysis pipeline. Implementing accurate stain normalization step can help the deep learning model to encode more accurate diagnostic/prognostic features that are less biased by the presence of noise and artifacts in the images.

On other hand, as a deep learning solution, generative models (such as generative adversarial [GAN]-based models) (Niazi et al., 2019) have been recently presented to cope with the stain normalization problem. The idea of these models is to preserve the morphological and architectural structure of the images while manipulating color information. This is by using the color system matrix as input to the generator of GANs to encourage the model to learn the color space of histopathology images and consequently transfer accurately the color from the reference image(s). Recently, a model called StainGAN (Shaban et al., 2019) has been proposed based on CycleGAN to perform stain normalization using the concepts of cycle consistency of adversarial networks. StainGAN showed good performance due to its capability to map images to a specified color model, and at the same time preserve the same histopathology tissue structure. Likewise, a Self-Attentive Adversarial Stain Normalization (SAASN) method (Shrivastava et al., 2021) has been designed in a way to incorporate the self-attention mechanism in CycleGAN to achieve better finer details of the generated images. Another GAN-based model called conditional Generative Adversarial Networks (cGAN) (Senaras et al., 2018) has been adopted to force the generator to learn the underlying distribution of the histopathology images from the training data. In (Cong et al., 2021), a semi-supervised GAN-based stain normalization method has been proposed to learn how to transfer or colorize input images using both labeled (target domain images) and unlabeled (source domain images).

4.3 | Uncertainty quantification in digital pathology

As an initial stage to introduce explainability to classification and segmentation methods of histopathology images, it is crucial to measure the uncertainty of the predictions. During uncertainty quantification, random transformations to the input images or the network architecture generates predictive distribution. This idea leads to figuring out different regions of uncertainty that may be clinically instructive. Moreover, the measure of uncertainty is utilized to help pathologists to rank images that should be prioritized for pathology annotations. Generally, uncertainty quantification can be applied using two different techniques. First, measuring the uncertainty of an input image prediction (or a probability distribution extracted from an image classification model). This method helps in understanding the level of randomness in the probability distribution of an input sample. To adopt this approach, Shannon Entropy (Shannon, 1948) is one of the popular techniques to measure uncertainty in the input samples. Image prediction (probability distribution) generated by deep learning model for image classification is used by Shannon entropy as an input to produce an uncertainty score. This uncertainty score indicates the level of randomness occurring in the probability distribution. For example, having a low uncertainty score value implies that a deep learning model is confident about its decision. Shannon Entropy can be represented as:

$$E(X) = E(p_1, ..., p_c) = -\sum_{i=1}^{c} p_i \log_2 p_i \ , \tag{1}$$

where $E(X)$ is Shannon entropy for input sample X and p_1, ..., p_c presents probability distribution for sample X on c class labels.

Instead of having single scalar of probability distribution for a given input image to measure the uncertainty score, it is possible to have multi-scalar of probability distributions for a particular image. This approach is another method to measure the level of uncertainty for an input image. Bayesian approximation using Monte-Carlo (MC) dropout (Gal & Ghahramani, 2016) is one of the popular strategies to measure the level of ambiguity in the input sample. This method works by adding dropout layers to the deep learning network for image classification. Then, during testing stage of an input image, the dropout layers are activated. This approach helps in having different versions from the deep learning model by randomly dropping different neurons from the network, and hence we can have different probability distributions for a specific input image. A proper number of test passes through the deep neural network generates a number of probability distributions. The mean calculated from the list of probability distributions represents the final image prediction, while standard deviation indicates a measure of uncertainty for the input image. We can represent the formula for calculating mean and standard deviation as:

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstine), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$\mu = \frac{1}{k}\sum_{i=1}^{k} \alpha(\Phi_i(X); W), \tag{2}$$

$$\sigma = \frac{1}{k}\sum_{i=1}^{k} (\alpha(\Phi_i(X); W) - \mu)^2, \tag{3}$$

where μ and σ define the mean image prediction and the uncertainty, respectively. k is the number of MC dropout test passes. The function α defines the deep neural network for image classification with input X and network weights W. Φ_i defines a test pass i using MC dropout to the input image X.

Recent work has introduced the use of uncertainty measure in their models (Fraz et al., 2018, 2020; Graham et al., 2018; Liang et al., 2020; Raczkowski et al., 2019). The work conducted by (Graham et al., 2018) proposed a measure of uncertainty in their CNN-based model by applying random transformations during test time for an enhanced segmentation result that simultaneously creates an instability map, highlighting zones of equivocalness. Fraz et al. (2018) proposed uncertainty driven pooling network for micro-vessel segmentation in routine histology images. Their model is based on an aligned Xception model for feature extraction across many scales, followed by atrous spatial pyramid pooling. Their design overcomes the difficulty of segmenting blood veins with different morphological appearances. Random transformations are used at test time to integrate uncertainty for a better segmentation result and simultaneous uncertainty map development, highlighting confusing locations. Another work by Fraz et al. (2020) proposed a feature attention block-based pyramid pooling deep neural network which applies random transformations to the test images to capture model uncertainty. The work in (Liang et al., 2020) proposed a calibration approach that preserves the overall classification accuracy, and improves model calibration. Their proposed method depends on Expected Calibration Error (ECE), which is a common metric for quantifying mis-calibration. For classification of colorectal cancer images, (Raczkowski et al., 2019) proposed an accurate, reliable, and active Bayesian network termed (ARA-CNN). Their model is built using residual networks and variational dropout to evaluate the uncertainty of the input data.

4.4 | Explainability in histopathology

Recent developments in AI have generated lots of promising algorithms capable of achieving high accuracy in detecting and classifying multiple pathologies. These endeavors have so far proved the effectiveness of machine learning algorithms in a quantitative manner. The growth in digitization of pathology workload is accelerating the development of AI models and approaches that allow for systems that can act as a decision support mechanism for pathologists. By integrating these solutions in the workflow, potential additional improvements to results in turn-around time and better detection quality. For such solutions to be implemented in a clinical workflow, the quality of theses methods' predictions need to be assessed and improved. One way of addressing this problem is to introduce techniques that can give an insight into the model's predictions.

Explainable artificial intelligence (XAI) aims to increase the trustworthiness of AI solutions by giving an insight into the predictions of machine learning models. This allows clinical staff to gain more qualitative evidence of what led the algorithm to make these decisions.

According to a recent survey in the field (Pocevičiūtė et al., 2020), there are three main roles that XAI can help enhance their workflow. These roles include AI developers who design models for application; clinical end-users who act as the medical staff using the AI solution to reach a diagnosis decision; and quality assurance personal responsibility for ensuring and assessing how well a solution is integrated into the workflow of a clinical setting.

Explanations can be presented in various ways. Some representations offer more relevant insight depending on who is viewing it. For example, explanations that show the inner workings of a model can be of assistance to developers as it allows them to detect certain biases that a model might be suffering from. On the other hand, clinical staff might find these explanations to be confusing as it requires some machine learning knowledge to interpret. Yet they might benefit from an explanation that shows which parts of the image are, for example, "cancerous" and thus led the model to classify this image as positive. These situations show the great variety of scenarios where pathology diagnostics would greatly benefit from interpretability and transparency of AI methods used in digital pathology. Figure 15 shows the different ways explanation techniques have been used and their respective target of explanation. From the figure, we can see that there is a trend in using projections on input to explain a network's prediction.

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstinet), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

This section goes over several existing approaches that aim at helping build clear and interpretable AI solutions. These approaches are designed for visual identification tasks, with some easily extended to segmentation tasks. A list of the available resources for XAI is provided in Table 3.

Visualizations on input images are very helpful in giving intuitive insight into an algorithm's prediction by highlighting areas in the image that influenced the decision making process. Methods in this group are categorized under three types: heat maps, patch selection, and receptive fields. Heat maps offer a graphical representation of how much each pixel contributes to the prediction. This is communicated by overlaying a color gradient over the original input image. For example, Grad-CAM (Selvaraju et al., 2019) is a technique that utilizes the gradients flowing through the last layer of a CNN to obtain an importance map for the target class. The resulting maps indicate which pixels contributed mostly to the prediction of the model. The second type of these visualizations by Ribeiro et al. (2016) focuses on the patches or regions in the input image that contributes most to the prediction by iteratively blocking parts of the image and recording the changes in activation in the network. The results reveal some regions that, depending on the dataset, resemble some objects in the image which can be intuitively assessed to determine if the network is basing its prediction on the correct discriminative features.

A similar visual interpretation can be also achieved through attention-based methods which allow for a form of explainable visualization that have become a useful concept in deep learning. The basic idea of attention is inspired by the way humans pay attention to different parts of an image or other data sources and analyze them. Sun, Darbehani, et al. (2020b) proposed an interpretable version of the U-Net (Ronneberger et al., 2015) called SAUNet. They added a parallel secondary shape stream to capture important shape-based information along with the regular texture features of the images. The architecture used an attention module in the decoder part of the U-Net. The spatial and shape attention maps were generated using SmoothGrad (Smilkov et al., 2017) to visualize the high activation region of the images.

A presentation of this degree allows the interpretability of certain biases that a trained model might be suffering from. In histopathology, this can be applied to present the contributing factors to the pathologist, allowing them to understand the "logic" behind the system's diagnosis.

A second set of visualizations that aim at understanding neural networks better are synthetic visualizations. They enable the assessment of how sensitive a model is to domain variations, which can be used to detect new data that a model will have problems generalizing to. These approaches are known as activation maximization and include Guided Back-Propagation (Zeiler & Fergus, 2014), DeConvNet (Springenberg et al., 2015), Layer-wise Relevance Propagation (Bach et al., 2015), integrated Gradients (Sundararajan et al., 2017), and SmoothGrad (Smilkov et al., 2017).

By optimizing for images that maximally activate the neuron in a network, Stacke et al. (2019) demonstrated that they are able to assess the generalization of a convolutional neural network, trained for tumor classification on H&E stained images, by analyzing the representations generated by the network and recording which features the network responds to.

It has been demonstrated that color can be a crucial feature in the prediction of a neural network (Dosovitskiy & Brox, 2016). Their experiments utilized the feature vectors generated by the deep neural network layers to reconstruct

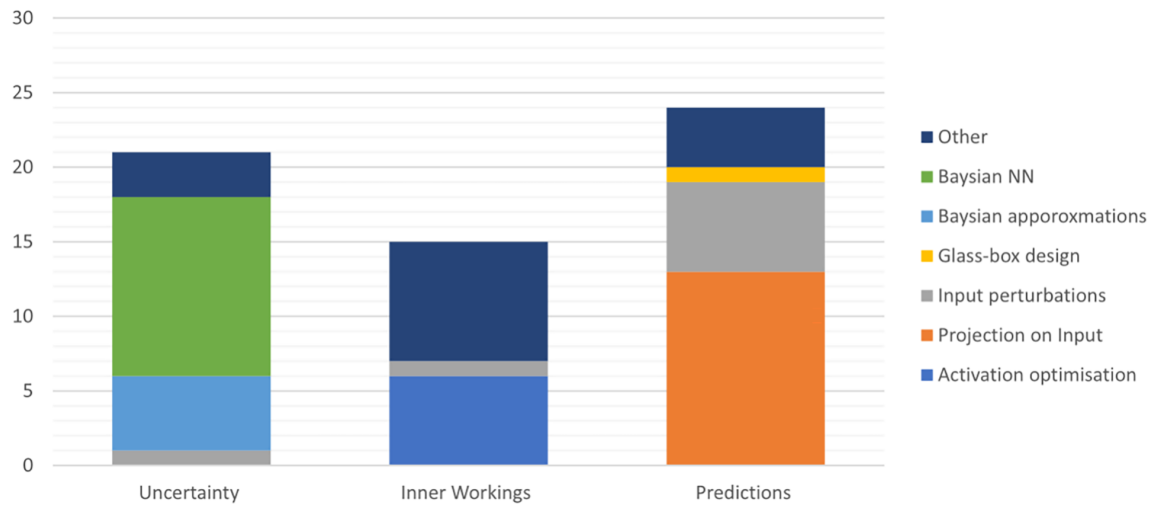

FIGURE 15 Summary of papers grouped based on target of explanation technique.

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstinet), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 List of resources for XAI repositories and software tools

Name	Type	References	Details
ExplainX	framework	explainX.ai (2019)	A light-weight and scalable explainable AI framework.
Explainable AI	framework	Google (2020)	A set of tools and packages developed by Google Cloud.
EthicalML/XAI	Library	EthicalML (2019)	Machine Learning library with AI explainability in its core.
tf-explain	Library	sicara (2019)	A tensorflow library with interpretability methods.
InterpretML	Library	Nori et al. (2019)	Python library (of explainable machine learning models) developed by Microsoft.
sklearn_explain	Library	antoinecarme (2019)	An experimental tool to provide a score explanation for scikit-learn models.
iNNvestigate	Library	Alber et al. (2019)	A software library with various implementations, including Saliency, Deconvnet, GuidedBackprop, IntergratedGradients, LRP, and PatternNet.
Skater	Library	Oracle (2018)	Python Library for Model Interpretation/Explanations.
AI Fairness 360	Library	Bellamy et al. (2018)	Python library developed by IBM to help in detecting/removing bias in machine learning models.
Heatmapping repository	Repository	Fraunhofer HHI (2019)	A webpage aims to regroup publications and software to help in better understanding nonlinear predictions of machine learning models.
XAI Resouces	Repository	pbiecek (2019)	A Github repository with various XAI articles and tools.
Interpretable Machine Learning repository	Repository	lopusz (2019)	Opinionated list of resources facilitating model interpretability.

the input image to get a better grasp of how neural network works. By comparing the reconstructions to the input image, an assessment can show how competent the feature extractors of a neural network are. While on their own, these works do not present much in terms of real-life implementation, they serve as a great tool for AI developers and QA staff for the assessment of deep models by offering an insight on how the models reach their decisions.

The main drawback of the aforementioned methods is that they are not class specific, that is, they do not provide a specific "reason" for why the model predicted a class. They do, however, give an insight into the patterns that are important in general.

Another category of explainable techniques for neural networks is clustering explanations. These methods present a more intuitive output for human review. The intuition here relies on the idea that visualizations of clusters of input images can reveal the potential rationale behind the neural network decisions. Consequently, these visualizations act as an explanation for how a CNN, for example, might classify images to a certain class based on certain visible features that can be quickly identified when viewed in a cluster of images. The effectiveness of the clustering approach that relied on t-SNE embedding of VGG19 network has been demonstrated in (Faust et al., 2018) to visualize the high-dimensional relationships between 13 classes of lesional tissue. The visualizations allowed them to detect "outliers" or out of distribution images which helps in alerting the user of images the model is uncertain about.

These methods have a variety of potential applications. They could help in creating machine learning solutions that are able to classify and provide an interpretable visualization to the end user. It also can be utilized by developers to detect biases in the data. The influence that some concepts might have on neural networks has been explored in (Bau et al., 2017).

Other methods use examples generated from the original dataset. In (Pocevičiūtė et al., 2020), a GAN network has been developed to transform image patches that contain tumor cells to healthy ones. By recording the neural network confidence probability, they were able to show for each tumor patch another patch that resembles the original but transformed into a healthy sample. The idea is that they can boost pathologist's confidence in model predictions by showing an example of how would the same slide look if it was healthy.

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstine), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

4.5 | Actionability in histopathology

This section goes over the next steps after explainable model predictions. To fully realize interpretable models, their predictions need to present an "actionable" intent. For example, based on a model's prediction, a proposed treatment plan based on the detected condition can act as a step forward in enhancing the user experience. Figure 16 shows an example for actionability workflow pipeline for automated diagnosis (image classification) deep learning model.

While up until now, almost all of the machine learning applications are very capable of achieving high diagnosis accuracy, some of them are still quite complex to explain their diagnostic decisions. However, the ability to act, that is, the ability to turn predictions into action, continues to be absent in many applications. Users would not only need reliable, interpretable predictions in real-world applications, but also instructions that are actionable to move the current output to a better, more desirable target. For instance, a model may provide a diagnosis in a medical workflow that is interpretable to technicians with a suggested treatment plan, advice on more scans, or other immediate actions. This pipeline offers every department in a medical hospital with invaluable insight. Actionability presents itself in any implementation of AI solutions as a more general problem. To provide value to users, a model needs to provide actionable insight, based on trustworthy, accountable, reliable data. They can provide real world value by translating these insights into practical decisions or actions. Models attempting such approaches are currently often complicated and lack optimal solutions. Current efforts to extract such actionable recommendations are restricted and limited to simple action models, which are restricted to modifying only one attribute for each action.

A novel approach has been proposed in (Lyu et al., 2016), which achieves actionability by combining learning with planning. Specifically, they proposed a system for the extraction of information from random forests, one of the most commonly used and best off-the-shelf classifiers. Their approach addresses actionability by formulating a sub-optimal action planning (SOAP) problem. To solve this problem the model tries to find a plan to alter some features of the input such that the random forest can produce a desirable output. Another work has been proposed in (Senousy, Abdelsamea, Mohamed, & Gaber, 2021b) which introduces actionability to deep learning models. They proposed an entropy-based elastic ensemble model termed (3E-Net) for grading breast invasive carcinoma samples where it applies an ensemble of image-wise models supported by a patch-wise CNN (DenseNet-161) for feature extraction. The model uses Shannon Entropy to measure the level of randomness in the obtained probability distribution associated to an input image. Moreover, 3E-Net introduces actionability to by taking decision of excluding poor samples based on their uncertainty scores, see Figure 17. Likewise, another work (Senousy, Abdelsamea, Gaber, et al., 2021a) proposed an actionable model for classification of breast cancer images named Multi Context and Uncertainty awareness (MCUa). Their model introduces an ensemble of multi-scale and multi-architecture models for patch-wise feature extraction where it applies diversity in feature learning. Then, the model introduces different levels of context-awareness for learning spatial dependencies between image patches. Then the model applies Bayesian approximation using MC dropout for measuring uncertainty and it excludes images based on the trust rates generated from the models in the ensemble architecture. Here, the authors come to the conclusion that uncertainty aware models (such as their previously developed 3E-Net

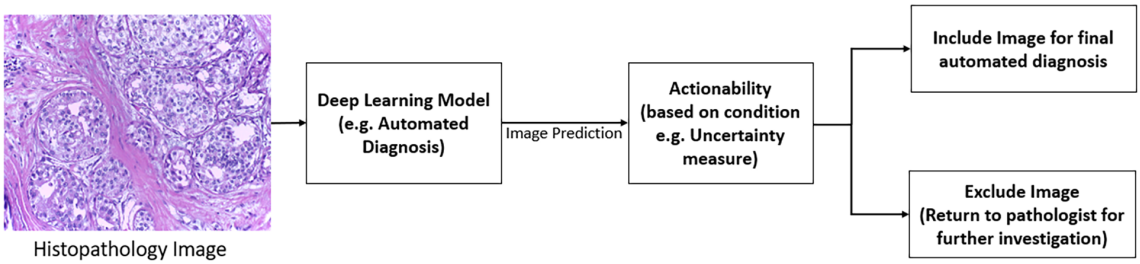

FIGURE 16 Actionability workflow pipeline. To introduce actionability in histopathology, it is crucial to process the predictions extracted from deep learning model (e.g., automated diagnosis model). For instance, the above diagram presents the workflow for a deep learning model for automated diagnosis which takes a histopathology image as input and generates image prediction (e.g., class labels based on probability distribution). The image prediction can be utilized by an actionability component which acts as an active learner that takes decision/action based on certain criteria/condition. The action can be an uncertainty measure which reflects the level of confidence associated with the image prediction. Based on this confidence condition, the image prediction can be used as the final automated diagnosis generated by the deep learning model or to exclude the image from the automated diagnosis process due to image uncertainty. The excluded image is then returned to pathologists for further investigation and re-annotation.

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstine), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

and MCUa) not only help in taking actions by excluding low-quality images or correct annotations but also can help in better understanding the behavior of the deep learning model. This provides an opportunity to build a dynamic ensemble model based on the most confident deep learning models, which can boost the performance of the overall model and at the same improve its trustworthiness.

5 | EVALUATION OF EXPLAINABLE TECHNIQUES

With the introduction of interpretability methods, comes the trivial task of finding a widely accepted quantitative metric to compare them.

Evaluation methods for explainable artificial intelligence (XAI) systems seem to fall into two categories: human-based and computational methods (see Table 4). The measures assess the explanations of XAI interface based on how a human qualitatively scores it or a computational assessment scores it. In order to verify and compare these methods, we need to quantify the "Quality" of the produced explanations.

5.1 | Human evaluated methods

In the world of HCI (Human Computer Interaction), researchers study user's mental models to gauge their understanding of the system they are dealing with. Some methods attempted as assessment of such mental models to evaluate if the XAI explanations led to a correct understanding of how the AI works. Furthermore, this can help in verifying the effectiveness of explanations provided by XAI models.

Kulesza et al. (2013) studied the effect of explanations on user's mental models. By using a combination of short-answer and Likert scale questions, they measured the impact of soundness and completeness on the fidelity user's mental model, their trust in the explanations, and cost/benefit of using these explanations. They defined these two factors as soundness (how truthful each element in an explanation is with respect to the underlying system) and completeness (the extent to which an explanation describes all of the underlying system). Their findings showed that when compared with sound explanations, complete ones were associated with the best mental models and highest perceived benefits with lowest cost. Users also tended to trust explanations that are highly sound and complete.

Ribeiro et al. (2018) experimented with an evaluation of the user's mental model through predictions of model output. In their experiments, users were asked to predict the behavior of the classifier based on examples that included predictions of the models with and without explanations. They demonstrated how explanations can reduce human overestimation of model's accuracy and allow user's to better understand the coverage of some explanations by gaining a high understanding of the behavior for the model. While interview questions can provide researches to examine certain aspects of a user's mental model, they are highly dependent on the structure of the questions which can vary resulting in different impacts on the user. Similarly, most qualitative assessments of mental models such as Think-Aloud, task reflection, etc. provide rich information about the end-user's mental models, but suffer from variability and high time consumption in carrying out the full process of design, implementation, and result analysis.

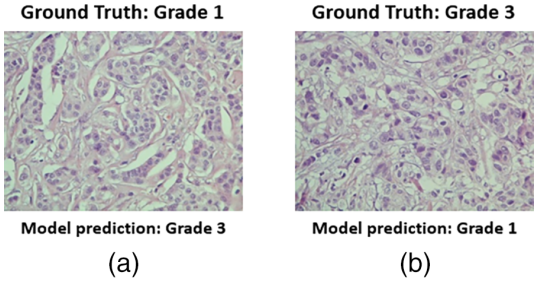

FIGURE 17 Highly uncertain excluded images from final image classification using actionable 3E-Net (Senousy, Abdelsamea, Mohamed, & Gaber, 2021b). (a) Shows an excluded image which was graded as invasive breast carcinoma grade 3 and the actual label was grade 1, while (b) shows an excluded image graded as grade 1 with actual ground truth label grade 3.

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstine), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 4 List of evaluation methods split into two categories: Human evaluated methods that rely on a human to score the AI's performance

	Human evaluated					Computational			
Paper	Interview/ think-aloud	Likert-scale	Response time	Model output prediction	Perceived competence	Simulated experiments	Human generated	Comparative	Sanity check
Lage et al. (2019)		●	●						
Kay et al. (2016)		●		●					
Kulesza et al. (2013)	●	●							
Bussone et al. (2015)	●	●							
Lage et al. (2019)		●	●						
Yin et al. (2019)									
Samek et al. (2015)								●	
Das et al. (2016)							●		
Islam et al. (2019)									
Ribeiro et al. (2018)				●		●			
Krause et al. (2017)				●					
Ross et al. (2017)								●	
Kindermans et al. (2017)									●
Mohseni & Ragan (2018)							●		
Lin et al. (2020)						●			

Note: Computational methods involve a computer generated testing of XAI. Each category is then split into different techniques based on the different methodologies that have been used.

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstinet), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Another quality of interest to evaluate in the user's perception of the model explanations is satisfaction and usefulness. Measuring usefulness can be done through better response time, overall task performance, or measuring cognitive load. User Satisfaction, on the other hand, can be measured in similar qualitative ways of mental models such as interview, questionnaires, etc.

Lage et al. (2019) measured response time as the number of seconds it took the subject to submit a task, while measuring satisfaction through a 5-point scale Likert. After submitting their response to each question, but before being told if their answer was correct, the participant was asked to rate subjectively how difficult it was for the model to answer the question on a scale from 1 to 5 where one was very easy to use, three was neutral, and five was very difficult to use. They found that greater complexity, in terms of cognitive chunks or model size, leads to slower response time and less user satisfaction.

User trust and reliance is another widely departed topic in terms of how to assess such qualities in the AI field. Prior knowledge and beliefs can influence the initial state of trust of the user but overtime response changes depending on the user interaction with the system (Mohseni & Ragan, 2018). Some proposed ways to measure trust and reliance try to enquire about the user's opinion during and after the system interaction (Bussone et al., 2015). Though, these methods can often suffer from user bias and impressions of the user based on the stated accuracy and the perceived accuracy as (Yin et al., 2019) found in their experiment, where they measured trust and reliance on AI system based on the user agreement with the system's recommended action. Their results show that higher stated accuracy of models tends to persuade the users to rely on using them. An effect that gets amplified when the user's own performance improves from reliance on the model.

Finally, a more general way of calculating the effect of the explanations on the workflow of the user is by analyzing the time taken by the user before the introduction of explanations.

5.2 | Computational methods

Reliance on human evaluation can result in the development of model explanations that are more focused on convincing the user rather than offering transparent interface with the underlying machine learning model. Thus, becoming unfaithful to the inner workings of the models. Such implicit biases can be problematic to interpretable machine learning. It limits our ability to provide purely descriptive explanations that are accurate and transparent. Ethically, this becomes more prominent as we move towards more "convincing" explanations, we fail to satisfy our ethical goals of offering transparency of underlying model predictions.

Computational measures aim to mediate this problem by offering solutions that are more reliant. By assessing the consistency of explanations after same input transformations, comparing with inherently interpretable models (gold standards), and assessing similarity with human generated explanations, scoring of explanations can become more "metric" based rather than human-based.

By comparing to inherently interpretable models, some methods can assess the fidelity of explanations to the underlying model. Ribeiro et al. (2016) showed that by creating a gold standard, retrieved from decision trees, and compared their proposed ad-hoc method to it. While this can act as an indicator of interpretability of the method, it is limited by the ability to generate a gold standard that fits the domain space of the field.

Hase and Bansal (2020) proposed a multifaceted metrics to compare explainers based on their correctness, consistency, and confidence. By using masked areas from the input of a high probability data point, they evaluated correctness as the increase in accuracy on the new masked input. Consistency is evaluated based on the ability of the explainer to consistently identify the same relevant components in the image even after input transformation.

Clinically, a physician needs to trust and depend on the system at hand without falling into: self-reliance where clinician rejects the recommendations of the system entirely or over-reliance where they consider suggestions even though they are incorrect. The many variables that this evaluation challenge poses may make it so that a compromise needs to be made, however, with properly researched requirements, a solution that satisfies them would be feasible.

6 | SUMMARY

In this article, we discussed the recent progress in state-of-the-art AI systems applicable on histopathology with an emphasis on deep learning models. Deep learning networks have demonstrated high performance in several

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstine), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

applications in the field of histopathology image analysis, ranging from low-level tasks (such as stain normalization) to high-level tasks (such as explainability and actionability of AI tools). However, these models need a large dataset with detailed annotation to be introduced into the networks. Network architecture and tuning parameters are not trivial tasks. More specifically, the parameters in deep neural networks are difficult to interpret in clinical contexts. With these limitations, the authors believe that an self-supervised and/or weakly supervised learning-based approachs, based solely on limited or partially annotated datasets, would support the histopathology image analysis field. It is also important to consider the relationship between the machine-learned features and the clinical diagnostic criteria.

In order to integrate these algorithms into healthcare systems, there is a problem of reliability and interpretability that need to be addressed. For the histopathology field there are some open problems that still need to be solved. The first problem is introduced by the nature of the explanations and their utility to the user experiences. To gain a deeper understanding of the predictive algorithm, it is important to derive a causal explanation as to why the algorithm made the prediction. Though the XAI techniques may help improve the understandability of our AI solutions, at the current state we do not have a scientifically backed evaluation framework that would tell us how well they perform and what drawbacks they might have. This problem exists, because we still lack a solid definition of what makes an explanation "good", thus making the majority of outputs by any XAI technique uncertain. Nonetheless, the emergence of many explainability techniques makes it clear that many approaches can be considered when developing XAI solutions.

An additional consideration when developing new algorithms is making sure they are actionable in real world situations. In several applications, the ability to act, that is, the ability to transform predictions into action, continues to be absent. Users will need not only accurate, interpretable predictions in real-world applications, but also instructions to move the current output to a better, more desirable target.

AUTHOR CONTRIBUTIONS

Mohammed Abdelsamea: Conceptualization (lead); supervision (lead); writing – original draft (equal). **Usama Zidan:** Writing – original draft (equal). **Zakaria Senousy:** Writing – original draft (equal); writing – review and editing (equal). **Mohamed Medhat Gaber:** Conceptualization (supporting); supervision (equal); writing – review and editing (equal). **Emad Rakha:** Writing – review and editing (supporting). **Mohammad Ilyas:** Writing – review and editing (supporting).

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analysed in this study.

ORCID

Mohammed M. Abdelsamea https://orcid.org/0000-0002-2728-1127
Mohamed Medhat Gaber https://orcid.org/0000-0003-0339-4474

RELATED WIREs ARTICLE

Causability and explainability of artificial intelligence in medicine

REFERENCES

Abbas, A., Abdelsamea, M. M., & Gaber, M. M. (2020). Detrac: Transfer learning of class decomposed medical images in convolutional neural networks. *IEEE Access*, *8*, 74901–74913.

Abbasi-Sureshjani, S., Yüce, A., Schönenberger, S., Skujevskis, M., Schalles, U., Gaire, F. and Korski, K. (2021), Molecular subtype prediction for breast cancer using h&e specialized backbone, pp. 1–9.

Abdelsamea, M. M., Grineviciute, R. B., Besusparis, J., Cham, S., Pitiot, A., Laurinavicius, A., & Ilyas, M. (2019a). Tumour parcellation and quantification (tupaq): A tool for refining biomarker analysis through rapid and automated segmentation of tumour epithelium. *Histopathology*, *74*(7), 1045–1054.

Abdelsamea, M. M., Pitiot, A., Grineviciute, R. B., Besusparis, J., Laurinavicius, A., & Ilyas, M. (2019b). A cascade-learning approach for automated segmentation of tumour epithelium in colorectal cancer. *Expert Systems with Applications*, *118*, 539–552.

Ahmad, N., Asghar, S., & Gillani, S. A. (2021). Transfer learning-assisted multi-resolution breast cancer histopathological images classification. *The Visual Computer*. https://doi.org/10.1007/s00371-021-02153-y

Alber, M., Lapuschkin, S., Seegerer, P., Hägele, M., Schütt, K. T., Montavon, G., Samek, W., Müller, K.-R., Dähne, S., Kindermans, P.-J., & Müller, K.-R. (2019). Innvestigate neural networks! *Journal of Machine Learning Research*, *20*, 1–8. http://jmlr.org/papers/v20/18-540.html

Ali, S., & Madabhushi, A. (2012). An integrated region-, boundary-, shape-based active contour for multiple object overlap resolution in histological imagery. *IEEE Transactions on Medical Imaging*, *31*(7), 1448–1460.

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstinet), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Al-Janabi, S., Huisman, A., & Van Diest, P. J. (2012). Digital pathology: Current status and future perspectives. *Histopathology*, *61*(1), 1–9.

antoinecarme (2019), Model explanation provides the ability to interpret the effect of the predictors on the composition of an individual score. https://github.com/antoinecarme/sklearn-explain

Aryal, M. and Soltani, N. Y. (2022), Context-aware graph-based self-supervised learning of whole slide images. In ICASSP 2022–2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 3553–3557.

Awan, R., Koohbanani, N., Shaban, M., Lisowska, A., & Rajpoot, N. (2018). Context-aware learning using transferable features for classification of breast cancer histology images.

Awan, R. and Rajpoot, N. (2018), Deep autoencoder features for registration of histology images. In Annual Conference on Medical Image Understanding and Analysis, Springer 371–378.

Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K., & Samek, W. (2015). On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLoS One*, *10*, e0130140.

Bagci, U., & Bai, L. (2010). Automatic best reference slice selection for smooth volume reconstruction of a mouse brain from histological images. *IEEE Transactions on Medical Imaging*, *29*(9), 1688–1696.

Ballerini, L., Bocchi, L. and Johansson, C. B. (2004), Image segmentation by a genetic fuzzy c-means algorithm using color and spatial information, in 'EvoWorkshops', Springer, pp. 260–269.

Banerjee, I., Ling, Y., Chen, M. C., Hasan, S. A., Langlotz, C. P., Moradzadeh, N., Chapman, B., Amrhein, T., Mong, D., Rubin, D. L., Farri, O., & Lungren, M. P. (2019). Comparative effectiveness of convolutional neural network (CNN) and recurrent neural network (rnn) architectures for radiology text report classification. *Artificial Intelligence in Medicine*, *97*, 79–88. http://www.sciencedirect.com/science/article/pii/S0933365717306255

Barisoni, L., Lafata, K. J., Hewitt, S. M., Madabhushi, A., & Balis, U. G. (2020). Digital pathology and computational image analysis in nephropathology. *Nature Reviews Nephrology*, *16*, 1–17.

Bau, D., Zhou, B., Khosla, A., Oliva, A. and Torralba, A. (2017), Network dissection: Quantifying interpretability of deep visual representations. In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3319–3327.

Belhassen, S., & Zaidi, H. (2010). A novel fuzzy c-means algorithm for unsupervised heterogeneous tumor quantification in pet. *Medical Physics*, *37*(3), 1309–1324.

Bellamy, R. K. E., Dey, K., Hind, M., Hoffman, S. C., Houde, S., Kannan, K., Lohia, P., Martino, J., Mehta, S., Mojsilovic, A., Nagar, S., Ramamurthy, K. N., Richards, J., Saha, D., Sattigeri, P., Singh, M., Varshney, K. R. and Zhang, Y. (2018), 'AI fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias'. URL: https://arxiv.org/abs/1810.01943

Borovec, J., Kybic, J., Arganda-Carreras, I., Sorokin, D. V., Bueno, G., Khvostikov, A. V., Bakas, S., Eric, I., Chang, C., Heldmann, S., et al. (2020). Anhir: Automatic non-rigid histological image registration challenge. *IEEE Transactions on Medical Imaging*, *39*(10), 3042–3052.

Borovec, J., Kybic, J., Bušta, M., Ortiz-de Solórzano, C. and Munoz-Barrutia, A. (2013), Registration of multiple stained histological sections, In:2013 IEEE 10th International Symposium on Biomedical Imaging, IEEE, pp. 1034–1037.

Borovec, J., Munoz-Barrutia, A. and Kybic, J. (2018), Benchmarking of image registration methods for differently stained histological slides, In 2018 25th IEEE International Conference on Image Processing (ICIP), IEEE, pp. 3368–3372.

Bueno, G., Fernández-Carrobles, M. M., Deniz, O., & García-Rojo, M. (2016). New trends of emerging technologies in digital pathology. *Pathobiology*, *83*(2–3), 61–69.

Bulten, W., Bándi, P., Hoven, J., Loo, R., Lotz, J., Weiss, N., Laak, J., Ginneken, B., Hulsbergen-Kaa, C., & Litjens, G. (2019). Epithelium segmentation using deep learning in h&e-stained prostate specimens with immunohistochemistry as reference standard. *Scientific Reports*, *9*(1), 1–10.

Bussone, A., Stumpf, S., & O'Sullivan, D. (2015). *The role of explanations on trust and reliance in clinical decision support systems* (pp. 160–169). Institute of Electrical and Electronics Engineers Inc.

Bychkov, D., Joensuu, H., Nordling, S., Tiulpin, A., Kücükel, H., Lundin, M., Sihto, H., Isola, J., Lehtimäki, T., Kellokumpu-Lehtinen, P.-L., Smitten, K., Lundin, J., & Linder, N. (2022). Outcome and biomarker supervised deep learning for survival prediction in two multicenter breast cancer series. *Journal of Pathology Informatics*, *13*(1), 9.

Campanella, G., Hanna, M. G., Geneslaw, L., Miraflor, A., Werneck Krauss Silva, V., Busam, K. J., Brogi, E., Reuter, V. E., Klimstra, D. S., & Fuchs, T. J. (2019). Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature Medicine*, *25*(8), 1301–1309. https://doi.org/10.1038/s41591-019-0508-1

Chang, H., Loss, L. A. and Parvin, B. (2012), Nuclear segmentation in H&E sections via multi-reference graph cut (MRGC). In International symposium biomedical imaging.

Chen, H., Li, C., Li, X., Rahaman, M. M., Hu, W., Li, Y., Liu, W., Sun, C., Sun, H., Huang, X., & Grzegorzek, M. (2022). Il-mcam: An interactive learning and multi-channel attention mechanism-based weakly supervised colorectal histopathology image classification approach. *Computers in Biology and Medicine*, *143*, 105265.

Chen, H., Qi, X., Yu, L., Dou, Q., Qin, J., & Heng, P.-A. (2017). Dcan: Deep contour-aware networks for object instance segmentation from histology images. *Medical Image Analysis*, *36*, 135–146.

Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A. L. and Zhou, Y. (2021a), 'Transunet: Transformers make strong encoders for medical image segmentation', https://arxiv.org/abs/2102.04306

Chen, P., Liang, Y., Shi, X., Yang, L., & Gader, P. (2021b). Automatic whole slide pathology image diagnosis framework via unit stochastic selection and attention fusion. *Neurocomputing*, *453*, 312–325.

Chen, P., Shi, X., Liang, Y., Li, Y., Yang, L., & Gader, P. D. (2020). Interactive thyroid whole slide image diagnostic system using deep representation. *Computer Methods and Programs in Biomedicine*, *195*, 105630.

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstinet), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Chen, T. and Chefdhotel, C. (2014), Deep learning based automatic immune cell detection for immunohistochemistry images, In Machine Learning in Medical Imaging, Springer, pp. 17–24.

Chennamsetty, S. S., Safwan, M., & Alex, V. (2018). Classification of breast cancer histology image using ensemble of pre-trained neural networks. In A. Campilho, F. Karray, & B. ter Haar Romeny (Eds.), *Image analysis and recognition* (pp. 804–811). Springer International Publishing.

Cifci, D., Foersch, S., & Kather, J. N. (2022). Artificial intelligence to identify genetic alterations in conventional histopathology. *The Journal of Pathology*, *257*, 430–444.

Ciga, O., Xu, T., & Martel, A. L. (2022). Self supervised contrastive learning for digital histopathology. *Machine Learning with Applications*, *7*, 100198.

Cong, C., Liu, S., Ieva, A. D., Pagnucco, M., Berkovsky, S. and Song, Y. (2021), Semi-supervised adversarial learning for stain normalisation in histopathology images. In International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, pp. 581–591.

Courtiol, P., Maussion, C., Moarii, M., Pronier, E., Pilcer, S., Sefta, M., Mnceron, P., Toldo, S., Zaslavskiy, M., le Stang, N., et al. (2019). Deep learning-based classification of mesothelioma improves prediction of patient outcome. *Nature Medicine*, *25*(10), 1519–1525.

Cruz-Roa, A., Gilmore, H., Basavanhally, A., Feldman, M., Ganesan, S., Shih, N., Tomaszewski, J., Madabhushi, A., & González, F. (2018). High-throughput adaptive sampling for whole-slide histopathology image analysis (hashi) via convolutional neural networks: Application to invasive breast cancer detection. *PLoS One*, *13*, e0196828.

Cui, Y., Zhang, G., Liu, Z., Xiong, Z. and Hu, J. (2018), A deep learning algorithm for one-step contour aware nuclei segmentation of histopathological images. http://arxiv.org/abs/1803.02786

Das, A., Agrawal, H., Zitnick, C. L., Parikh, D. and Batra, D. (2016), Human attention in visual question answering: Do humans and deep networks look at the same regions?. http://arxiv.org/abs/1606.03556

Datar, M., Padfield, D. and Cline, H. (2008), Color and texture based segmentation of molecular pathology images using hsoms. In 2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, IEEE, pp. 292–295.

Dhivya, P., & Vasuki, S. (2019). An ensemble approach for classification of breast histopathology images. *IETE Journal of Research*, *68*, 1–10. https://doi.org/10.1080/03772063.2019.1644974

Do, H. M., Spear, L. G., Nikpanah, M., Mirmomen, S. M., Machado, L. B., Toscano, A. P., Turkbey, B., Bagheri, M. H., Gulley, J. L., & Folio, L. R. (2020). Augmented radiologist workflow improves report value and saves time: A potential model for implementation of artificial intelligence. *Academic Radiology*, *27*(1), 96–105. Special Issue: Artificial Intelligence. http://www.sciencedirect.com/science/article/pii/S1076633219304477

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J. and Houlsby, N. (2020), An image is worth 16x16 words: Transformers for image recognition at scale, CoRR abs/2010.11929. https://arxiv.org/abs/2010.11929

Dosovitskiy, A. and Brox, T. (2016), Inverting visual representations with convolutional networks. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4829–4837.

Echle, A., Rindtorff, N. T., Brinker, T. J., Luedde, T., Pearson, A. T., & Kather, J. N. (2021). Deep learning in cancer pathology: A new generation of clinical biomarkers. *British Journal of Cancer*, *124*(4), 686–696.

Ehteshami Bejnordi, B., Zuidhof, G., Balkenhol, M., Hermsen, M., Bult, P., Ginneken, B., Karssemeijer, N., Litjens, G., & van der Laak, J. (2017). Context-aware stacked convolutional neural networks for classification of breast carcinomas in whole-slide histopathology images. *Journal of Medical Imaging*, *4*, 1–8.

EthicalML (2019), 'explainable machine learning - xai - explainable ai 0.1 documentation'. https://ethicalml.github.io/xai/index.html.

explainX.ai (2019), explainX.ai|explainable ai for data scientists. https://www.explainx.ai/.

Fatakdawala, H., Xu, J., Basavanhally, A., Bhanot, G., Ganesan, S., Feldman, M., Tomaszewski, J. E., & Madabhushi, A. (2010). Expectation–maximization-driven geodesic active contour with overlap resolution (emagacor): Application to lymphocyte segmentation on breast cancer histopathology. *IEEE Transactions onBiomedical Engineering*, *57*(7), 1676–1689.

Faust, K., Xie, Q., Han, D., Goyle, K., Volynskaya, Z., Djuric, U., & Diamandis, P. (2018). Visualizing histopathologic deep learning classification and anomaly detection using nonlinear feature space dimensionality reduction. *BMC Bioinformatics*, *19*(1), 173.

Feng, Z., Li, A., Gong, H., & Luo, Q. (2016). An automatic method for nucleus boundary segmentation based on a closed cubic spline. *Frontiers in Neuroinformatics*, *10*, 1–10.

Fernandez-Gonzalez, R., Jones, A., Garcia-Rodriguez, E., Chen, P. Y., Idica, A., Lockett, S. J., Barcellos-Hoff, M. H., & Ortiz-De-Solorzano, C. (2002). System for combined three-dimensional morphological and molecular analysis of thick tissue specimens. *Microscopy Research and Technique*, *59*(6), 522–530.

Feuerstein, M., Heibel, H., Gardiazabal, J., Navab, N. and Groher, M. (2011), Reconstruction of 3-d histology images by simultaneous deformable registration. In International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, pp. 582–589.

Folio, L. R., Machado, L. B., & Dwyer, A. J. (2018). Multimedia-enhanced radiology reports: Concept, components, and challenges. *Radiographics*, *38*(2), 462–482. https://doi.org/10.1148/rg.2017170047

Fraunhofer HHI, (2019), 'Heatmapping, TU Berlin'. http://heatmapping.org/.

Fraz, M., Khurram, S. A., Graham, S., Shaban, M., Loya, A., & Rajpoot, N. (2020). Fabnet: Feature attention-based network for simultaneous segmentation of microvessels and nerves in routine histology images of oral cancer. *Neural Computing and Applications*, *32*, 9915–9928.

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstinet), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Fraz, M. M., Shaban, M., Graham, S., Khurram, S. A., & Rajpoot, N. M. (2018). Uncertainty driven pooling network for microvessel segmentation in routine histology images. In D. Stoyanov, Z. Taylor, F. Ciompi, Y. Xu, A. Martel, L. Maier-Hein, N. Rajpoot, J. van der Laak, M. Veta, S. McKenna, D. Snead, E. Trucco, M. K. Garvin, X. J. Chen, & H. Bogunovic (Eds.), *Computational pathology and ophthalmic medical image analysis* (pp. 156–164). Springer International Publishing.

Fu, H., Qiu, G., Shu, J., & Ilyas, M. (2014). A novel polar space random field model for the detection of glandular structures. *IEEE Transactions on Medical Imaging*, *33*(3), 764–776.

Gal, Y. and Ghahramani, Z. (2016), Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In international conference on machine learning, pp. 1050–1059.

Google (2020), Explainable ai|google cloud. https://cloud.google.com/explainable-ai/.

Graham, S., Chen, H., Dou, Q., Heng, P. and Rajpoot, N. M. (2018), Mild-net: Minimal information loss dilated network for gland instance segmentation in colon histology images, CoRRabs/1806.01963. http://arxiv.org/abs/1806.01963

Graham, S. and Rajpoot, N. M. (2018), Sams-net: Stain-aware multi-scale network for instance-based nuclei segmentation in histology images. In 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), pp. 590–594.

Graham, S., Vu, Q. D., Raza, S. E. A., Azam, A., Tsang, Y. W., Kwak, J. T., & Rajpoot, N. (2019). Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical Image Analysis*, *58*, 101563. https://www.sciencedirect.com/science/article/pii/S1361841519301045

Greenwald, N. F., Miller, G., Moen, E., Kong, A., Kagel, A., Dougherty, T., Fullaway, C. C., McIntosh, B. J., Leow, K. X., Schwartz, M. S., Pavelchek, C., Cui, S., Camplisson, I., Bar-Tal, O., Singh, J., Fong, M., Chaudhry, G., Abraham, Z., Moseley, J., ... van Valen, D. (2022). Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Nature Biotechnology*, *40*(4), 555–565.

Gunduz-Demir, C., Kandemir, M., Tosun, A. B., & Sokmensuer, C. (2010). Automatic segmentation of colon glands using object-graphs. *Medical Image Analysis*, *14*(1), 1–12.

Gupta, L., Klinkhammer, B. M., Boor, P., Merhof, D. and Gadermayr, M. (2018), Stain independent segmentation of whole slide images: A case study in renal histology. In 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), IEEE, pp. 1360–1364.

Gurcan, M. N., Boucheron, L. E., Can, A., Madabhushi, A., Rajpoot, N. M., & Yener, B. (2009). Histopathological image analysis: A review. *IEEE Reviews in Biomedical Engineering*, *2*, 147–171.

Hafiane, A., Bunyak, F. and Palaniappan, K. (2008), Fuzzy clustering and active contours for histopathology image segmentation and nuclei detection. In Advanced concepts for intelligent vision systems, Springer, pp. 903–914.

Han, C., Lin, J., Mai, J., Wang, Y., Zhang, Q., Zhao, B., Chen, X., Pan, X., Shi, Z., Xu, Z., et al. (2022). Multi-layer pseudo-supervision for histopathology tissue semantic segmentation using patch-level classification labels. *Medical Image Analysis*, *80*, 102487.

Hase, P., & Bansal, M. (2020). Evaluating explainable AI: Which algorithmic explanations help users predict model behavior? *arXiv Preprint*, arXiv:2005.01831.

Hiary, H., Alomari, R. S., Saadah, M., & Chaudhary, V. (2013). Automated segmentation of stromal tissue in histology images using a voting Bayesian model. *Signal, Image and Video Processing*, *7*(6), 1229–1237.

Hirra, I., Ahmad, M., Hussain, A., Ashraf, M. U., Saeed, I. A., Qadri, S. F., Alghamdi, A. M., & Alfakeeh, A. S. (2021). Breast cancer classification from histopathological images using patch-based deep learning modeling. *IEEE Access*, *9*, 24273–24287.

Hsu, W.-W., Guo, J.-M., Pei, L., Chiang, L.-A., Li, Y.-F., Hsiao, J.-C., Colen, R., & Liu, P. (2022). A weakly supervised deep learning-based method for glioma subtype classification using wsi and mpmris. *Scientific Reports*, *12*(1), 6111. https://doi.org/10.1038/s41598-022-09985-1

Huang, Y. and Chung, A. C. (2018), Improving high resolution histology image classification with deep spatial fusion network, CoRRabs/1807.10552. http://arxiv.org/abs/1807.10552

Ibrahim, A., Gamble, P., Jaroensri, R., Abdelsamea, M. M., Mermel, C. H., Chen, P.-H. C., & Rakha, E. A. (2020). Artificial intelligence in digital breast pathology: Techniques and applications. *The Breast*, *49*, 267–273.

Iizuka, O., Kanavati, F., Kato, K., Rambeau, M., Arihiro, K., & Tsuneki, M. (2020). Deep learning models for histopathological classification of gastric and colonic epithelial tumours. *Scientific Reports*, *10*(1), 1–11.

Islam, S. R., Eberle, W., & Ghafoor, S. K. (2019). Towards quantification of explainability in explainable artificial intelligence methods. In *The thirty-third international flairs conference*, *2020*. pp. 75–81. http://arxiv.org/abs/1911.10104

Kainz, P., Pfeiffer, M., & Urschler, M. (2015). Semantic segmentation of colon glands with deep convolutional neural networks and total variation segmentation. *arXiv preprint*, arXiv:1511.06919.

Kanavati, F., Toyokawa, G., Momosaki, S., Rambeau, M., Kozuma, Y., Shoji, F., Yamazaki, K., Takeo, S., Iizuka, O., & Tsuneki, M. (2020). Weakly-supervised learning for lung carcinoma classification using deep learning. *Scientific Reports*, *10*(1), 9297. https://doi.org/10.1038/s41598-020-66333-x

Kanavati, F., Toyokawa, G., Momosaki, S., Takeoka, H., Okamoto, M., Yamazaki, K., Takeo, S., Iizuka, O., & Tsuneki, M. (2021). A deep learning model for the classification of indeterminate lung carcinoma in biopsy whole slide images. *Scientific Reports*, *11*(1), 8110. https://doi.org/10.1038/s41598-021-87644-7

Kanavati, F., & Tsuneki, M. (2021). Breast invasive ductal carcinoma classification on whole slide images with weakly-supervised and transfer learning. *Cancers*, *13*(21), 1–14. https://www.mdpi.com/2072-6694/13/21/5368

Kang, Q., Lao, Q., & Fevens, T. (2019). Nuclei segmentation in histopathological images using two-stage learning. In D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, P.-T. Yap, & A. Khan (Eds.), *Medical image computing and computer assisted intervention – MICCAI 2019* (pp. 703–711). Springer International Publishing.

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstinet), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

H. Kassani, Hosseinzadeh Kassani, P., Wesolowski, M., Schneider, K. and Deters, R. (2019). Classification of histopathological biopsy images using ensemble of deep learning networks.

Katouzian, A., Karamalis, A., Sheet, D., Konofagou, E., Baseri, B., Carlier, S. G., Eslami, A., König, A., Navab, N., & Laine, A. F. (2012). Iterative self-organizing atherosclerotic tissue labeling in intravascular ultrasound images and comparison with virtual histology. *IEEE Transactions on Biomedical Engineering*, *59*(11), 3039–3049.

Kay, M., Kola, T., Hullman, J. R. and Munson, S. A. (2016, May). When (ish) is my bus? user-centered visualizations of uncertainty in everyday, mobile predictive systems. In *Proceedings of the 2016 chi conference on human factors in computing systems*. (pp. 5092–5103). ACM digital Library.

Khan, A. M., El-Daly, H. and Rajpoot, N. M. (2012), A gamma-Gaussian mixture model for detection of mitotic cells in breast cancer histopathology images. In Pattern Recognition (ICPR), 2012 21st International Conference on, IEEE, pp. 149–152.

Kindermans, P.-J., Hooker, S., Adebayo, J., Alber, M., Schütt, K. T., Dähne, S., Erhan, D. and Kim, B. (2017), The (un)reliability of saliency methods', Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)11700 LNCS, 267–280. http://arxiv.org/abs/1711.00867

Ko, B. C., Gim, J.-W., & Nam, J.-Y. (2011). Automatic white blood cell segmentation using stepwise merging rules and gradient vector flow snake. *Micron*, *42*(7), 695–705. http://www.sciencedirect.com/science/article/pii/S0968432811000576

Kohl, M., Walz, C., Ludwig, F., Braunewell, S. and Baust, M. (2018), Assessment of breast cancer histology using densely connected convolutional networks, CoRRabs/1804.04595. http://arxiv.org/abs/1804.04595

Komura, D., & Ishikawa, S. (2018). Machine learning methods for histopathological image analysis. *Computational and Structural Biotechnology Journal*, *16*, 34–42.

Koné, I. and Boulmane, L. (2018), Hierarchical resnext models for breast cancer histology image classification, CoRRabs/1810.09025. http://arxiv.org/abs/1810.09025

Koohbanani, N. A., Unnikrishnan, B., Khurram, S. A., Krishnaswamy, P., & Rajpoot, N. (2021). Self-path: Self-supervision for classification of pathology images with limited annotations. *IEEE Transactions on Medical Imaging*, *40*(10), 2845–2856.

Krause, J., Dasgupta, A., Swartz, J., Aphinyanaphongs, Y. and Bertini, E. (2017), A workflow for visual diagnostics of binary classifiers using instance-level explanations, 2017 IEEE Conference on Visual Analytics Science and Technology, VAST 2017: Proceedings. pp. 162–172. http://arxiv.org/abs/1705.01968

Kulesza, T., Stumpf, S., Burnett, M., Yang, S., Kwan, I. and Wong, W.-K. (2013), Too much, too little, or just right? ways explanations impact end users mental models, pp. 3–10. http://openaccess.city.ac.uk/

Lage, I., Chen, E., He, J., Narayanan, M., Kim, B., Gershman, S. J. and Doshi-Velez, F. (2019), Human evaluation of models built for interpretability, Proceedings of the AAAI Conference on Human Computation and Crowdsourcing **7**, 59–67. www.aaai.org

Lakhani, P., Kim, W., & Langlotz, C. P. (2012). Automated extraction of critical test values and communications from unstructured radiology reports: An analysis of 9.3 million reports from 1990 to 2011. *Radiology*, *265*(3), 809–818. https://doi.org/10.1148/radiol.12112438

Lal, S., Das, D., Alabhya, K., Kanfade, A., Kumar, A., & Kini, J. (2021). Nucleisegnet: Robust deep learning architecture for the nuclei segmentation of liver cancer histopathology images. *Computers in Biology and Medicine*, *128*, 104075. https://www.sciencedirect.com/science/article/pii/S0010482520304066

Lee, K., Niemeijer, M., Garvin, M. K., Kwon, Y. H., Sonka, M., & Abramoff, M. D. (2010). Segmentation of the optic disc in 3-d oct scans of the optic nerve head. *IEEE Transactions on Medical Imaging*, *29*(1), 159–168.

Li, B., Li, Y. and Eliceiri, K. W. (2021a), Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 14318–14328.

Li, C., Chen, H., Zhang, L., Xu, N., Xue, D., Hu, Z., Ma, H., & Sun, H. (2019a). Cervical histopathology image classification using multilayer hidden conditional random fields and weakly supervised learning. *IEEE Access*, *7*, 90378–90397.

Li, G., Raza, S. E. A., & Rajpoot, N. M. (2017). Multi-resolution cell orientation congruence descriptors for epithelium segmentation in endometrial histology images. *Medical Image Analysis*, *37*, 91–100.

Li, J., Chen, W., Huang, X., Yang, S., Hu, Z., Duan, Q., Metaxas, D. N., Li, H., & Zhang, S. (2021b). Hybrid supervision learning for pathology whole slide image classification. In M. de Bruijne, P. C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, & C. Essert (Eds.), *Medical image computing and computer assisted intervention: MICCAI 2021* (pp. 309–318). Springer International Publishing.

Li, J., Li, W., Sisk, A., Ye, H., Wallace, W. D., Speier, W., & Arnold, C. W. (2021c). A multi-resolution model for histopathology image classification and localization with multiple instance learning. *Computers in Biology and Medicine*, *131*, 104253. https://www.sciencedirect.com/science/article/pii/S0010482521000470

Li, J., Lin, T., & Xu, Y. (2021d). Sslp: Spatial guided self-supervised learning on pathological images. In M. de Bruijne, P. C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, & C. Essert (Eds.), *Medical image computing and computer assisted intervention: MICCAI 2021* (pp. 3–12). Springer International Publishing.

Li, X., Radulovic, M., Kanjer, K., & Plataniotis, K. N. (2019b). Discriminative pattern mining for breast cancer histopathology image classification via fully convolutional autoencoder. *IEEE Access*, *7*, 36433–36445.

Li, Y., Wu, J., & Wu, Q. (2019c). Classification of breast cancer histology images using multi-size and discriminative patches based on deep learning. *IEEE Access*, *7*, 21400–21408.

Li, Y., Wu, X., Li, C., Li, X., Chen, H., Sun, C., Rahaman, M. M., Yao, Y., Zhang, Y., & Jiang, T. (2022). A hierarchical conditional random field-based attention mechanism approach for gastric histopathology image classification. *Applied Intelligence*, *52*, 9717–9738. https://doi.org/10.1007/s10489-021-02886-2

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstinet), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Liang, G., Zhang, Y., Wang, X. and Jacobs, N. (2020), Improved trainable calibration method for neural networks on medical imaging classification. In British machine vision conference (BMVC).

Lin, A., Chen, B., Xu, J., Zhang, Z., & Lu, G. (2022). Ds-transunet: Dual swin transformer u-net for medical image segmentation. In *IEEE Transactions on Instrumentation and Measurement,* (Vol. 71, pp. 1–15). arXiv:2106.06716.

Lin, Y.-S., Lee, W.-C., & Celik, Z. B. (2020). What do you see? Evaluation of explainable artificial intelligence (xai) interpretability through neural backdoors. *arXiv*, 1–10. http://arxiv.org/abs/2009.10639

Liu, C., Shang, F., Ozolek, J. A., & Rohde, G. K. (2016). Detecting and segmenting cell nuclei in two-dimensional microscopy images. *Journal of Pathology Informatics*, *7*, 42.

lopusz (2019), 'Awesome interpretable machine learning'. https://github.com/lopusz/awesome-interpretable-machine-learning.

Lu, M. Y., Chen, R. J., Kong, D., Lipkova, J., Singh, R., Williamson, D. F., Chen, T. Y., & Mahmood, F. (2022). Federated learning for computational pathology on gigapixel whole slide images. *Medical Image Analysis*, *76*, 102298.

Lu, M. Y., Chen, R. J., & Mahmood, F. (2020). Semi-supervised breast cancer histology classification using deep multiple instance learning and contrast predictive coding (conference presentation). In J. E. Tomaszewski & A. D. Ward (Eds.), *Medical Imaging 2020: Digital Pathology. SPIE Proceedings, SPIE* (Vol. 11320, 113200J). SPIE Digital Library.

Lyu, Q., Chen, Y., Li, Z., Cui, Z., Chen, L., Zhang, X. and Shen, H. (2016), Extracting actionability from machine learning models by sub-optimal deterministic planning, arXivabs/1611.00873.

Madabhushi, A., & Lee, G. (2016). Image analysis and machine learning in digital pathology: Challenges and opportunities. *Medical Image Analysis*, *33*, 170–175.

Makarchuk, G., Kondratenko, V., Pisov, M., Pimkin, A., Krivov, E. and Belyaev, M. (2018), Ensembling neural networks for digital pathology images classification and segmentation, CoRRabs/1802.00947. http://arxiv.org/abs/1802.00947

Malon, C., Miller, M., Burger, H. C., Cosatto, E. and Graf, H. P. (2008), Identifying histological elements with convolutional neural networks. In Proceedings of the 5th international conference on soft computing as transdisciplinary science and technology, ACM, 450–456.

Man, R., Yang, P., & Xu, B. (2020). Classification of breast cancer histopathological images using discriminative patches screened by generative adversarial networks. *IEEE Access*, *8*, 155362–155377.

Mohseni, S. and Ragan, E. D. (2018), A human-grounded evaluation benchmark for local explanations of machine learning. https://github.com/SinaMohseni/ML-Interpretability

Muhammad, N., Shaban, M., Snead, D., Khurram, S. A. and Rajpoot, N. (2018), A multi-resolution deep learning framework for lung adenocarcinoma growth pattern classification: 22nd Conference, MIUA 2018, Southampton, UK, July 9–11, 2018, Proceedings, pp. 3–11.

Naik, S., Doyle, S., Agner, S., Madabhushi, A., Feldman, M. and Tomaszewski, J. (2008), Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology. In Biomedical Imaging: From Nano to Macro, 2008. 5th IEEE International Symposium on ISBI 2008, 284–287.

Nazeri, K., Aminpour, A. and Ebrahimi, M. (2018), Two-stage convolutional neural network for breast cancer histology image classification, CoRRabs/1803.04054. http://arxiv.org/abs/1803.04054

Nguyen, K., Jain, A. K. and Allen, R. L. (2010), Automated gland segmentation and classification for Gleason grading of prostate tissue images. In pattern recognition (ICPR), 2010 20th International Conference on IEEE, pp. 1497–1500.

Nguyen, L., Gao, R., Lin, D., & Lin, Z. (2019). Biomedical image classification based on a feature concatenation and ensemble of deep cnns. *Journal of Ambient Intelligence and Humanized Computing*, 1–13.

Niazi, M. K. K., Parwani, A. V., & Gurcan, M. N. (2019). Digital pathology and artificial intelligence. *The Lancet Oncology*, *20*(5), e253–e261.

Nori, H., Jenkins, S., Koch, P., & Caruana, R. (2019). Interpretml: A unified framework for machine learning interpretability. *arXiv*, 1–18.

Obando, D. F. G., Frafjord, A., Øynebråten, I., Corthay, A., Olivo-Marin, J.-C. and Meas-Yedid, V. (2017), Multi-staining registration of large histology images. In '2017 IEEE 14th International symposium on biomedical imaging (ISBI 2017)', IEEE, 345–348.

oracle (2018), 'Skater: Python library for model interpretation/explanations'. https://github.com/oracle/Skater.

Panchal, P., Panchal, S., & Shah, S. (2013). A comparison of sift and surf. *International Journal of Innovative Research in Computer and Communication Engineering*, *1*(2), 323–327.

Pati, P., Jaume, G., Fernandes, L. A., Foncubierta-Rodríguez, A., Feroce, F., Anniciello, A. M., Scognamiglio, G., Brancati, N., Riccio, D., Di Bonito, M., De Pietro, G., Botti, G., Goksel, O., Thiran, J.-P., Frucci, M., & Gabrani, M. (2020). Hact-net: A hierarchical cell-to-tissue graph neural network for histopathological image classification. In C. H. Sudre, H. Fehri, T. Arbel, C. F. Baumgartner, A. Dalca, R. Tanno, K. van Leemput, W. M. Wells, A. Sotiras, B. Papiez, E. Ferrante, & S. Parisot (Eds.), *Uncertainty for safe utilization of machine learning in medical imaging, and graphs in biomedical image analysis* (pp. 208–219). Springer International Publishing.

pbiecek (2019), Interesting resources related to xai (explainable artificial intelligence). https://github.com/pbiecek/xai-resources/blob/master/README.md#theses.

Phoulady, H. A., Zhou, M., Goldgof, D. B., Hall, L. O. and Mouton, P. R. (2016), Automatic quantification and classification of cervical cancer via adaptive nucleus shape modeling. In Image Processing (ICIP), 2016 IEEE International Conference on IEEE, pp. 2658–2662.

Plissiti, M. E., Nikou, C. and Charchanti, A. (2010), Accurate localization of cell nuclei in pap smear images using gradient vector flow deformable models., *in* 'BIOSIGNALS', pp. 284–289.

Pocevičiūtė, M., Eilertsen, G. and Lundström, C. (2020), Survey of xai in digital pathology. In Artificial intelligence and machine learning for digital pathology, Lecture Notes in Computer Science. Springer, 56–88.

Raczkowski, L., Mozejko, M., Zambonelli, J., & Szczurek, E. (2019). Ara: Accurate, reliable and active histopathological image classification framework with bayesian deep learning. *Scientific Reports*, *9*, 14347.

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstinet), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Rakha, E. A., Toss, M., Shiino, S., Gamble, P., Jaroensri, R., Mermel, C. H., & Chen, P.-H. C. (2021). Current and future applications of artificial intelligence in pathology: A clinical perspective. *Journal of Clinical Pathology*, *74*(7), 409–414.

Razavi, S., Khameneh, F. D., Nouri, H., Androutsos, D., Done, S. J., & Khademi, A. (2022). Minugan: Dual segmentation of mitoses and nuclei using conditional gans on multi-center breast h&e images. *Journal of pathology informatics*, *13*, 100002.

Ribeiro, M. T., Singh, S. and Guestrin, C. (2016), "Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. (pp. 1135–1144). ACM digital Library.

Ribeiro, M. T., Singh, S. and Guestrin, C. (2018), Anchors: High-precision model-agnostic explanations. In Proceedings of the AAAI Conference on Artificial Intelligence, *32*.

Robertson, S., Azizpour, H., Smith, K., & Hartman, J. (2018). Digital image analysis in breast pathology-from image processing techniques to artificial intelligence. *Translational Research*, *194*, 19–35.

Ronneberger, O., Fischer, P. and Brox, T. (2015), U-net: Convolutional networks for biomedical image segmentation. In 'International Conference on Medical image computing and computer-assisted intervention', Springer, pp. 234–241.

Ross, A. S., Hughes, M. C. and Doshi-Velez, F. (2017), Right for the right reasons: Training differentiable models by constraining their explanations, International Joint Conferences on Artificial Intelligence, pp. 2662–2670. http://arxiv.org/abs/1703.03717

Rossetti, B. J., Wang, F., Zhang, P., Teodoro, G., Brat, D. J. and Kong, J. (2017), Dynamic registration for gigapixel serial whole slide images. In '2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017)', IEEE, pp. 424–428.

Roy, K., Banik, D., Bhattacharjee, D., & Nasipuri, M. (2019). Patch-based system for classification of breast histology images using deep learning. *Computerized Medical Imaging and Graphics*, *71*, 90–103. https://www.sciencedirect.com/science/article/pii/S0895611118302039

Sahasrabudhe, M., Christodoulidis, S., Salgado, R., Michiels, S., Loi, S., André, F., Paragios, N., & Vakalopoulou, M. (2020). Self-supervised nuclei segmentation in histopathological images using attention. In A. L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M. A. Zuluaga, S. K. Zhou, D. Racoceanu, & L. Joskowicz (Eds.), *Medical image computing and computer assisted intervention: MICCAI 2020* (pp. 393–402). Springer International Publishing.

Saillard, C., Dehaene, O., Marchand, T. and Moindrot, O. (2021), '*Self supervised learning improves dmmr/msi detection from histology slides across multiple cancers*', pp. 1–16.

Salvi, M., & Molinari, F. (2018). Multi-tissue and multi-scale approach for nuclei segmentation in h&e stained images. *Biomedical Engineering Online*, *17*, 89.

Samek, W., Binder, A., Montavon, G., Bach, S., & Müller, K.-R. (2015). Evaluating the visualization of what a deep neural network has learned. *IEEE Transactions on Neural Networks and Learning Systems*, *28*, 2660–2673. http://arxiv.org/abs/1509.06321

Schirris, Y., Gavves, E., Nederlof, I., Horlings, H. M., & Teuwen, J. (2022). Deepsmile: Contrastive self-supervised pre-training benefits msi and hrd classification directly from h&e whole-slide images in colorectal and breast cancer. *Medical Image Analysis*, *79*, 102464.

Schultz, S., Krüger, J., Handels, H. and Ehrhardt, J. (2019), Bayesian inference for uncertainty quantification in point-based deformable image registration. In Medical Imaging 2019: Image Processing, Vol. 10949, International Society for Optics and Photonics, 109491S.

Schwier, M., Böhler, T., Hahn, H. K., Dahmen, U., & Dirsch, O. (2013). Registration of histological whole slide images guided by vessel structures. *Journal of pathology informatics*, *4*, 10.

Sekhar, S. C., Aguet, F., Romain, S., Thévenaz, P. and Unser, M. (2008), Parametric b-spline snakes on distance maps-application to segmentation of histology images. In Signal processing conference, 2008 16th European, IEEE, pp. 1–5.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2019). Grad-cam: Visual explanations from deep networks via gradient-based localization. *International Journal of Computer Vision*, *128*(2), 336–359. https://doi.org/10.1007/s11263-019-01228-7

Senaras, C., Niazi, M. K. K., Sahiner, B., Pennell, M. P., Tozbikian, G., Lozanski, G., & Gurcan, M. N. (2018). Optimized generation of high-resolution phantom images using cgan: Application to quantification of ki67 breast cancer images. *PLoS One*, *13*(5), e0196846.

Senousy, Z., Abdelsamea, M., Gaber, M. M., Abdar, M., Acharya, R. U., Khosravi, A., & Nahavandi, S. (2021a). Mcua: Multi-level context and uncertainty aware dynamic deep ensemble for breast cancer histology image classification. *IEEE Transactions on Biomedical Engineering*, *69*(2), 818–829.

Senousy, Z., Abdelsamea, M. M., Mohamed, M. M., & Gaber, M. M. (2021b). 3e-net: Entropy-based elastic ensemble of deep convolutional neural networks for grading of invasive breast carcinoma histopathological microscopic images. *Entropy*, *23*(5), 1–21. https://www.mdpi.com/1099-4300/23/5/620

Sertel, O., Kong, J., Catalyurek, U. V., Lozanski, G., Saltz, J. H., & Gurcan, M. N. (2009). Histopathological image analysis using model-based intermediate representations and color texture: Follicular lymphoma grading. *Journal of Signal Processing Systems*, *55*(1–3), 169–183.

Shaban, M., Awan, R., Fraz, M., Azam, A., Snead, D., & Rajpoot, N. (2020). Context-aware convolutional neural network for grading of colorectal cancer histology images. *IEEE Transactions on Medical Imaging*, *39*(7), 2395–2405.

Shaban, M. T., Baur, C., Navab, N., & Albarqouni, S. (2019). Staingan: Stain style transfer for digital histological images. In 2019 IEEE 16th international symposium on biomedical imaging (Isbi 2019), IEEE. pp. 953–956.

Shahul Hameed, K., Banumathi, A., & Ulaganathan, G. (2017). A simple method of immunostained tissue scoring based on maximal separation techniques. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, *6*, 1–7.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423.

Shao, Z., Bian, H., Chen, Y., Wang, Y., Zhang, J., Ji, X., & Zhang, Y. (2021). Transmil: Transformer based correlated multiple instance learning for whole slide image classification. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, & J. W. Vaughan (Eds.), *Advances in neural information processing systems* (Vol. 34, pp. 2136–2147). Curran Associates, Inc.. https://proceedings.neurips.cc/paper/2021/file/10c272d06794d3e5785d5e7c5356e9ff-Paper.pdf

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstinet), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Sharma, Y., Shrivastava, A., Ehsan, L., Moskaluk, C. A., Syed, S., & Brown, D. (2021). Cluster-to-conquer: A framework for end-to-end multi-instance learning for whole slide image classification. In M. Heinrich, Q. Dou, M. de Bruijne, J. Lellmann, A. Schläfer, & F. Ernst (Eds.), *Proceedings of the Fourth Conference on Medical Imaging with Deep Learning', Vol. 143 of Proceedings of Machine Learning Research, PMLR* (pp. 682–698). PMLR. https://proceedings.mlr.press/v143/sharma21a.html

Shen, Y., & Ke, J. (2020). A deformable crf model for histopathology whole-slide image classification. In A. L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M. A. Zuluaga, S. K. Zhou, D. Racoceanu, & L. Joskowicz (Eds.), *Medical image computing and computer assisted intervention: MICCAI 2020* (pp. 500–508). Springer International Publishing.

Shrivastava, A., Adorno, W., Sharma, Y., Ehsan, L., Ali, S. A., Moore, S. R., Amadi, B., Kelly, P., Syed, S. and Brown, D. E. (2021), Self-attentive adversarial stain normalization, In International conference on pattern recognition, springer, pp. 120–140.

sicara (2019), tf-explain: tf-explain documentation. https://tf-explain.readthedocs.io/en/latest/.

Sirinukunwattana, K., Raza, S. E. A., Tsang, Y.-W., Snead, D. R., Cree, I. A., & Rajpoot, N. M. (2016). Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. *IEEE Transactions on Medical Imaging*, *35*(5), 1196–1206.

Sirinukunwattana, K., Snead, D. R., & Rajpoot, N. M. (2015). A stochastic polygons model for glandular structures in colon histology images. *IEEE Transactions on Medical Imaging*, *34*(11), 2366–2378.

Smilkov, D., Thorat, N., Kim, B., Viégas, F. and Wattenberg, M. (2017), *Smoothgrad: Removing noise by adding noise*, pp. 1–10.

Sommer, C., Fiaschi, L., Hamprecht, F. A. and Gerlich, D. W. (2012), Learning-based mitotic cell detection in histopathological images. In Pattern recognition (ICPR), 2012 21st International Conference on, IEEE, 2306–2309.

Sornapudi, S., Stanley, R., Stoecker, W., Almubarak, H., Long, L., Antani, S., Thoma, G., Zuna, R. and Frazier, S. (2018), Deep learning nuclei detection in digitized histology images by superpixels, *9*, 5.

Springenberg, J. T., Dosovitskiy, A., Brox, T. and Riedmiller, M. (2015), '*Striving for simplicity: The all convolutional net*', pp. 1–14.

Srinidhi, C. L., Ciga, O., & Martel, A. L. (2021). Deep neural network models for computational histopathology: A survey. *Medical Image Analysis*, *67*, 101813.

Srinidhi, C. L., Kim, S. W., Chen, F.-D., & Martel, A. L. (2022). Self-supervised driven consistency training for annotation efficient histopathology image analysis. *Medical Image Analysis*, *75*, 102256. https://www.sciencedirect.com/science/article/pii/S1361841521003017

Stacke, K., Eilertsen, G., Unger, J. and Lundström, C. (2019), '*A closer look at domain shift for deep learning in histopathology*', pp. 1–8.

Sun, C., Xu, A., Liu, D., Xiong, Z., Zhao, F., & Ding, W. (2020a). Deep learning-based classification of liver cancer histopathology images using only global labels. *IEEE Journal of Biomedical and Health Informatics*, *24*(6), 1643–1651.

Sun, J., Darbehani, F., Zaidi, M. and Wang, B. (2020b), Saunet: Shape attentive u-net for interpretable medical image segmentation. In International conference on medical image computing and computer-assisted intervention, Springer, pp. 797–806.

Sundararajan, M., Taly, A. and Yan, Q. (2017). Axiomatic attribution for deep networks. In *International conference on machine learning*. (pp. 3319–3328). PMLR.

Takamatsu, M., Yamamoto, N., Kawachi, H., Nakano, K., Saito, S., Fukunaga, Y., & Takeuchi, K. (2022). Prediction of lymph node metastasis in early colorectal cancer based on histologic images by artificial intelligence. *Scientific Reports*, *12*(1), 1–11.

Tosta, T. A. A., de Faria, P. R., Neves, L. A., & Nascimento, M. Z. (2019). Computational normalization of h&e-stained histological images: Progress, challenges and future potential. *Artificial Intelligence in Medicine*, *95*, 118–132.

van den Oord, A., Li, Y., & Vinyals, O. (2018). Representation learning with contrastive predictive coding. *arXiv e-prints*, arXiv-1807, pp. 1–13.

Vang, Y. S., Chen, Z., & Xie, X. (2018). Deep learning framework for multi-class breast cancer histology image classification. In A. Campilho, F. Karray, & B. ter Haar Romeny (Eds.), *Image analysis and recognition* (pp. 914–922). Springer International Publishing.

Venet, L., Pati, S., Yushkevich, P., & Bakas, S. (2019). Accurate and robust alignment of variable-stained histologic images using a general-purpose greedy diffeomorphic registration tool. *arXiv preprint*, arXiv:1904.11929, 1–3.

Veta, M., Pluim, J. P., Van Diest, P. J., & Viergever, M. A. (2014). Breast cancer histopathology image analysis: A review. *IEEE Transactions on Biomedical Engineering*, *61*(5), 1400–1411.

Wan, T., Zhao, L., Feng, H., Li, D., Tong, C., & Qin, Z. (2020). Robust nuclei segmentation in histopathology using asppu-net and boundary refinement. *Neurocomputing*, *408*, 144–156.

Wang, J., Chen, R. J., Lu, M. Y., Baras, A. and Mahmood, F. (2020), Weakly supervised prostate tma classification via graph convolutional networks. In 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), pp. 239–243.

Wang, P., Li, P., Li, Y., Xu, J., & Jiang, M. (2022). Classification of histopathological whole slide images based on multiple weighted semi-supervised domain adaptation. *Biomedical Signal Processing and Control*, *73*, 103400. https://www.sciencedirect.com/science/article/pii/S1746809421009976

Wang, X., Chen, H., Gan, C., Lin, H., Dou, Q., Huang, Q., Cai, M. and Heng, P. (2018), Weakly supervised learning for whole slide lung cancer image classification. In *Medical Imaging with Deep Learning*. https://openreview.net/forum?id=SJwod1hjz

Wang, X., Yang, S., Zhang, J., Wang, M., Zhang, J., Huang, J., Yang, W., & Han, X. (2021). Transpath: Transformer-based self-supervised learning for histopathological image classification. In M. de Bruijne, P. C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, & C. Essert (Eds.), *Medical image computing and computer assisted intervention – MICCAI 2021* (pp. 186–195). Springer International Publishing.

Wang, Y., Crookes, D., Diamond, J., Hamilton, P. and Turner, R. (2007), Segmentation of squamous epithelium from ultra-large cervical histological virtual slides. In Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE, IEEE, pp. 775–778.

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstinet), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Wodzinski, M., & Skalski, A. (2019). Automatic nonrigid histological image registration with adaptive multistep algorithm. *arXiv preprint*, 1–4. arXiv:1904.00982.

Wu, H.-S., Xu, R., Harpaz, N., Burstein, D., & Gil, J. (2005). Segmentation of intestinal gland images with iterative region growing. *Journal of Microscopy*, *220*(3), 190–204.

Wulczyn, E., Steiner, D. F., Xu, Z., Sadhwani, A., Wang, H., Flament-Auvigne, I., Mermel, C. H., Chen, P.-H. C., Liu, Y., & Stumpe, M. C. (2020). Deep learning-based survival prediction for multiple cancer types using histopathology images. *PLoS One*, *15*(6), e0233678.

Xiang, T., Song, Y., Zhang, C., Liu, D., Chen, M., Zhang, F., Huang, H., O'Donnell, L., & Cai, W. (2022). Dsnet: A dual-stream framework for weakly-supervised gigapixel pathology image analysis. *IEEE Transactions on Medical Imaging*, 1–11.

Xie, X., Li, Y., Zhang, M., & Shen, L. (2019). Robust segmentation of nucleus in histopathology images via mask r-cnn. In A. Crimi, S. Bakas, H. Kuijf, F. Keyvan, M. Reyes, & T. van Walsum (Eds.), *Brainlesion: Glioma, multiple sclerosis, stroke and traumatic brain injuries* (pp. 428–436). Springer International Publishing.

Xing, F., Xie, Y., Su, H., Liu, F., & Yang, L. (2017). Deep learning in microscopy image analysis: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, *29*(10), 4550–4568.

Xu, J., Janowczyk, A., Chandran, S., & Madabhushi, A. (2011). A high-throughput active contour scheme for segmentation of histopathological imagery. *Medical Image Analysis*, *15*(6), 851–862.

Xue, Y., Ye, J., Zhou, Q., Long, L. R., Antani, S., Xue, Z., Cornwell, C., Zaino, R., Cheng, K. C., & Huang, X. (2021). Selective synthetic augmentation with histogan for improved histopathology image classification. *Medical Image Analysis*, *67*, 101816. https://www.sciencedirect.com/science/article/pii/S1361841520301808

Yan, K., Wang, X., Lu, L., & Summers, R. M. (2018). DeepLesion: Automated mining of large-scale lesion annotations and universal lesion detection with deep learning. *Journal of Medical Imaging*, *5*(3), 1–11. https://doi.org/10.1117/1.JMI.5.3.036501

Yan, R., Ren, F., Zihao, W., Wang, L., Zhang, T., Liu, Y., Rao, X., Zheng, C., & Zhang, F. (2020). Breast cancer histopathological image classification using a hybrid deep neural network. *Methods*, *173*, 52–60.

Yang, P., Hong, Z., Yin, X., Zhu, C., & Jiang, R. (2021). Self-supervised visual representation learning for histopathological images. In M. de Bruijne, P. C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, & C. Essert (Eds.), *Medical image computing and computer assisted intervention: MICCAI 2021* (pp. 47–57). Springer International Publishing.

Yang, Z., Ran, L., Zhang, S., Xia, Y., & Zhang, Y. (2019). Ems-net: Ensemble of multiscale convolutional neural networks for classification of breast cancer histology images. *Neurocomputing*, *366*, 46–53. http://www.sciencedirect.com/science/article/pii/S0925231219310872

Yin, M., Vaughan, J. W., & Wallach, H. (2019). *Understanding the effect of accuracy on trust in machine learning models*. Association for Computing Machinery, pp. 1–12. https://doi.org/10.1145/3290605.3300509

Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. In D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars, (Eds.), *Computer Vision: ECCV 2014. ECCV 2014. Lecture Notes in Computer Science* (Vol. 8689). Springer. https://doi.org/10.1007/978-3-319-10590-1_53.

Zeng, Z., Xie, W., Zhang, Y., & Lu, Y. (2019). Ric-unet: An improved neural network based on unet for nuclei segmentation in histology images. *IEEE Access*, *7*, 21420–21428.

Zhang, X., Zhu, X., Tang, K., Zhao, Y., Lu, Z., & Feng, Q. (2022). Ddtnet: A dense dual-task network for tumor-infiltrating lymphocyte detection and segmentation in histopathological images of breast cancer. *Medical Image Analysis*, *78*, 102415.

Zhao, Y., Yang, F., Fang, Y., Liu, H., Zhou, N., Zhang, J., Sun, J., Yang, S., Menze, B., Fan, X. and Yao, J. (2020), Predicting lymph node metastasis using histopathological images based on multiple instance learning with deep graph convolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

Zhou, C., Jin, Y., Chen, Y., Huang, S., Huang, R., Wang, Y., Zhao, Y., Chen, Y., Guo, L., & Liao, J. (2021). Histopathology classification and localization of colorectal cancer using global labels by weakly supervised deep learning. *Computerized Medical Imaging and Graphics*, *88*, 101861. https://www.sciencedirect.com/science/article/pii/S0895611121000094

Zhou, Y., Graham, S., Koohbanani, N. A., Shaban, M., Heng, P.-A., & Rajpoot, N. (2019a). Cgc-net: Cell graph convolutional network for grading of colorectal cancer histology images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops* (pp. 1–11). https://doi.org/10.48550/arXiv.1909.01068.

Zhou, Y., Onder, O. F., Dou, Q., Tsougenis, E., Chen, H., & Heng, P.-A. (2019b). Cia-net: Robust nuclei instance segmentation with contour-aware information aggregation. In A. C. S. Chung, J. C. Gee, P. A. Yushkevich, & S. Bao (Eds.), *Information processing in medical imaging* (pp. 682–693). Springer International Publishing.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Abdelsamea, M. M., Zidan, U., Senousy, Z., Gaber, M. M., Rakha, E., & Ilyas, M. (2022). A survey on artificial intelligence in histopathology image analysis. *WIREs Data Mining and Knowledge Discovery*, *12*(6), e1474. https://doi.org/10.1002/widm.1474

19424795, 2022, 6, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1474 by Egyptian National Sci. Network (Enstinet), Wiley Online Library on [10/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License